



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Universite d'Ottawa

Modelisation spatio-dynamique
de la structure urbaine et
du systeme de transport

Par

Stéphane Brice

Département de Géographie

Faculté des Arts

Thèse présentée à

l'Ecole des Etudes Supérieures et de la Recherche
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)



Stéphane Brice, Ottawa, Canada, 1990



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-62328-4

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

TABLE DES MATIERES

	page
Sommaire	x
Abstract	xi
Remerciements	xii
<u>1 INTRODUCTION</u>	1
1.1 CARACTERISTIQUES DE LA THESE	1
1.2 LES PREMICES D'UNE INTEGRATION	2
1.2.1 Problematique generale	2
1.2.2 Le contexte de recherche	3
1.3 LES MODELES URBAINS INTEGRES	8
1.4 PRESENTATION DES CHAPITRES	9
Première partie	
MODELISER LE SYSTEME DE TRANSPORT URBAIN	
<u>2 ASPECTS THEORIQUES DE LA PROGRAMMATION MATHEMATIQUE</u>	15
2.1 NECESSITE D'UNE INTRODUCTION THEORIQUE	15
2.2 PROPRIETES GENERALES D'UN PROBLEME D'OPTIMISATION	17
2.2.1 Elements et structures d'un probleme d'optimisation mathematique	17
2.2.2 Maxima et minima d'une fonction	18
2.2.3 Proprietes de convexite et de concavite	20
2.2.4 Typologie des problemes d'optimisation	22
2.3 DETERMINATION DES CONDITIONS D'OPTIMALITE	25
2.3.1 Optimisation sans contrainte	25
2.3.2 Optimisation avec contraintes d'egalite	27
2.3.3 Optimisation avec contraintes d'inegalite	30
<u>3 MODELISATION DE LA DEMANDE DE TRANSPORT</u>	33
3.1 INTRODUCTION	33

3.2	GENERATION DES DEPLACEMENTS	35
3.2.1	Definition du deplacement	35
3.2.2	Approche par activite urbaine	36
3.2.3	Mobilite individuelle et caracteristiques de population	39
3.2.4	Approche mixte	41
3.3	DISTRIBUTION SPATIALE DES DEPLACEMENTS	44
3.3.1	Modele analogique	44
3.3.2	Modeles issus de la maximisation d'entropie	45
3.3.3	Modeles issus d'un probleme d'optimisation	48
3.3.4	Signification geographique des multiplicateurs de Lagrange	52
3.4	MODELISATION DE LA REPARTITION MODALE	53
3.4.1	Aspects fondamentaux	53
3.4.2	Determination de la fonction d'utilite	54
3.4.3	Modeles econometriques	55
3.4.4	Modele derivant d'un probleme de maximisation d'utilite	60
3.5	INTEGRATION DES DISTRIBUTIONS SPATIALE ET MODALE	62
3.5.1	Distributions simultanees	62
3.5.2	Partage modal prioritaire	63
3.5.3	Structure hierarchisee	66
3.6	MODELES D'AFFECTION ROUTIERE	71
3.6.1	Approches deterministes	71
3.6.2	Approche par trafic d'equilibre	73
3.7	INTEGRATION DE LA DISTRIBUTION SPATIALE, DE LA REPARTITION MODALE ET DE L'AFFECTION ROUTIERE	78
3.7.1	Inadéquation de l'approche par interface	78
3.7.2	Le probleme combinant distribution et affectation	78
3.7.3	Version multi-modale du probleme de distribution - affectation	81
3.8	METHODES D'ESTIMATION POUR LES MODELES DE DEMANDE DE TRANSPORT	85
3.8.1	Introduction	85
3.8.2	Analyse de regression multiple	86
3.8.3	Calibration des modeles de distributions spatiale et modale	90
3.8.4	Algorithme de solution du modele integre	96
3.A	APPENDICE A: Derivation du modele dogit de repartition modale a partir d'un programme mathematique	103
3.B	APPENDICE B: Formulation d'un probleme d'affectation routiere avec agregation de reseau	105

<u>4 MODELISATION DE L'OFFRE DE TRANSPORT</u>	109
4.1 CONCEPTUALISATION	109
4.1.1 Definition de l'offre de transport	109
4.1.2 Un cadre conceptuel pour l'offre de transport urbain	112
4.1.3 Structure d'un modele generique d'offre de transport	116
4.2 PROBLEME DE CONCEPTION DE RESEAU	119
4.2.1 Formulation avec coûts fixes de transport	119
4.2.2 Formulation avec réseau congestionné	120
4.2.3 Incorporation d'une demande inter-zonale variable	122
4.2.4 Application multi-modale	124
4.3 ALGORITHMES POUR LE PROBLEME DE CONCEPTION DE RESEAU	127
4.3.1 Solution du problème avec coûts fixes	127
4.3.2 Algorithme avec trafic d'équilibre	131
4.3.3 Distribution spatiale et repartition modale	133
4.4 CONCLUSION SUR LE CADRE DE MODELISATION DU TRANSPORT	136

Deuxième partie
MODELISER LA STRUCTURE URBAINE

<u>5 ANALYSE DES SYSTEMES SPATIO-TEMPORELS</u>	139
5.1 INTRODUCTION	139
5.2 LES SYSTEMES PUREMENT TEMPORELS	141
5.2.1 Introduction	141
5.2.2 Processus autoregressifs	141
5.2.3 Processus de moyennes mobiles	146
5.2.4 Processus de décalage distribué et fonctions de transfert	148
5.2.5 Systèmes temporels multivariés	152
5.3 MODELISATION DE PROCESSUS SPATIO-TEMPORELS	155
5.3.1 Représentation générale de systèmes spatio-temporels	155
5.3.2 Spécification de la structure spatiale	156
5.3.3 Formulation de modèles purement spatiaux	160
5.3.4 Conciliation des séries temporelles et des modèles spatiaux	164

5.4	ESTIMATION D'UN SYSTEME SPATIO-TEMPOREL	168
5.4.1	Paramètres de systemes temporels	168
5.4.2	Paramètres de modèles spatiaux	172
5.4.3	Les problèmes liés à la structure spatiale	177
6	<u>MODELISATION SPATIO-DYNAMIQUE DE LA STRUCTURE URBAINE</u>	182
6.1	INTRODUCTION	182
6.2	MODELES CLASSIQUES DE LOCALISATION	184
6.2.1	Un modèle urbain de synthèse; le modèle de Lowry	184
6.2.2	Modèles de localisation commerciale	187
6.2.3	Modèles de localisation résidentielle	189
6.3	PROBLEMES D'OPTIMISATION ET LOCALISATION D'ACTIVITE	192
6.3.1	Formulation normative	192
6.3.2	Intégration de l'entropie dans un problème de localisation	195
6.3.3	Benefices de localisation	198
6.3.4	Le modèle de Lowry en tant que modèle de bénéfice de localisation	207
6.3.5	Etat d'équilibre de la structure urbaine	208
6.4	CONVERSION SPATIO-DYNAMIQUE DU MODELE DE LOWRY	216
6.4.1	Tentatives d'incorporation de la dimension du temps	216
6.4.2	Dynamique d'un système gradient	218
6.5	ANALYSE DYNAMIQUE ET PROPRIETES DE BIFURCATION	225
6.5.1	Quelques notions	225
6.5.2	Stabilité structurelle des modèles de localisation	232
6.5.3	Stabilité dynamique d'un processus co-évolutif de localisation	241
6.5.4	Notes concernant un modèle opérationnel	248
6.6	CONCILIATION DES MODELES DE TRANSPORT ET DE LOCALISATION	250
6.6.1	Cadre conceptuel	250
6.6.2	Cadre opérationnel	252
6.6.3	Les implications d'une conciliation	253
6.A	APPENDICE A: Equilibre statique du modèle dynamique de localisation	257
6.B	APPENDICE B: Solution d'un système dynamique approximant	260

Troisième partie
UNE APPLICATION DU CADRE DE MODELISATION

<u>7 OPERATIONNALISATION DU CADRE DE MODELISATION</u>	263
7.1 INTRODUCTION A L'APPLICATION	263
7.1.1 Objectifs de l'application	263
7.1.2 La region metropolitaine de Montreal comme terrain d'essai	265
7.2 DEFINITION ET SPECIFICATION DE LA BASE DE DONNEES	267
7.2.1 Systeme spatial	267
7.2.2 Horizon temporel et periodes	268
7.2.3 Definition des reseaux de transport	269
7.2.4 Enumeration des itineraires	273
7.2.5 Données de demande de transport	274
7.2.6 Données d'activites urbaines	277
7.3 PRESENTATION DU MODELE OPERATIONNALISE	282
7.3.1 Introduction	282
7.3.2 Generation spatiale des deplacements	282
7.3.3 Distributions spatiale, modale et routiere	285
7.3.4 Solution du probleme de demande de transport	290
7.3.5 Phase de modification des reseaux de transport	292
7.3.6 Relocalisation des activites	293
7.A ANNEXE: Liste des zones de la region etudiee	299
7.B ANNEXE: Identification du reseau routier retenu - L(1)	302
7.C ANNEXE: Identification du reseau de transport collectif retenu - L(2)	304
<u>8 APPLICATION A LA REGION DE MONTREAL</u>	305
8.1 LES TROIS SCENARIOS DE SIMULATION	305
8.1.1 Introduction	305
8.1.2 Description du scenario A	307
8.1.3 Description du scenario B	309
8.1.4 Description du scenario C: le plan "Côté"	312

8.2	COMMENTAIRES SUR LES SIMULATIONS	316
8.2.1	Développement des réseaux	316
8.2.2	Localisation future des activités	323
8.2.3	Répartition modale	340
8.2.4	Congestion routière et coût de transport	351
8.3	COMMENTAIRES SUR LES PERFORMANCES DU MODELE	368
8.3.1	L'agregation spatiale	368
8.3.2	L'agregation des réseaux	371
8.3.3	Période d'étude et agregation temporelle	373
8.3.4	Les sous-modèles de transport	375
8.3.5	Les sous-modèles de localisation	381
8.3.6	Integration des systèmes de transport et de la localisation d'activité	385
8.4	REFLEXIONS SUR L'UTILITE EMPIRIQUE DE LA MODELISATION	388
8.4.1	Les limites	388
8.4.2	Les avantages	390
9	<u>CONCLUSION</u>	392
9.1	CONTEXTE DE LA MODELISATION	392
9.1.1	Cadre general	392
9.1.2	Modelisation de la demande de transport	393
9.1.3	Modelisation de l'offre de transport	395
9.1.4	Le processus de localisation	397
9.1.5	La conversion vers un système dynamique	398
9.1.6	Conciliation de la structure urbaine et du système de transport	401
9.1.7	L'application a la région de Montréal	402
9.2	PERSPECTIVES DE DEVELOPPEMENT	405
9.2.1	Vers des formulations plus indulgentes	405
9.2.2	Bonification des sous-modèles	408
9.2.3	Vers de nouvelles interpretations	411
9.2.4	Algorithmes de solution et procédures d'estimation	413
	<u>REFERENCES BIBLIOGRAPHIQUES</u>	417

LISTES DES TABLEAUX

7.1	Resultats de l'estimation du modele d'origine	284
7.2	Resultats de l'estimation du modele de destination	285
7.3	Resultats de la calibration modale-spatiale	290
7.4	Parametres des modeles spatio-temporels de localisation	296
8.1	Liste des liens projetés de transport en commun L ^P (2)	308
8.2	Liste des liens routiers projetés L ^P (1)	310
8.3	Projets de transport du «Plan d'action 1988-1998»	315
8.4	Statistiques globales des simulations: scenario A	317
8.5	Statistiques globales des simulations: scenario B	320
8.6	Statistiques globales des simulations: scenario C	321
8.7	Evolution de l'emploi exogene des sous-regions selon les previsions de Lamonde et al. [1988]	327
8.8	Evolution démographique des sous-regions selon les scenarios A, B et C	328
8.9	Evolution de l'emploi de service des sous-regions selon les scenarios A, B et C	329
8.10	Niveau simulé de congestion du réseau routier: scenario A	353
8.11	Niveau simulé de congestion du réseau routier: scenario B	354
8.12	Niveau simulé de congestion du réseau routier: scenario C	355

LISTE DES FIGURES

5.1	Représentation d'une fonction de transfert temporelle	151
6.1	Trajectoire d'un système dynamique convergent	221
6.2	Formes génériques d'un polynôme cubique selon différentes valeurs de a : $f(x) = ax^3 + bx^2 + cx + d$	228
6.3	Formes génériques d'un polynôme quartique selon différentes valeurs paramétriques de $f(y) = ay^4 + by^3 + cy^2 + dy + e$	228
6.4	Courbes R_1-S_1 versus S_1	239
6.5	Courbes d'équilibre pour " α ": a) zone avantageusement localisée; b) zone marginale	239
6.6	Courbes d'équilibre pour " ψ ": a) $\alpha < 1$; b) $\alpha = 1$; c) $\alpha > 1$	239
6.7	Courbes d'équilibre pour " α " pour un système à 4 zones	239
6.8	Trajectoire dynamique des variables d'emploi de service (S_1) et de population (P_1) tirées d'une même zone: a) sans compétition pour le sol; b) activité commerciale «prédatrice»; c) compétition pour le sol	247
8.1	Séquence de modélisation	307

LISTE DES CARTES GEOGRAPHIQUES

7.1	Montréal métropolitain - Numéro des zones de la région d'étude	298
7.2	Montréal métropolitain - Réseau routier principal	301
7.3	Montréal métropolitain - Réseau de transport collectif	303
8.1	Population 1981 - Part relative (%)	324
8.2	Emploi total 1981 - Part relative (%)	325
8.3	Population 1981-2001 - Scénario de référence - Taux annuel moyen	332
8.4	Services 1981-2001 - Scénario de référence - Taux annuel moyen	333
8.5	Population 2001 - Scénario A - Écart (%) du scénario C	334
8.6	Services 2001 - Scénario A - Écart (%) du scénario C	335
8.7	Population 2001 - Scénario B - Écart (%) du scénario C	337
8.8	Services 2001 - Scénario B - Écart (%) du scénario C	338
8.9	Origines 1981 - Transport en commun - Répartition modale (%)	341
8.10	Destinations 1981 - Transport en commun - Répartition modale (%)	342
8.11	Transport en commun 1981-2001 - Scénario A - Transfert modal (%)	345
8.12	Transport en commun 1981-2001 - Scénario B - Transfert modal (%)	348
8.13	Transport en commun 1981-2001 - Scénario C - Transfert modal (%)	350
8.14	Indice d'accessibilité à l'emploi 1981 (selon coût composé)	361
8.15	Variation d'accessibilité 1981-2001 - Scénario A (selon l'emploi 1981)	362
8.16	Variation d'accessibilité 1981-2001 - Scénario B (selon l'emploi 1981)	365
8.17	Variation d'accessibilité 1981-2001 - Scénario C (selon l'emploi 1981)	366

Sommaire

Il paraît difficile d'interpréter l'organisation spatiale d'une grande métropole sans y référer à son système de transport tout comme l'analyse des déplacements des personnes et la planification de l'offre de transport ne peuvent se dissocier de la compréhension de la structure urbaine. Pour l'économiste, le transport reste une demande dérivée de l'activité économique. Pour le géographe, les interactions humaines sont davantage qu'un produit mais un facteur organisant ces activités dans la ville.

On est alors tenté par un cadre de modélisation pouvant concilier ces deux réalités si souvent analysées isolément. Des obstacles fondamentaux se dressent déjà. La modélisation du système de transport doit avant tout être intégrée au niveau de ses propres composantes de la demande et de l'offre. Quant à la structure urbaine, il faut chercher une représentation capable d'appréhender la répartition et l'évolution de plusieurs activités interdépendantes. L'unification demande aussi la définition de principes communs ou, tout au plus, compatibles, des formulations uniformisées et un accommodement des échelles temporelles.

Ce projet d'intégration n'oblige pas à «retirer le monde» mais plutôt à mettre à contribution tous les cadres théoriques, approches et modèles concernés par la structure et le transport urbains. L'exercice de synthèse mène à la formulation d'un modèle spatio-dynamique d'une région urbaine.

Une application à la région métropolitaine de Montréal met à jour les possibilités analytiques, les ajustements nécessaires et les limites d'une telle modélisation.

Abstract

it appears difficult to explain the spatial organization of a major urban centre without referring to its transportation system just as travel demand analysis and transport planning cannot be dissociated from the understanding of urban structure. Moreover, in economics, transportation is seen as a demand resulting from economic activities while in geography, human interactions are considered as an organizing factor rather than a product of such activities.

Consequently we are interested by a modelling framework which can conciliate both realities often analysed separately. Several fundamental hindrances are therefore encountered. Transportation system analysis must initially be integrated within its own components of demand and supply. As for the urban structure, we must seek a representation suited to apprehend the spatial distribution and the evolution of several interdependent activities. Unification also requires definitions of common or at least consistent principles, standardized formulations and an adjustment of temporal scales.

This attempt of integration does not imply starting over but rather using all theoretical frameworks, approaches and models concerned by urban structure and transportation. This comprehensive exercise leads to a formulation of a space-time model of a urban region.

An empirical application to the Montreal metropolitan area unveils the analytical possibilities, the necessary adjustment and the limitations of this modelling.

Remerciements

Au programme de bourses de doctorat du Conseil de recherches en sciences humaines du Canada, au programme conjoint de bourses du Fonds F.C.A.R. et du Ministère des transports du Québec ainsi qu'à l'École des études supérieures et de la recherche de l'Université d'Ottawa pour leur soutien financier sans quoi cette recherche n'aurait pu être réalisée.

Au département de géographie de l'Université d'Ottawa qui m'a offert un excellent encadrement.

A Leon Floeckaerts, professeur titulaire et directeur de cette thèse, qui m'a constamment appuyé et aidé.

A André Landiois, professeur agrégé, avec qui des idées fructueuses ont été échangées.

A Louise, ma tendre épouse, qui n'a jamais cessé de croire à ce projet et dont les encouragements répétés ont permis de le mener à terme.

A Gabriel, mon fils, qui a grandi avec ce travail et qui a su égayer les moments difficiles de la rédaction.

Enfin, à tous les parents et amis qui ont su démontrer leur sollicitude.

Chapitre 1 INTRODUCTION

1.1 CARACTERISTIQUES DE LA THESE

Le mot "these" s'accompagne de deux definitions distinctes mais compatibles. L'une d'elles concerne l'ouvrage compose en vue de l'obtention d'un grade tandis que l'autre se rattache à une proposition que l'on peut énoncer et soutenir. La these de doctorat presentee ici s'accorde avec ces deux definitions.

Nous avons voulu aussi lui donner un sens qui depasse ceux-ci, celui d'une «synthese» i.e. un exercice de conciliation. Certes, la these, au sens academique du terme, est souvent vue comme un ouvrage qui propose du nouveau, des decouvertes, des remises en question, des solutions, des regards neufs sur de vieux problemes. Pourtant, ce nouveau ne s'appuie que sur une idee originale alors que la grande partie d'une these est consacree à rassembler des elements theoriques, methodologiques et empiriques deja connus afin de construire un cadre duquel jaillira cette idee neuve.

Dans l'etude de processus complexes, la synthese semble la cle de la comprehension (en anglais, synthese peut d'ailleurs se traduire par "comprehensive"). Elle ne consiste pas seulement de reunir des composantes apparentees d'un phenomene etudie. Elle cherche aussi à les concilier voire les integrer. D'ailleurs toutes ces composantes sont ni plus ni moins des compartiments imagines par l'homme dans un but de simplification. Pour comprendre davantage, il n'appartient qu'à l'homme d'imaginer des compartiments encore plus grands et englobants, peut-être moins detailles.

Ce discours semble emprunte à celui de l'analyse systemique dans lequel on insiste sur l'importance des proprietes de l'ensemble plutot que les caracteristiques de ses elements. Dans ce travail, nous ne reprenons pas le langage de l'analyse des systemes mais retenons cependant cette preoccupation de la conciliation et de l'integration.

1.2 LES PREMICES D'UNE INTEGRATION

1.2.1 Problématique générale

Le titre annonce déjà plusieurs tentatives d'intégration. D'abord, l'adjectif composé "spatio-dynamique" témoigne d'une conciliation des deux grandes dimensions, l'espace et le temps. L'une d'elles ne peut échapper au champ d'intérêt du géographe: les lieux sont la base d'observation privilégiée de la discipline. Une analyse purement spatiale ne peut toutefois que s'intéresser à des phénomènes strictement statiques. Plus souvent qu'autrement, la staticité paraît une présomption plutôt qu'une réalité. L'axe du temps permet quant à lui d'exprimer le mouvement et l'évolution qui entourent la plupart des phénomènes étudiés en géographie.

Ensuite, le titre évoque la structure urbaine et le système de transport. Ici, la structure urbaine adopte le sens de l'organisation spatiale des fonctions de la ville. Ces fonctions peuvent être représentées par des utilisations du sol, des bâtiments ou des activités humaines. Quant au système de transport, il concerne d'une part les infrastructures, les équipements et les services permettant les déplacements à l'intérieur d'une région urbaine. D'autre part, il comprend aussi l'ampleur et la distribution des déplacements.

Il va sans dire que la structure urbaine, l'offre et la demande de transport sont des découpages d'une même réalité. La structure interne d'une ville est un arrangement résultant d'interactions spatiales entre activités. Ces activités génèrent des déplacements dont la configuration influence le déploiement des réseaux de transport ainsi que la qualité du service. Ces relations rétroactives restent bien connues malgré qu'elles soient difficiles à décrire et davantage à modéliser.

Dans le domaine de la planification des transports urbains, on eut souvent du mal à évaluer les impacts sur le développement urbain. Jusqu'à tout récemment, les infrastructures de transport étaient déployées en fonction de strictes facteurs de circulation. On ignorait leurs effets sur la demande d'abord et par la suite sur l'organisation de la ville. Cette négligence avait ironiquement pour effet d'annuler tous les améliorations

escomptées des conditions de transport. Aujourd'hui, on reconnaît à ce type de planification une bonne responsabilité à l'égard de la suburbanisation incontrôlée des métropoles et de la motorisation accrue des déplacements urbains.

Evidemment, on accusera les planificateurs de transport et leur préoccupation trop orientée sur la voirie. Quant à nous, nous expliquons cette négligence par un cadre théorique déficient et à l'absence d'une méthodologie adéquate capable d'appréhender "synthétiquement" un système de transport et ses relations avec la structure spatio-économique qui le supporte et le justifie.

Notre titre fait aussi mention du terme "modélisation". Cela signifie que la synthèse envisagée adoptera le langage des modèles mathématiques. D'ailleurs, il existe des modèles appropriés à chacune des facettes de la structure urbaine et du système de transport. Notre vue synthétique semble toutefois se heurter au cloisonnement et à la spécialisation de ces formulations. Pourtant, beaucoup de modèles partagent des formes fonctionnelles apparentes, parfois des principes analogues.

Avec le temps et avec le souci de synthèse des chercheurs, les modèles vinrent à se ressembler voire à s'unifier. Issus initialement de principes théoriques plus ou moins éparses et rigoureux, les modèles urbains et les modèles de transport se constituèrent des cadres davantage intégrés. Nous avons identifié ici quelques uns de ces cadres qui permettent une grande conciliation.

1.2.2 Le contexte de recherche

D'abord, le concept d'«entropie», né avec la seconde loi de la thermodynamique mais repris dans un contexte statistique, fit à certains égards une révolution de la modélisation en tant que mesure d'information d'une distribution. Dépourvu de présomption comportementale, le principe de maximisation de l'entropie tel que présenté par Wilson [1967] put donner une justification statistique commune aux modèles gravitaires, de répartition modale, de distribution des flux et de localisation. Il apporta également une certaine cohérence au sein des composantes du modèle urbain

de Lowry [1964] dont on vantait la perspicacité tout en reprochant sa création "ad hoc". Dans cette préoccupation unificatrice, le concept d'entropie a l'avantage d'être utilisable dans un programme non-linéaire convexe.

La «théorie de l'utilité aléatoire», théorie économétrique du comportement individuel, devint un complément aux approches d'entropie maximale. Il s'agit de déterminer les probabilités de choix d'un consommateur ou d'un voyageur en regard des attributs propres à chacune des alternatives et des perceptions imparfaites du sujet que l'on représente par une distribution de probabilité. Cochrane [1967] démontra que le modèle de gravité peut être dérivé selon une telle approche tandis que McFadden [1973; 1978], Domencich et McFadden [1975], et Williams [1977] développèrent des cadres appropriés pour engendrer des modèles de distribution de flux, de partage modal et de localisation. L'intérêt de cette approche pour nous est que les modèles qui en émanent sont isomorphes à ceux issus de l'entropie. De plus, elle fournit une certaine explication comportementale aux modèles qui est absente dans l'approche entropique.

Dans la modélisation du transport, on doit à Beckmann, McGuire et Winsten [1956] la contribution la plus novatrice. Ils reprirent une idée mise de l'avant par Wardrop [1952] et montrèrent que les flux d'équilibre et des temps de déplacement variant selon la congestion pouvaient être calculés en solutionnant un problème de programmation non-linéaire. Ceci donna le coup d'envoi à d'innombrables développements des modèles d'équilibre de réseau.

Peut-être involontairement, le problème de Beckmann et al. devint un premier exemple d'intégration. Leur programme comprenait un terme supplémentaire servant à répartir les déplacements selon la distance séparant les origines des destinations. Bruynoogne [1969] et Tomlin [1971] spécifierent une fonction de distance avec laquelle dérivait-on un modèle de gravité de distribution des flux. L'incorporation de l'entropie [Florian, Nguyen, Ferland, 1975; Evans, 1976] clarifia la formulation et fit incidemment le rapprochement entre les idées de Wilson et de Beckmann et al. . Il n'en fallait pas plus pour faciliter l'intégration de la répartition modale [Florian, Nguyen, 1978].

Jusqu'ici, nous n'avons évoqué que les conciliations possibles des modèles descriptifs de la demande de transport. Qu'en est-il des modèles d'offre de transport? Il existe différentes formulations de modèle d'offre selon la nature et le niveau d'agrégation de l'application. Parmi celles-ci, se distinguent les «modèles de conception de réseau», terme générique désignant un ensemble de formulations et d'algorithmes destiné à résoudre des problèmes de formation de réseau. À première vue, les problèmes classiques de réseau semblent disparates (problème du plus court chemin, du voyageur de commerce, du réseau minimal, etc...) mais ils peuvent être vus comme des cas spéciaux d'un modèle général de conception de réseau [Magnanti, Wong, 1984].

Le problème général consiste à optimiser les améliorations d'un réseau en rapport avec des critères de bénéfice et de coût. Il s'agit d'un problème de programmation entière qui peut être intégré avec un problème d'affectation d'équilibre [LeBlanc, 1975]. Du fait, il devient compatible avec un problème de distribution de la demande: à cet égard, le concept d'entropie permet l'intégration complète [Boyce, Janson, 1980].

Il existe une forte parenté entre les modèles classiques de localisation [Lowry, 1964; Lakshmanan, Hansen, 1965] et certains modèles de transport. Entre autre, ils partagent leur structure gravitaire. Pourtant, un grand nombre de modèles de localisation ne réfèrent pas, du moins explicitement, à ce mécanisme d'interaction basé sur la distance des lieux. Pensons aux modèles d'allocation optimale de Koopmans et Beckmann [1957] et de localisation résidentielle d'Herbert et Stevens [1960].

Il faut trouver ailleurs les liens qui unissent les modèles de transport et de localisation. L'une des possibilités de conciliation se situe dans leur formulation en problème d'optimisation. Avec la maximisation de l'entropie, les modèles de type gravitaire se réécrivent en problème non-linéaire [Wilson, 1974]. Les modèles de Koopmans-Beckmann et d'Herbert-Stevens sont déjà des programmes mathématiques. Le programme linéaire d'allocation TOPAZ [Brotchie, 1969] assisté d'un modèle d'interaction spatiale est équivalent à un programme non-linéaire doté d'un terme d'entropie [Coelho, Wilson; 1977]. En fait, à travers leur formulation en problème d'optimisation, les modèles de localisation appartiennent à une même

famille que ceux de transport. Par surcroît, les principes d'optimisation se ressemblent: minimisation des coûts ou maximisation des bénéfices soit pour l'utilisateur, soit pour le système.

Les problèmes d'optimisation telles qu'on les utilise sont généralement statiques. Leurs solutions fournissent des distributions selon des conditions et contraintes d'une période particulière. Les applications prétendues dynamiques restent en fait des applications «statiques-comparatives» où l'on transpose les conditions d'une unité de temps à un autre. Une véritable modélisation dynamique autorise des délais de réaction, une permanence des effets, une mémorisation des comportements antérieures.

Seules les modèles de «l'analyse des séries temporelles» admettent de telles propriétés parce qu'ils renferment une structure de décalage. Malheureusement, ces modèles sont typiquement aspatiaux au sens qu'ils ne traitent pas des interactions entre des lieux. Par ailleurs, le décalage de temps peut être assimilé par analogie à la contiguïté ou à la proximité géographique; on est alors en présence de modèles d'autocorrélation spatiale [Cliff, Ord; 1970; 1973; 1980].

Du point de vue théorique, rien n'empêche d'unir au sein d'une même formulation le temps et l'espace. Les autoregressions temporelle et spatiale peuvent se confondre avec d'autres types de relations linéaires [Hordijk, Nijkamp, 1977; 1978]. La grande distinction réside dans la spécification de la structure spatiale qui, au contraire de celle du temps, est multilatérale, bidirectionnelle et multiforme. Bennett [1979] presenta une représentation générale de «système spatio-temporel» et, du même coup, systematisa les formulations. Son approche nous apparut très intégratrice: les systèmes spatio-temporels admettent des relations régressives et autoregressives, des relations simultanées comme décalées, des interactions spatiales de toutes sortes. Par surcroît, ils accueillent les matrices de structures spatiales telles que proposées par Cliff et Ord tout comme ils ont leurs équivalents en équations de différence, fonctions de transfert et fonctions d'impulsion-réponse [Box, Cox, 1970].

En fait, le cadre de modélisation de Bennett possède toutes les caractéristiques pour représenter des processus de développement urbain. Habituellement, ces processus sont décrits d'une manière statique-comparative avec des modèles de localisation. Il reste à concilier les deux approches.

Le recours à des formulations spatio-temporelles de modèles urbains nous invite à explorer leurs propriétés dynamiques. Les textes de Poston et Wilson [1977], d'Harris et Wilson [1978], de Wilson [1981] d'Harris et al. [1982] et de Clarke [1985] nous apprennent que la version "dynamisée" du modèle de localisation commerciale fait surgir des «catastrophes» i.e. des évolutions discontinues. Dans la réalité urbaine, ces catastrophes existent bel et bien: elles font émerger ou disparaître des organisations spatiales. Les modèles d'évolution intra-urbaine peuvent exhiber d'autres types de «bifurcations» [Dendrinos, Mullaly, 1985; Hazewinkel et al. 1986]. Celles-ci réfèrent à la stabilité et à configuration de l'évolution conjointe de plusieurs activités urbaines.

Le rapprochement entre ces analyses dynamiques, les systèmes spatio-temporels et les problèmes d'optimisation reste envisageable. Dans un cas, on privilégie les systèmes d'équations différentielles tandis que dans l'autre, on recourt à des modèles de décalage: les deux approches apparaissent conciliables. La solution d'un problème mathématique peut être considérée comme une distribution d'équilibre dans un environnement dynamique. De même, l'équation dynamique (dérivée du temps) peut s'inspirer de la formulation de l'équation solution (dérivée de la fonction objective en respect de la variable à localiser). Enfin, la spécification des interactions spatiales d'un système spatio-temporel peut être fournie par un modèle d'interaction, lui-même dérivé d'un programme mathématique équivalent.

Cette proximité des modèles nous a incité à faire un effort d'intégration permettant de formuler un modèle général du système de transport et de la structure urbaine. La tâche n'apparaît pas si simple. Elle consiste d'abord à uniformiser les notations, à redéfinir les principes de chacun dans une forme assimilable et à chercher une formulation englobante. Par ailleurs, le modèle général devra être opérationnel, c'est-à-dire qu'il soit applicable à une région urbaine.

1.3 LES MODELES URBAINS INTEGRES

Nous ne sommes sûrement pas les premiers à s'intéresser à la formulation d'un modèle urbain synthétisant l'organisation spatiale et le transport dans une ville. L'ancêtre de ces modèles reste sans aucun doute le modèle de Lowry [1964], un système d'équations non-linéaires et d'identités comptables avec lequel on alloue les activités résidentielles et de services dans l'espace urbain. Le mécanisme d'allocation se base sur des relations d'accessibilité de type gravitaire et c'est la seule référence au transport. Par une série d'itérations, on parvient à une distribution d'équilibre entre les activités. Cette distribution peut être plus facilement déterminée par des équations linéaires [Garin, 1966] et peut être dérivée par maximisation de l'entropie [Wilson, 1971b].

Bien que tous les modèles de localisation qui suivirent étaient fondés sur l'accessibilité, il faut attendre Putnam [1975] et Berechman [1976] pour connaître les premières véritables tentatives de combinaison des modèles d'utilisation du sol et de transport. Ces auteurs privilégièrent l'usage itératif de modèles de localisation de type Lowry et de méthodes de trafic d'équilibre. Il ne s'agit pas d'une intégration mais plutôt d'un "interface" de modèles. Une telle approche ne garantit pas la convergence vers un équilibre valide de même qu'elle peut conduire à des solutions différentes selon la solution initiale.

Los [1979] et Anas [1984] présentèrent des modèles pleinement intégrés sous la forme d'un programme mathématique d'optimisation, l'un à partir du modèle résidentiel d'Herbert et Stevens [1960], l'autre selon un schéma de type Lowry. L'existence et l'unicité d'une solution fut démontrée et les méthodes de calcul nécessaires à son obtention furent identifiées.

Les solutions issues de tels problèmes d'optimisation fournissent néanmoins des distributions urbaines d'équilibre. Dans une application de demande de transport, l'état d'équilibre reste un état envisageable à court terme puisque les voyageurs réagissent rapidement à des modifications des conditions de déplacement. Quand on parle d'activités économiques ou d'utilisation du sol, l'équilibre spatial ne peut être atteint qu'à long terme et ce, en absence de perturbation. Un tel équilibre n'est donc

vraisemblable qu'en théorie.

Dans le cadre de cette thèse, nous voulons proposer une modélisation intégrée de la structure urbaine et du système de transport qui diffère des autres déjà proposées sur les points suivants. D'abord, bien que nous privilégions la programmation mathématique comme moyen d'intégration, nous ne considérons pas toujours la solution comme l'état réalisable du système. Dans le cas de la demande de transport où les délais de réaction sont ordinairement courts, les distributions solutions sont jugées représentatives. Par contre, au chapitre du développement urbain, l'état d'équilibre est vu comme un état ultime auquel parvient-on qu'après un certain temps. L'introduction de décalages permet de reproduire cette évolution progressive vers l'état d'équilibre; en pratique, il paraît peu probable qu'un tel état soit atteint, le système urbain étant trop sujet à des perturbations.

Parmi les perturbations les plus influentes, comptent les modifications des réseaux de transport qui transforment le patron d'accessibilité intra-urbaine. Bien que dans tous les modèles urbains de synthèse, les réseaux de transport soient considérés exogènes, nous avons voulu internaliser le processus décisionnel des planificateurs de transport. Dans de grands systèmes, ce processus peut être imité par un programme mathématique d'optimisation [Magnanti, Wong, 1984]. Or, la formulation d'un tel problème peut se concilier avec celle de la demande. Il devient possible, du moins théoriquement, de modéliser les améliorations probables d'un réseau en fonction de certains critères et des conditions observées et/ou appréhendées de déplacement.

1.4 PRESENTATION DES CHAPITRES

La thèse est organisée en trois parties, chacune regroupant deux à trois chapitres. Les chapitres sont structurés par sous-chapitre et section selon un système décimal.

Les deux premières parties restent essentiellement théoriques. Elles revoient la littérature jugée pertinente, identifient les modèles et les méthodes susceptibles de concerner un éventuel modèle urbain général et

tendent de dresser des principes d'intégration. La dernière partie est consacrée à une application à la région métropolitaine de Montréal qui vise à faire ressortir les possibilités, les limites et les problèmes associés à une telle modélisation.

La première partie, comprenant les chapitres 2, 3 et 4, aborde la modélisation du transport. Puisque la plupart des modèles sont formulées en problème d'optimisation et que cette méthodologie est omniprésente dans la thèse, nous avons décidé de commencer avec un court chapitre introduisant à la programmation mathématique. Le lecteur déjà initié y trouvera un rappel des éléments fondamentaux de la programmation tandis que le non-initié se familiarisera avec les concepts et la terminologie afin de lui faciliter la lecture des chapitres consécutifs. Cependant, ce chapitre reste très sommaire et que pour une véritable introduction, il est préférable de consulter un bon manuel de programmation mathématique.

Le troisième chapitre aborde la modélisation de la demande de transport selon les quatre phases conventionnelles i.e. la génération des déplacements, la distribution spatiale des flux, la répartition modale et l'affectation sur le réseaux. Pour chacune des phases, les formulations classiques sont reprises et nous tentons d'y établir des principes communs. Certains sous-chapitres se consacrent spécifiquement à l'unification de certaines phases. Finalement, nous présentons un modèle intégré de demande qui exclut toutefois la génération zonale des déplacements, phase particulière qui peut être considérée comme une relation indépendante. Au terme du chapitre, des méthodes conventionnelles d'estimation et de calibration ainsi que les algorithmes appropriés de solution sont proposés.

La quatrième chapitre aborde la modélisation de l'offre de transport. Puisque l'offre peut être traitée de différentes manières, nous avons choisi un cadre conceptuel spécifiquement adapté d'une part, à un contexte urbain et, d'autre part, aux problèmes typiques de réseaux de transport en recherche opérationnelle. Une typologie des problèmes de conception de réseau est ensuite dressée. Progressivement, nous pouvons en arriver à une formulation parfaitement consistante avec un problème d'optimisation de la demande. Il importe ensuite d'examiner les implications d'ordre algorithmiques.

Nous pouvons alors conclure à un cadre général de modélisation du transport. Le cadre reste toutefois statique du fait que les problèmes d'optimisation associés sont dépourvus de références temporelles. Cette lacune, nous tenterons de la combler dans la deuxième partie de la thèse, celle qui se consacre à la modélisation de la structure urbaine.

On commence la seconde partie par un chapitre introduisant à l'analyse des systèmes spatio-temporels tel que formalisés par Bennett [1979]. Les caractéristiques des processus purement temporels sont décrits et assimilés par la suite à des processus typiquement spatiaux. Une représentation d'un système spatio-temporel devient alors envisageable et, d'une certaine manière, applicable à certains phénomènes observables.

Le sixième chapitre est entièrement dédié aux modèles de localisation. Nous abordons tour à tour les formulations classiques "intuitives", les problèmes d'optimisation normatifs et descriptifs. En fait, les trois types montrent beaucoup d'affinités, les modèles classiques s'assimilant à des conditions d'optimalité tandis que les problèmes descriptif et normatif se voyant comme des cas spéciaux de l'un par rapport à l'autre et vice versa. Nous traitons aussi les modèles décrivant un processus d'allocation de deux activités ou plus. Leur solution mène conséquemment à un état d'équilibre spatial quoique statique de la structure urbaine.

Cette propriété de convergence vers un l'état d'équilibre stable apparaît fondamentale à l'égard de la "dynamisation" du processus. L'assimilation d'une trajectoire strictement mathématique définie par les conditions d'optimalité à une trajectoire dynamique permet de concevoir les modèles de localisation comme des systèmes spatio-temporels tels que formalisés par Bennett [1979]. Ensuite, nous faisons état de certaines caractéristiques dynamiques des modèles de localisation. L'analyse des propriétés de «catastrophe» et de «bifurcation» offre un éclairage nouveau sur les possibilités descriptives de ces modèles.

Enfin, au terme du chapitre, nous procédons à une conciliation des modèles de transport et de localisation. Des cadres conceptuel et opérationnel sont dressés.

La dernière partie concerne la mise en application du modèle général dans une authentique région urbaine, l'aire métropolitaine de Montréal. Au septième chapitre, nous devons d'abord définir et spécifier la base de données car elle influence grandement la portée de l'application. Puis, la version opérationnalisée du modèle est présentée: on y reproduit les résultats abrégés des estimations et calibrations et fait état des procédures algorithmiques utilisées.

Le chapitre suivant est entièrement consacré à l'analyse de trois simulations de l'évolution intra-urbaine de Montréal selon un horizon 2001. Chaque simulation est effectuée selon notre modélisation et réfère à un scénario distinct de développement du transport. Un scénario définit en quelque sorte une politique générale de développement du transport urbain. Pour nous, il détermine des paramètres de base du modèle intégré.

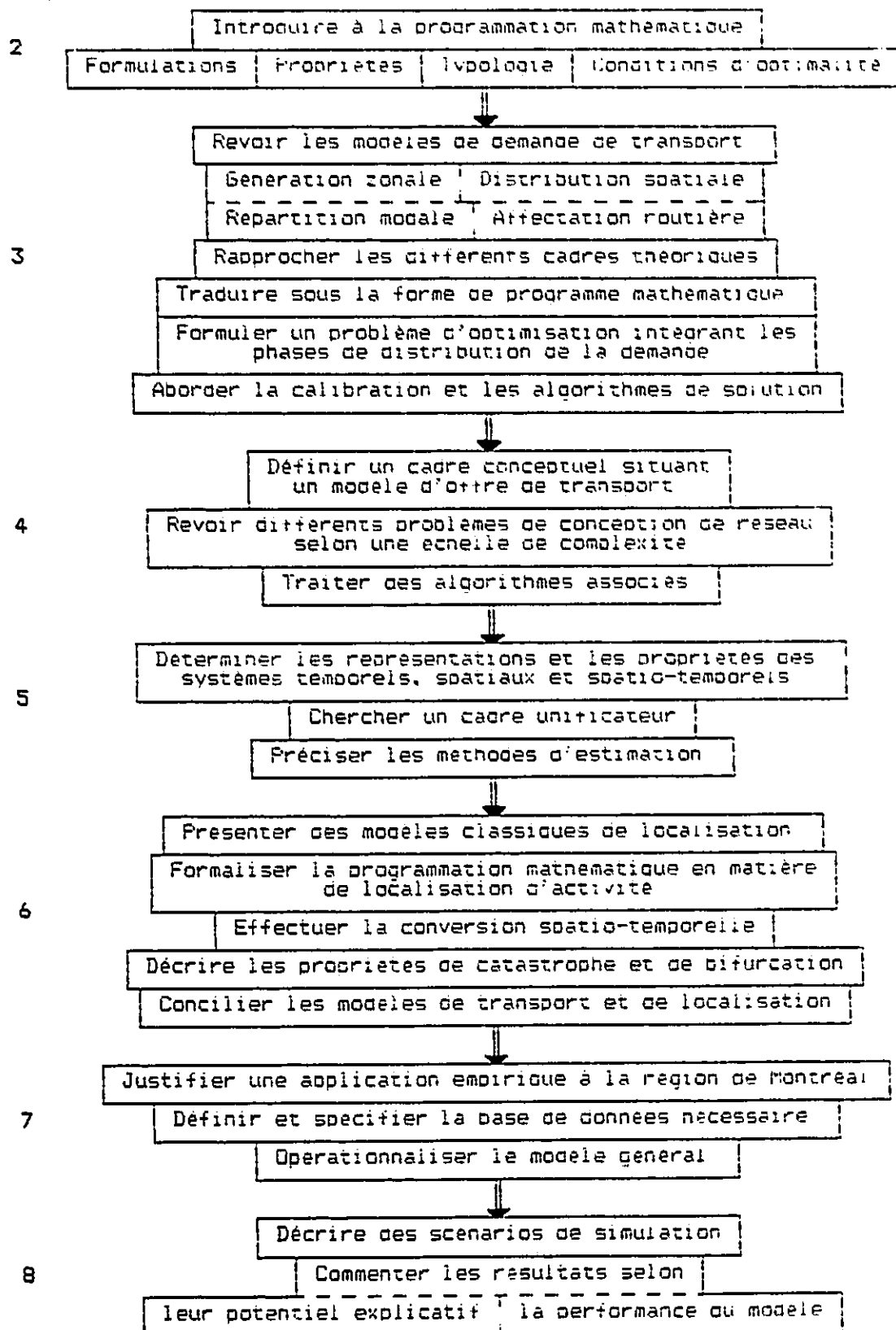
Les résultats empiriques sont commentés au sous-chapitre 8.2. Nous y examinons les développements alternatifs des réseaux et leurs conséquences sur la localisation future des activités, la répartition modale, la congestion routière ainsi que les coûts de transport des voyageurs. Il va sans dire que tous ces éléments sont extrêmement interdépendants et, par le fait même, sont mis en relief dans une modélisation intégrée qui focalise sur ces interrelations.

Nous évaluons les performances du modèle en regard de divers éléments auxquels le modèle peut paraître vulnérable. Entre autre, l'agregation qu'elle soit spatiale, temporelle ou des réseaux, influence passablement la qualité des résultats, leur fiabilité de même que les conclusions auxquels ils conduisent. Nous nous attardons aussi sur les problèmes associés aux formes fonctionnelles des sous-modèles et aux calibrations qu'elles supposent. Enfin, nous discutons des implications opérationnelles de l'intégration du système de transport et de la structure urbaine.

Dans le neuvième chapitre, nous tirons les principales conclusions de cette thèse. A la page suivante, est présenté un schéma identifiant les principaux jalons du cheminement scientifique adopté par l'auteur. Ce schéma aidera sans doute le lecteur à reconnaître la démarche et, du fait, à faciliter la compréhension du texte qui suit.

REPRESENTATION SCHEMATIQUE DE LA DEMARCHE

Chap.



Première partie:

**MODELISER LE SYSTEME
DE TRANSPORT URBAIN**

Chapitre 2

ASPECTS THEORIQUES DE LA PROGRAMMATION MATHÉMATIQUE

2.1 NECESSITE D'UNE INTRODUCTION THEORIQUE

Ce chapitre se veut essentiellement introductif. Il vise à faire un bref survol du domaine de la programmation mathématique concernée par la formulation, la description et la solution de problèmes d'optimisation. Il s'agit dans un premier temps de préciser les propriétés générales d'un programme. On identifiera les principales composantes d'un programme et décrira l'espace mathématique dans lequel évoluent les variables du programme. Par la suite, on abordera la caractérisation des optima, maxima et minima, les propriétés fondamentales de convexité des ensembles et de concavité / convexité des fonctions et, enfin, une typologie sommaire des problèmes d'optimisation usuels. Dans un deuxième temps, on s'intéressera aux théories relatives à la détermination des conditions d'optimalité. Celles-ci régissent en fait les procédures d'obtention de la solution.

La mise en place d'un chapitre d'introduction à la programmation mathématique dans une thèse de doctorat en géographie peut surprendre le lecteur. Elle se justifie, d'une part, par le fait que la formulation de problèmes d'optimisation apparaît un thème central dans cette recherche; l'intégration et la conciliation des modèles de transport et de localisation repose surtout sur ce cadre à la fois théorique et méthodologique. D'autre part, la programmation mathématique reste malheureusement peu courante en géographie. En effet, la géographie dite «quantitative» semble dominée par le paradigme de la statistique; les analyses multivariées linéaires (de régression, discriminantes, factorielles, etc...) occupent une place prépondérante dans le paysage quantitatif de cette discipline. Pourtant, les méthodes d'optimisation peuvent se concilier à ces analyses et sont même munies d'un cadre théorique plus englobant (et moins restrictif). De plus, ils conviennent davantage à des problèmes spatiaux.

Cette sous-utilisation pourrait s'expliquer par la méfiance qu'ont les chercheurs en sciences humaines envers toute méthode impliquant une notion de «rationalité» surtout connotant avec «optimalité». Il s'agit à tout le

moins d'une méprise car, ici, l'optimisation possède essentiellement un sens mathématique et non comportemental. De même, il ne faut pas confondre les problèmes «normatifs» avec lesquels on cherche des allocations totalement rationnelles (par exemple, planification des équipements, contrôle de production) et les problèmes «descriptifs» avec lesquels des distributions observables sont établies. Il se peut par ailleurs qu'un problème descriptif stipule un comportement rationnel mais il devient fallacieux de l'inférer à l'individu plutôt que de l'attribuer à une distribution numérique. De la même manière, les modèles statistiques prônés en géographie sont criblés de présomptions qui, dans bien des cas, se défendent difficilement.

Il paraît faux tout de même de prétendre que les géographes méconnaissent les problèmes d'optimisation. Le «problème de transport de programmation linéaire» d'Hitchcock [1941], le «problème d'allocation quadratique» de Koopmans et Beckmann [1957], le «problème du plus court chemin» de Dantzig [1960], les «méthodes de maximisation d'entropie» introduites par Wilson [1970] et le «modèle TOPAZ d'allocation optimale» de Brotchie [1969] se révèlent des exemples classiques de programmation mathématique dans un contexte spatial. Néanmoins, les adeptes en géographie se restreignent à des applications spécifiques et négligent souvent les possibilités d'unification et de description qu'offre le cadre théorique de la programmation.

La programmation mathématique s'entoure d'une littérature scientifique abondante et très développée, notamment en ce qui concerne les monographies consacrées à un type de programmation. Pour une introduction à la programmation linéaire, citons Dantzig [1963] et Taha [1976]. Quant à la programmation non-linéaire, les livres de Zangwill [1969] et Simmons [1975] couvrent adéquatement le domaine en se distinguant par leur approche unifiée; le chapitre présent réfère largement à ces deux volumes. Pour un traitement spécifique des programmes convexes et quadratiques, voir respectivement Rockafellar [1969] et Van de Pamme [1975]. Dans un contexte d'analyse de transport et localisation, l'oeuvre collective de Wilson, Coelho, Macgill et Williams [1981] expose maintes applications de l'optimisation qui se rapprochent beaucoup de celles considérées dans cette recherche.

2.2 PROPRIETES GENERALES D'UN PROBLEME D'OPTIMISATION

2.2.1 Eléments et structure d'un problème d'optimisation mathématique

Les problèmes d'optimisation concernent la «maximisation» ou la «minimisation» d'une fonction algébrique, appelée «fonction objective», d'une ou de plusieurs variables. En général, les valeurs prises par les variables sont restreintes par un ensemble d'équations et/ou d'inégalités, appelées «contraintes». Par conséquent, il s'agit de trouver les meilleures valeurs permises par les contraintes plutôt que les meilleures possibles. Dans les applications qui concernent cette recherche, les variables étant généralement «continues», on omettra ici le traitement des «programmes entiers».

On peut formuler un problème d'optimisation suffisamment général pour qu'il puisse représenter tout genre de problème. Designons d'abord un vecteur $x \mid x' = \{x_1, x_2, \dots, x_N\}$ rassemblant les variables assujetties à l'optimisation. Représentons la fonction objective par $f(x)$, les fonctions de contrainte par un vecteur-fonction $g(x) \mid [g(x)]' = \{g_1(x), g_2(x), \dots, g_M(x)\}$ et les constantes de contrainte par un vecteur $b \mid b' = \{b_1, b_2, \dots, b_M\}$. Alors un problème d'optimisation ($\text{opt} \equiv \text{min ou max}$) s'écrira:

$$\begin{array}{l} \text{opt } f(x) \\ \{x\} \end{array} \quad \text{sujet à } g_k(x) \leq, =, \geq b_k \quad k=1,2,\dots,M \quad (2.1a-2.2a)$$

ou bien

$$\begin{array}{l} \text{opt } f(x) \\ \{x\} \end{array} \quad \text{sujet à } g(x) \{ \leq, =, \geq \} b \quad (2.1b-2.2b)$$

Les fonctions $f(x)$ et $g(x)$ peuvent être linéaires et/ou non-linéaires; si aucune est non-linéaire, alors le problème réfère à la «programmation linéaire» tandis si l'une d'entre elles l'est, on traite avec la «programmation non-linéaire». Cette distinction ne se révèle fondamentale que lorsque des méthodes de solution sont sollicitées, sinon un problème linéaire n'est qu'un cas particulier d'un problème non-linéaire.

Une allocation numérique particulière du vecteur x est dite une «solution» du problème s'il désigne un point dans l'espace Euclidien à "N" dimensions E^N . Cette solution est dite «faisable» si elle satisfait sans exception toutes les contraintes du problème. L'union de toutes les solutions faisables dans E^N constitue la «région faisable». Dans un problème

d'optimisation dépourvu de contraintes, la région faisable correspond à l'espace multi-dimensionnel E^N et il n'y a pas de région infaisable.

Une «solution optimale» reste avant tout faisable et permet d'obtenir une valeur de la fonction objective $f(x)$ supérieure ou inférieure à toute autre solution faisable. Cette définition ne garantit cependant pas l'unicité et l'existence de l'optimum. Certains problèmes «à optima multiples» sont résolus par plusieurs configurations numériques de x . Par ailleurs, il est fort possible qu'aucune solution faisable existe car aucune valeur de x ne puisse satisfaire la totalité des contraintes; dans ce cas le problème est considéré «infaisable». Le problème peut aussi être «sans borne» i.e. donnant une solution optimale de valeur infinie de par ses contraintes qui définissent une région faisable illimitée ou de par le type de fonction objective.

2.2.2 Maxima et minima d'une fonction

Résoudre un problème d'optimisation signifie évidemment d'identifier une solution optimale et de s'assurer de son optimalité. Le terme "optimum" réfère comme on le sait déjà au terme "maximum" ou "minimum". Le théorème des valeurs extrêmes stipule que si une fonction $f(x)$ est continue dans un intervalle $[a,b]$, alors f a une valeur maximale absolue et une valeur minimale absolue dans $[a,b]$. Les extrema absolus sont nécessairement parmi les extrema relatifs lesquels ne peuvent apparaître qu'aux points frontières "a" et "b" et aux points critiques "c" tel que $df/dx = 0$ ($x=c$). En programmation mathématique, la définition des optima s'assimile à celle des extrema bien qu'usuellement il faut considérer la multi-dimensionalité de l'espace. Certains concepts émanant de la théorie des ensembles permettent de comprendre davantage la définition des optima.

Définissons d'abord proprement la région faisable. Soient deux points $x \mid x' = \{x_1, \dots, x_N\}$, $y \mid y' = \{y_1, \dots, y_N\}$ dans un espace Euclidien E^N . La distance $|x-y|$ est définie par:

$$\left[\sum_{i=1}^N (x_i - y_i)^2 \right]^{1/2}$$

Un point x est un «point intérieur» de l'ensemble P dans E^N si $x \in P$ et s'il existe un nombre positif ξ tel que tous les points dans E^N à une

distance ξ sont aussi contenus dans P . Un point y est un «point frontière» d'un ensemble P dans E^N si pour tout $\xi > 0$, il existe à moins d'une distance ξ de y au moins un point dans P et au moins un point hors de P . Le point y n'a toutefois pas besoin d'être dans P .

Un «ensemble fermé» est un ensemble contenant la totalité de ses points frontières. Si $g_k(x)$ est une fonction continue, alors la collection des points x dans E^N satisfaisant $g_k(x) \leq b_k$ (ou $g_k(x) \geq b_k$, $g_k(x) = b_k$) constitue un ensemble fermé. L'intersection d'un nombre fini d'ensembles fermés est également un ensemble fermé. Par conséquent, tout programme:

$$\begin{array}{ll} \text{opt } f(x) & \text{sujet à } g(x) \{ \leq, =, \geq \} b \\ \{x\} & \end{array} \quad (2.3-2.4)$$

garantit une région faisable qui est un ensemble fermé du fait que les symboles logiques sont restreints à $\leq$, $=$ et $\geq$.

Dans la résolution d'un problème d'optimisation, le but consiste alors à identifier l'«optimum global» (ou absolu) parmi les points faisables. Un maximum global (ou minimum global) de la fonction $f(x)$ dans un ensemble fermé P se trouve au point x^* dans P si et seulement si $f(x) \leq f(x^*)$ (ou $f(x) \geq f(x^*)$) pour tout point x dans P . Cette définition convient autant à une «solution intérieure» qu'à une «solution frontière». Il faut distinguer en surcroît les «optima locaux» (ou relatifs) qui surviennent de deux manières. D'une part, un maximum local (ou minimum local) de $f(x)$ se rencontre à x^* dans P si à une distance $\delta > 0$ de x^* , tout point x dans P est tel que $f(x) \leq f(x^*)$ (ou $f(x) \geq f(x^*)$). D'autre part, généralement, tout point frontière est au moins un optimum local.

L'identification de l'optimum ou des optima repose sur celle des points critiques du problème d'optimisation. Pour l'instant, il y a deux concepts essentiels à expliciter; la «vecteur gradient» et la «matrice Hessienne». Le vecteur gradient de $f(x)$, ∇f , est un vecteur de taille N dont les éléments sont les dérivées partielles de f en respect de chaque élément de x . Ce vecteur s'avère très utile dans les procédures de recherche de solution car il donne la direction multi-dimensionnelle de croissance maximale de f :

$$\nabla f = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_N \end{bmatrix} \equiv \{\partial f / \partial x_i\} \quad N \cdot 1 \quad (2.4)$$

Par définition, le vecteur ∇f s'annule à tout point critique de $f(x)$ i.e. à tout maximum, minimum et point de selle. La matrice Hessienne est la version matricielle de la dérivée seconde; pour les points critiques d'une fonction univariée $f(x)$ tel que $\partial f / \partial x = 0$, celle-ci permet de distinguer un minimum ($\partial^2 f / \partial x^2 > 0$) d'un maximum ($\partial^2 f / \partial x^2 < 0$). La matrice Hessienne se compose donc des dérivées secondes de f en respect des croisements des éléments de x :

$$H = \begin{bmatrix} \partial^2 f / \partial (x_1)^2 & \partial^2 f / \partial x_1 \partial x_2 & \dots & \partial^2 f / \partial x_1 \partial x_N \\ \partial^2 f / \partial x_2 \partial x_1 & \partial^2 f / \partial (x_2)^2 & \dots & \partial^2 f / \partial x_2 \partial x_N \\ \vdots & \vdots & \ddots & \vdots \\ \partial^2 f / \partial x_N \partial x_1 & \partial^2 f / \partial x_N \partial x_2 & \dots & \partial^2 f / \partial (x_N)^2 \end{bmatrix} \equiv \{\partial^2 f / \partial x_i \partial x_j\} \quad N \cdot N \quad (2.5)$$

H est négatif semi-défini si le produit $x^T H x \leq 0$ et positif semi-défini si $x^T H x \geq 0$. Lorsque $f(x)$ possède de nombreux points critiques, le signe de la matrice Hessienne sert à distinguer les maxima des minima.

2.2.3 Propriétés de convexité et de concavité

Cette section fera un survol des concepts d'ensemble convexe et de fonctions concave et convexe. Un «ensemble convexe» possède la propriété qu'un segment linéaire unissant toute paire de points de l'ensemble est aussi dans l'ensemble. Plus exactement, un ensemble C dans E^N est convexe si:

$$y = \omega x_1 + (1-\omega)x_2 \in C, \quad x_1, x_2 \in C, \quad 0 \leq \omega \leq 1 \quad (2.6)$$

La fonction paramétrisée par ω constitue un segment de ligne dans l'espace E^N et est appelée une combinaison convexe des points x_1 et x_2 . L'intersection de plusieurs ensembles convexes devient un ensemble convexe [Zangwill, 1969, chap.3]. Dans un espace bi-dimensionnel, un ensemble ayant une forme circulaire, ellipsoïdale ou polygonale régulière est convexe tandis qu'un ayant une forme de "U", de "L" (ou toute autre forme dotée d'une saillie) est non-convexe car il existe des segments liant des points de l'ensemble qui ne sont pas complètement dans l'ensemble.

Une fonction $f(x)$ est dite «concave» si:

$$f[\omega x_1 + (1-\omega)x_2] \geq \omega f(x_1) + (1-\omega)f(x_2), \quad x_1, x_2 \in C, \quad 0 \leq \omega \leq 1 \quad (2.7)$$

et «convexe» si:

$$f[\omega x_1 + (1-\omega)x_2] \leq \omega f(x_1) + (1-\omega)f(x_2), \quad x_1, x_2 \in C, \quad 0 \leq \omega \leq 1 \quad (2.8)$$

Une telle caractérisation n'est significative que si $f(x)$ prévaut dans un ensemble convexe car sinon le point $y = \omega x_1 + (1-\omega)x_2$ n'est assurément pas compris dans l'ensemble. Sur un graphique représentant la fonction $f(x)$, la concavité est visible par l'ouverture de la courbe orientée vers le bas tandis que pour la convexité, l'ouverture est vers le haut. Par exemple, les fonctions $f(x) = -x^2$, $f(x) = \ln(x)$ sont concaves alors que $f(x) = x^2$, $f(x) = \exp(\alpha x)$, $f(x) = a^x$ sont convexes. Des fonctions avec plus d'une courbure comme $f(x) = ax^3$ ne sont ni convexes ni concaves à moins d'en restreindre l'intervalle de manière à isoler une partie convexe ou concave. Une propriété essentielle est que la somme de fonctions concaves est aussi concave. Alors si les $f_i(x)$ sont concaves:

$$f(x) = \sum_i a_i f_i(x) \quad \text{est aussi concave.} \quad (2.9)$$

Aussi si $f(x)$ est concave, alors $-f(x)$ est convexe. Si $g(x)$ est convexe, alors, l'ensemble

$$U \equiv \{x \mid g(x) \leq b\} \quad (2.10)$$

est un ensemble convexe tandis que si $g(x)$ est concave, l'ensemble

$$V \equiv \{x \mid g(x) \geq b\} \quad (2.11)$$

est également un ensemble convexe. Par conséquent, le programme (2.3-2.4) possède une région faisable convexe si chacune des contraintes génère un ensemble convexe du fait que l'intersection de celles-ci reste nécessairement convexe. Il faut incidemment que dans les contraintes d'égalité du type $g_k(x) = b_k$, la fonction $g_k(x)$ soit en même temps convexe et concave; cela n'est possible que si $g_k(x)$ est linéaire. Ces conditions ne sont pas strictement nécessaires car il se peut que parmi les contraintes se réalisent certaines ne formant pas un ensemble convexe bien que la région faisable demeure convexe. Cette éventualité apparaît lorsque la région faisable est telle qu'elle exclut les parties qui contreviennent aux conditions.

Dans la résolution d'un problème d'optimisation, la connaissance de ces propriétés pour la région faisable et pour la fonction objective permet d'identifier plus facilement l'optimum global. Ainsi, si $f(x)$ est défini

et convexe à travers une région faisable convexe U , alors tout minimum local de $f(x)$ dans U est aussi le minimum global dans U . De même que si $f(x)$ est défini et concave à travers une région faisable convexe V , alors tout maximum local de $f(x)$ dans V est aussi le maximum global dans V [Simmons, 1975 chap.3]. Ces circonstances semblent favoriser la solution d'un problème de minimisation, dans un cas, et de maximisation, dans l'autre. Il devient alors aisé de conclure que, dans les cas contraires, si $f(x)$ est convexe dans U , le maximum global se situera à un point frontière de U tandis que si $f(x)$ est concave dans V , le minimum global se trouvera sur la frontière de V . Notons enfin que la matrice Hessienne (2.5) reste positive (négative) semi-définie dans U (dans V) si $f(x)$ est convexe dans U (concave dans V).

2.2.4 Typologie des problèmes d'optimisation

Le survol théorique qui a été fait précédemment vaut pour l'ensemble des problèmes d'optimisation. Cependant, la solution des problèmes peut différer selon le type de fonction objective et de contraintes. Ici, on se contentera de distinguer et de nommer les types afin que dans les chapitres suivants, on puisse reconnaître les appellations conventionnelles. D'abord, identifions les genres de contraintes possibles selon la relation logique sollicitée:

$$\begin{array}{ll} g_k(x) = b_k & \text{Contrainte d'égalité} \\ g_k(x) \geq b_k & \text{Contrainte de supériorité} \\ g_k(x) \leq b_k & \text{Contrainte d'infériorité} \end{array} \quad \left. \vphantom{\begin{array}{l} \\ \\ \end{array}} \right\} \text{ inégalité}$$

Les inégalités strictes (" $>$ ", " $<$ ") sont rarement employées car elles définissent pas des ensembles fermés. D'ailleurs, il est inaccoutumé qu'une restriction exclut une valeur spécifique lorsqu'on travaille avec des nombres réels (donc continus). Les fonctions $g_k(x)$ peuvent être linéaires ou non-linéaires. Parmi les contraintes linéaires les plus communes il y a certes la contrainte de «non-négativité»:

$$x_i \geq 0 \quad i=1,2,\dots,N \quad \text{ou} \quad x \geq 0 \quad (2.12)$$

Dans ce cas, $g_k(x) = 1 \cdot x_k$ ($k=i=1,2,\dots,N$). Les contraintes de somme seront très fréquentes dans les modèles présentées ultérieurement:

$$\sum_{i=1}^N x_i = X \quad \text{ou} \quad i'x = X \quad i' = \{1,1,\dots,1\} \quad 1 \cdot N \quad (2.13)$$

Si le vecteur x est une distribution de probabilité, alors $\sum x_i = 1$. Dans le cas où la variable du problème est une matrice $X = \{x_{ij}\}$ $N \times N$, les contraintes de «totaux marginaux» demeurent très utilisées:

$$\sum_{j=1}^N x_{ij} = X_i^T \quad \text{et} \quad \sum_{i=1}^N x_{ij} = X_j^F \quad (2.14)$$

Dans un problème comportant deux niveaux de variables endogènes, on rencontre souvent des contraintes de «couplage» du type:

$$\sum_{i=1}^N x_{mi} - y_m = 0 \quad m=1,2,\dots,N_m \quad (2.15)$$

Quant aux contraintes non-linéaires, il n'y a pas lieu ici de présenter les sortes les plus communes car elles sont ordinairement très particulières.

La typologie des programmes repose sur les genres de fonction objective et de contraintes. On trouve dans Kuhn [1981] une schématisation des types de programmation mathématique et leurs relations du point de vue de leurs structures. Un «programme linéaire» possède une fonction à optimiser strictement linéaire et des contraintes linéaires:

$$\min_{\{x\}} c'x \quad \text{sujet à} \quad Ax \geq b \quad x \geq 0 \quad (2.16-2.17)$$

où $c = \{c_i\}$ $N \times 1$, $x = \{x_i\}$ $N \times 1$, $A = \{a_{ki}\}$ $M \times N$, $b = \{b_k\}$ $M \times 1$

A titre d'exemple, ce programme est un problème d'allocation de quantités x_i qui vise à minimiser les coûts spécifiques c_i et à satisfaire les M requisitions b_k . Dans la famille des programmes non-linéaires, il y a le fameux «programme quadratique» qui consiste à optimiser une matrice de forme quadratique i.e. représentant un polynôme du second degré du vecteur endogène. Les contraintes sont linéaires:

$$\max_{\{x\}} \frac{1}{2}x'Qx + c'x \quad \text{sujet à} \quad Ax \leq b \quad x \geq 0 \quad (2.18-2.19)$$

où $Q = \{q_{ij}\}$ $N \times N$ où $Q = Q'$ (symétrique)

Ici, la forme quadratique permet d'introduire un effet d'interaction parmi les variables x_i . Parmi les problèmes non-linéaires, il s'agit généralement du plus facile à résoudre car il est légèrement plus complexe qu'un problème linéaire. On parle souvent aussi de «programmation convexe». Un problème convexe contient une fonction objective convexe dans le cas d'une minimisation et concave dans le cas d'une maximisation (un problème quadratique reste un cas particulier de problème convexe). Un problème con-

vexe peut se présenter ainsi:

$$\max_{\{x\}} f(x) \quad \text{sujet à } g(x) \leq b \quad x \geq 0 \quad (2.20-2.21)$$

où $f(x)$ est concave et $g(x)$ est convexe

Comme il a été mentionné dans 2.2.3, tout maximum local de $f(x)$ dans la région faisable est alors systématiquement le maximum global. On abordera souvent des problèmes où la fonction objective est une sommation de fonctions concaves identiques de chaque variable du type:

$$f(x) = \sum_i f(x_i) \quad (2.22)$$

Un type particulier de problème non-linéaire est le «programme géométrique» caractérisé par sa fonction objective «posynômiale». Des dérivations de modèles d'interaction spatiale ont déjà été faites à partir de problèmes géométriques [voir Nijkamp, 1972] mais ce genre de programmation ne sera pas exploré ici. Enfin, le problème non-linéaire généralisé apparaît sous la forme suivante:

$$\max_{\{x\}} f(x) \quad \text{sujet à } g(x) \leq b \quad x \geq 0 \quad (2.23-2.24)$$

Seule la propriété de continuité, non plus de concavité, est exigée pour f . Contraintes linéaires et non-linéaires peuvent se confondre. Tout comme pour le programme convexe (2.20-2.21), la contrainte de non-négativité n'est pas obligatoire bien que courante. Est aussi un problème non-linéaire, celui dont $f(x) = c'x$ (linéaire) mais pourvu d'au moins une contrainte non-linéaire. Dans tous les cas présentés, la maximisation peut être substituée par une minimisation (car $\max(f) = \min(-f)$) et que les contraintes de supériorité sont interchangeable par celles d'infériorité et vice et versa: $\{g_k(x) \geq b_k\} \equiv \{-g_k(x) \leq -b_k\}$. Les contraintes d'égalité peuvent être représentées par: $\{g_k(x) = b_k\} \equiv \{g_k(x) \geq b_k \text{ \& } g_k(x) \leq b_k\}$. De cette façon, les programmes génériques décrits de (2.16) à (2.24) conviennent à la plupart des formulations.

2.3 DETERMINATION DES CONDITIONS D'OPTIMALITE

Ce sous-chapitre traitera des méthodes habituellement utilisées en programmation mathématique dans l'identification des points optimaux. D'une part, il sera question des «conditions nécessaires» i.e. conditions garantissant qu'un point est un optimum local. La nécessité s'exprime par le caractère obligatoire que revêtent certains attributs de ces points critiques que sont les optima relatifs. D'autre part, nous aborderons les «conditions suffisantes», conditions avec lesquelles, soit on distingue les maxima locaux des minima locaux, soit on s'assure de l'optimalité globale d'un point. Nous étudierons séparément les problèmes d'optimisation sans contrainte, ceux avec contraintes d'égalité et, enfin, ceux avec contraintes d'inégalité. Cette division se justifie par les aspects théoriques fondamentaux qui diffèrent dans chacune des catégories de problème.

2.3.1 Optimisation sans contrainte

Un problème d'optimisation sans contrainte est celui qui cherche le maximum ou le minimum d'une fonction de la distribution $\{x_i\}$ sans imposer de restriction sur les valeurs que peuvent prendre les x_i . La région faisable correspond alors au domaine des nombres réels. Voici un problème générique:

$$\max_{\{x\}} f(x) \quad x \in \mathbb{R}^n \quad (2.25)$$

Parmi les types de programme mathématique qui ont été cités précédemment dans la section 2.2.4, les problèmes sans contrainte apparaissent les plus simples à résoudre. En général, il suffit de différencier la fonction objective en respect des variables et de préciser les valeurs où les dérivées partielles s'annulent. Si la fonction n'est pas strictement linéaire, les zéros des dérivées partielles coïncident aux optima locaux dont au moins un à l'optimum global. Par exemple, une solution optimale du problème (2.25) est obtenue si les conditions suivantes se vérifient:

$$\frac{\partial f(x)}{\partial x_i} = 0 \quad i=1,2,\dots,N \quad (2.26a)$$

ou selon une notation matricielle:

$$\nabla f = 0, \quad \nabla f' = \left[\frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \quad \dots \quad \frac{\partial f(x)}{\partial x_N} \right] \quad (2.26b)$$

Les équations ci-dessus représentent les conditions nécessaires pour qu'un point soit optimal. Elles s'avèrent insuffisantes car elles se vérifient pour tout point critique. Les conditions suffisantes d'un optimum local sont développées à partir de l'expansion de Taylor au voisinage du point optimal considéré [voir Simmons, 1975 chap.4]. Parce que le terme de premier ordre (contenant le vecteur gradient ∇f) est nul et que les termes d'ordre 3 et plus paraissent négligeables, seul le second terme contenant la matrice Hessienne H de dérivées secondes permet de définir les conditions suffisantes d'un optimum local. Lorsque x^* est un vecteur de valeurs critiques tel que

$\nabla f(x^*) = 0$, il est un maximum si:

$$H(x^*) \text{ est négatif semi-défini, soit } x'Hx \leq 0 \quad (2.27a)$$

et un minimum si:

$$H(x^*) \text{ est positif semi-défini, soit } x'Hx \geq 0 \quad (2.27b)$$

où $H = \{\partial^2 f(x)/\partial x_i \partial x_j\}$ $N \cdot N$

Quant à la détermination du maximum (minimum) global, il suffit de sélectionner le vecteur x^* qui donne le $f(x^*)$ le plus élevé (faible) parmi ceux qui satisfont les conditions de premier et de deuxième ordre.

Dans la pratique, il y a peu d'application où l'on ne contraint pas la distribution recherchée, ne serait-ce qu'une condition de non-négativité. Le cas le plus classique de méthode d'optimisation dépourvue de contrainte reste sans aucun doute la «méthode des moindres carrés» affiliée à l'analyse de régression linéaire (ou non-linéaire). Par exemple, le problème de régression simple consiste à minimiser les écarts au carré entre une série observée $\{y_i\}$ et une série estimée $\{\hat{y}_i\}$ décrite par une fonction linéaire d'une autre série $\{x_i\}$. La minimisation ne concerne que les paramètres de la fonction linéaire unissant y_i et x_i . Il s'agit d'un problème sans contrainte car aucune restriction existe à l'égard des deux paramètres. Voici le problème de minimisation en question:

$$\min_{\{\alpha, \beta\}} f = \sum_1 (y_i - \hat{y}_i)^2 \quad \text{où } \hat{y}_i = \alpha + \beta x_i \quad (2.28a)$$

$$\text{ou } \min_{\{\alpha, \beta\}} f = \sum_1 \left[(y_i)^2 - 2\alpha y_i - 2\beta x_i y_i + \alpha^2 + 2\alpha\beta x_i + \beta^2 (x_i)^2 \right] \quad (2.28b)$$

Les conditions d'optimalité pour α et pour β sont:

$$\frac{\partial f}{\partial \alpha} = -2(\sum_1 y_1) + 2N\alpha + 2\beta(\sum_1 x_1) = 0 \quad (2.29)$$

$$\frac{\partial f}{\partial \beta} = -2(\sum_1 x_1 y_1) + 2\alpha(\sum_1 x_1) + 2\beta(\sum_1 (x_1)^2) = 0 \quad (2.30)$$

La présence de α et de β dans leurs fonctions dérivées permet de définir leurs valeurs respectives:

$$\alpha = \frac{\sum_1 y_1}{N} - \beta \cdot \frac{\sum_1 x_1}{N} = \bar{y} - \beta \bar{x} \quad (2.31)$$

$$\beta = \frac{\sum_1 x_1 y_1 - N \bar{x} \bar{y}}{\sum_1 (x_1)^2 - N \bar{x}^2} \equiv \frac{\text{Cov}(x, y)}{\text{Var}(x)} \quad (2.32)$$

Etant donné que la fonction objective (2.28b) est strictement convexe, il n'y a qu'un minimum possible et les conditions nécessaires (2.29) et (2.30) sont suffisantes pour le minimum global.

2.3.2 Optimisation avec contraintes d'égalité

La présente section traitera des problèmes d'optimisation dotés de contraintes d'égalité. On aborde distinctivement les égalités et les inégalités car les conditions gouvernant la solution diffèrent qualitativement. En effet, les contraintes d'égalité, au contraire de celles d'inégalité, ne dressent pas de borne à la région faisable mais réduisent plutôt la dimensionnalité du problème en induisant des dépendances entre les variables. Ainsi, si le problème concerne un vecteur de longueur N et est muni de M contraintes, il y a en fait $N-M$ variables indépendantes. Voici une formulation d'un exemple générique:

$$\max_{\{x\}} f(x) \quad \text{sujet à } g_k(x) = b_k \quad k=1,2,\dots,M \quad (2.33-2.34)$$

où $x = \{x_1\} N \times 1 \quad N > M$

Les M fonctions $g_k(x)$ peuvent autant être linéaires que non-linéaires; pour être considérées, les fonctions ne doivent pas être redondantes, ni inconsistantes. De plus, si $M \geq N$, il n'y a aucune région faisable dans laquelle peut évoluer x_1 .

La théorie de l'optimisation avec contraintes d'égalité est fondée sur l'existence de certains nombres appelés «multiplicateurs de Lagrange», nommés en l'honneur du mathématicien français Joseph-Louis Lagrange. La méthode des multiplicateurs lagrangiens est employée pour enlever la considération explicite de la dépendance causée par les contraintes au dépens de l'introduction de variables additionnelles. Il suffit de reformuler le problème d'optimisation sous sa forme «lagrangienne», forme qui concilie la fonction objective et les contraintes. Par exemple, le problème (2.33-2.34) devient:

$$\max \mathcal{L}(x, \lambda) = f(x) - \sum_k \lambda_k (g_k(x) - b_k) \equiv f(x) - \lambda' [g(x) - b] \quad (2.35)$$

Le vecteur λ contient les M multiplicateurs lagrangiens associés à chacune des contraintes. Il faut noter, d'une part, que le deuxième terme n'affecte pas la fonction objective initiale car, à l'optimalité, il devient obligatoirement nul. D'autre part, la fonction lagrangienne permet d'absorber les contraintes au sein de la fonction à optimiser. Le rôle des multiplicateurs consiste alors à faire respecter les contraintes. Les valeurs stationnaires de x et λ sont alors obtenues, tout comme dans un problème non-contraint, en annulant leurs dérivées partielles. Soit:

$$\frac{\partial \mathcal{L}(x, \lambda)}{\partial x_i} = \frac{\partial f(x)}{\partial x_i} - \sum_k \lambda_k \frac{\partial g_k(x)}{\partial x_i} = 0 \quad i=1, 2, \dots, N \quad (2.36a)$$

$$\frac{\partial \mathcal{L}(x, \lambda)}{\partial \lambda_k} = g_k(x) - b_k = 0 \quad k=1, 2, \dots, M \quad (2.37b)$$

ou dans une notation matricielle:

$$\nabla f - \lambda \cdot \nabla g = 0 \quad g(x) - b = 0 \quad (2.36b-2.37b)$$

Ces équations désignent les conditions nécessaires pour un optimum local. Elles sont suffisantes pour un optimum global, dans le cas d'une maximisation, si $f(x)$ est concave et les $g_k(x)$ sont linéaires; pour une minimisation, il suffit de garantir la convexité de $f(x)$ au même titre que la linéarité des contraintes. Si tel n'est pas le cas, on se contentera d'établir les conditions suffisantes de 2^e ordre pour un optimum local en se rabattant sur les propriétés de la matrice Hessienne composée des dérivées secondes de la fonction lagrangienne. Cependant, cette matrice est dorénavant de dimension $(N+M) \cdot (N+M)$ car les multiplicateurs font maintenant partie du problème. Alors:

$$H = \begin{bmatrix} \nabla^2 L & \nabla g' \\ \nabla g & 0 \end{bmatrix} \quad \begin{aligned} \nabla^2 L &= \{\partial^2 L(x, \lambda) / \partial x_i \partial x_j\} \quad N \cdot N \\ \nabla g &= \{\partial g_k(x) / \partial x_j\} \quad M \cdot N \\ 0 &= \{0\} \quad M \cdot M \end{aligned} \quad (2.38)$$

Ici, imaginons, par exemple, un problème simple d'allocation de biens (ou de service) représentés par la variable x_i parmi un certain nombre de lieux $i=1, 2, \dots, N$. La distribution recherchée doit maximiser les bénéfices (variable z_i) mais on constate une certaine variabilité des allocations due à un grand nombre d'acteurs. Cette variabilité est mesurée par l'entropie présumée dans la distribution:

$$\max_{\{x\}} \sum_i x_i z_i \quad \text{sujet à} \quad -\sum_i x_i (\ln x_i - 1) = S \quad (2.39-2.40)$$

$$\sum_i x_i = X \quad (2.41)$$

Formulons la forme lagrangienne:

$$L(x, \gamma, \lambda) = \sum_i x_i z_i + \gamma \cdot \{-\sum_i x_i (\ln x_i - 1) - S\} - \lambda \cdot (\sum_i x_i - X) \quad (2.42)$$

Soit $\partial L / \partial x_i = z_i - \gamma \ln x_i - \lambda = 0$

et $\partial L / \partial \gamma$, $\partial L / \partial \lambda$ correspondant aux contraintes (2.40) et (2.41).

La distribution solution est:

$$x_i = \exp\{(z_i - \lambda) / \gamma\} \quad \text{où} \quad \lambda = -\ln X + \gamma \cdot \ln \sum_i \exp(z_i / \gamma) \quad (2.43)$$

et γ tel que (2.40) soit respecté.

Comme il a été dit précédemment, les multiplicateurs de Lagrange sont utilisés pour exprimer l'effet des contraintes sur les conditions d'optimalité de la distribution. La définition algébrique des multiplicateurs λ_k est claire lorsqu'une solution optimale est donnée. Supposons:

$$\max_{\{x\}} f(x) \quad \text{sujet à} \quad g(x) = b \quad (2.44-2.45)$$

avec solution optimale x^* , $\lambda^* = \{\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*\}$

Si le vecteur b est modifié, on s'attend à ce que les vecteurs optimaux soient également changés. Considérons x^* et λ^* des fonctions continues de b ; les dérivées partielles de l'optimum global $f^* = f(x^*)$ sont:

$$\frac{\partial f^*}{\partial b_k} = \sum_i \frac{\partial f}{\partial x_i} \cdot \frac{\partial x_i^*}{\partial b_k} \quad (2.46)$$

$$\frac{\partial g_r}{\partial b_k} = \sum_i \frac{\partial g_r}{\partial x_i} \cdot \frac{\partial x_i^*}{\partial b_k} = \begin{cases} 1 & \text{si } k = r \\ 0 & \text{sinon} \end{cases} \quad (2.47)$$

Pour chaque k spécifique, le système (2.47) représente M équations dont la multiplication avec les λ_k^* respectifs et la sommation sur k donnent:

$$\sum_i \frac{\partial x_i^*}{\partial b_k} \sum_k \lambda_k^* \frac{\partial g_k}{\partial x_i^*} = \lambda_k^* \quad (2.48)$$

La condition nécessaire d'optimalité (2.36a) rappelle que le deuxième terme égale la dérivée partielle de la fonction objective en respect de x_i . Alors, en se référant à (2.46), on obtient une définition pour λ_k^* :

$$\sum_i \frac{\partial x_i^*}{\partial b_k} \cdot \frac{\partial f}{\partial x_i^*} = \lambda_k^* = \frac{\partial f^*}{\partial b_k} \quad (2.49)$$

Ainsi, le multiplicateur de Lagrange obtenu à l'optimum global équivaut au changement marginal de la valeur optimale atteignable par la fonction objective en respect d'un changement de la valeur de la contrainte associée. Dans les applications géographiques, économiques et en planification urbaine et régionale, les multiplicateurs adoptent plusieurs sens particuliers. Dans un programme linéaire d'allocation minimisant les coûts, ils symbolisent des prix-fantômes rattachés à chacune des alternatives. Dans les modèles d'interaction, ils mesurent l'accessibilité relative de chacune des zones ou interprètent des paramètres de transport. Dans les problèmes de maximisation de l'utilité, ils désignent des «utilités composées» i.e. des valeurs de l'utilité de certain nombre d'alternatives regroupées. Les chapitres 3, 4 et 6 montreront quantité d'exemples où les multiplicateurs de Lagrange se révèlent plus que des variables arbitraires ou de simples instruments mathématiques mais décrivent des relations complexes unissant les variables, relations engendrées par l'imposition des contraintes.

2.3.3 Optimisation avec contraintes d'inégalité

Les contraintes inégalité apparaissent fréquemment dans les problèmes d'optimisation mais leur appréhension exige une méthode différente de celle des multiplicateurs de Lagrange bien que ceux-ci y sont encore sollicités. Elles ont comme rôle de restreindre la région faisable mais n'affecte pas la dimensionalité du problème. Voici un problème générique:

$$\max_{\{x\}} f(x) \quad \text{sujet à } g_k(x) \leq b_k \quad k=1,2,\dots,M \quad (2.50-2.51)$$

où $x \geq 0$ (non-négativité de x)

Cette formulation est générale car elle peut représenter un problème de minimisation car $\max f \equiv \min -f$ et que les contraintes peuvent être inversées en changeant les signes et le symbole logique. Les conditions nécessaires d'optimalité sont fournies par la théorie de «Kuhn-Tucker», du nom de ses protagonistes Kuhn et Tucker [1951]. Ces conditions tiennent à tous les optima locaux inscrits dans la région faisable définie par les contraintes d'inégalité et d'égalité. Par conséquent, la solution d'un programme mathématique peut théoriquement être trouvée en résolvant l'ensemble des équations et des inégalités représentant les conditions nécessaires de Kuhn-Tucker (K-T).

La théorie se base sur le fait que les inégalités tracent des bornes à la région où évolue x et sur lesquelles les dérivées partielles ne s'annuleront pas s'il s'agit d'une «solution frontière» plutôt que d'une «solution intérieure». Dans un problème de maximisation, si la solution optimale tombe sur une borne d'une contrainte d'infériorité, la dérivée est nécessairement positive (ou nulle, cas rare) puisque dans l'éventualité d'une dérivée négative, un déplacement négatif amènerait une valeur supérieure de la fonction objective et cela contreviendrait à l'optimalité de cette solution. Cela vaut inversement pour les contraintes de supériorité et pour les problèmes de minimisation.

La dérivation des conditions K-T s'effectue usuellement en introduisant des variables complémentaires dites «lâches» (slack variables) dans le problème initial de façon à générer des contraintes d'égalité. Reformulons le problème (2.50-2.51):

$$\begin{aligned} \max_{\{x\}} f(x) \quad & \text{sujet à } g_k(x) + s_k = b_k \quad k=1,2,\dots,M & (2.52-2.53) \\ & s \geq 0 & (2.54) \end{aligned}$$

Mis à part les restrictions de non-négativité de s_k ($k=1,2,\dots,M$), on travaille désormais avec des contraintes d'égalité mais aussi avec $M+N$ variables inconnues et M multiplicateurs lagrangiens. Réécrivons le problème sous sa forme lagrangienne:

$$\max L(x,s,\lambda) = f(x) - \lambda'[g(x)+s-b] \quad (2.55)$$

Avec comme conditions nécessaires:

$$\partial \mathcal{L} / \partial x = \nabla f - \lambda \cdot \nabla g = 0 \quad (2.56)$$

$$\partial \mathcal{L} / \partial \lambda = -g(x) - s + b = 0 \quad (2.57)$$

$$\partial \mathcal{L} / \partial s = -\lambda \leq 0 \quad (2.58)$$

La dernière condition est établie du fait que si s_k est positif (solution intérieure), sa dérivée est nulle alors que si s_k égale zéro, sa dérivée est indubitablement négative, soit:

$$s_k \frac{\partial \mathcal{L}}{\partial s_k} \equiv -s_k \lambda_k \equiv [b_k - g_k(x)] \lambda_k = 0 \quad \text{donc } \lambda_k \geq 0 \quad \text{car } s_k \geq 0 \quad (2.59)$$

En enlevant les variables lâches, on définit alors les conditions nécessaires de Kuhn-Tucker qui sont satisfaites pour tout optimum local, donc lors de la solution optimale globale:

$$\nabla f - \lambda \cdot \nabla g = 0 \quad (2.60)$$

$$g(x) \leq b \quad (2.61)$$

$$\lambda \geq 0 \quad \text{et} \quad \lambda \cdot [b - g(x)] = 0 \quad (2.62-2.63)$$

Ces conditions K-T s'avèrent suffisantes pour un programme convexe de maximisation tel que $f(x)$ est concave et les $g_k(x)$ définissent un ensemble convexe ou de minimisation où $f(x)$ et les $g^*(x)$ sont convexes. Sinon, on pourra comparer les diverses solutions vérifiant les conditions K-T et identifiant le point optimal global en évaluant, pour chaque solution, la fonction objective.

Dans les chapitres qui suivront, la détermination des conditions d'optimalité (de la fonction lagrangienne ou de Kuhn-Tucker) demeure fondamentale. Outre d'identifier les propriétés de la solution d'un programme mathématique, ces conditions sont appelées à définir des «modèles de distribution». Dans la plupart des applications, les problèmes seront convexes, garantissant ainsi une seule distribution optimale. Evidemment, les problèmes d'optimisation seront expressément formulés de manière à contenir les principaux facteurs affectant la ou les distributions recherchées. Cette démarche permet de dériver des modèles non seulement issus de principes compatibles mais aussi dotés d'une forme fonctionnelle qui se justifie mathématiquement.

Chapitre 3

MODELISATION DE LA DEMANDE DE TRANSPORT

3.1 INTRODUCTION

Dans un contexte de planification urbaine, l'analyse de la demande de transport est conventionnellement morcelée en quatre phases distinctes. Cette fragmentation, apparemment logique, s'appuie sur un scénario dynamique du processus de décision du voyageur. Dans un premier temps, l'individu se demande s'il doit effectuer un déplacement. Le motif de celui-ci lui commande le choix d'une destination. Il faut ensuite que le voyageur sélectionne le mode de transport qui conviendra et, après coup, de l'itinéraire approprié. Pour l'analyste, ces étapes correspondent plus ou moins aux phases connues sous les noms de «génération des déplacements», de «distribution spatiale», de «répartition modale» et d'«affectation routière».

La décomposition de la demande présuppose une sequentialité des processus qui fut facilement confondue avec l'application consecutive des modèles d'estimation de la demande. En effet, la démarche classique consistait à déterminer un modèle spécifique à chaque niveau de la demande de manière conditionnelle selon laquelle les niveaux inférieurs s'alimentent des résultats des niveaux supérieurs. Cette approche n'empêche pas l'emploi de modèles émanant de rationalités différentes.

Le qualificatif «séquentiel» apparaît trompeur dans le contexte présent. Il est souvent associé à des structures de modélisation «récur-sive» et «hiérarchique». Ici, il indique un arrangement de modèles séparés mais appliqués successivement. Quant à la dénomination réursive, elle se rattache à un processus de nature dynamique et doté de relations de décalage. Par contre, la structure hiérarchique n'a aucune référence temporelle et se concilie tant avec un cadre séquentiel que simultané.

L'essence du problème d'estimation de la demande de transport s'oppose néanmoins à un cadre séquentiel d'analyse. En effet, l'usage consécutif de modèles fait fi d'une part appréciable de l'interdépendance des niveaux.

Si on peut dire que la désagregation successive de la demande tient compte des distributions des niveaux supérieurs, il en va autrement des relations en sens inverse. Les phases plus désagrégées fournissent des mesures décrivant les conditions de transport qu'il importe de considérer chez les phases antérieures. Par exemple, la distribution spatiale (origine-destination) s'appuie sur le coût de transport inter-zonal; celui-ci dépend des coûts associés aux modes disponibles sur le trajet. Par ailleurs, en présence de réseau susceptible à la congestion, le coût modal de voyage varie selon le trafic affecté sur les voies. Il devient alors malaisé de préconiser une démarche séquentielle sans ajustement rétroactif.

Un cadre de modélisation consistant devrait assurer la cohérence des variables à travers tous les niveaux. La hiérarchisation de la demande ne contrevient pas à l'obtention de la consistance car il est possible de maintenir cette structure en adoptant une approche simultanée pour les modèles. Les méthodes de programmation mathématique permettent de formuler des problèmes d'optimisation desquels dérivent des modèles parfaitement intégrés. Ainsi, la structure conditionnelle des distributions n'implique pas obligatoirement une séquentialité des estimations; elle s'insère avantageusement dans un problème d'optimisation unifié.

Le présent chapitre abordera séparément les phases de la demande en exposant les modèles les plus adéquats. On insistera sur la formulation des programmes mathématiques équivalents et on cherchera une notation et une rationalité cohérentes. La phase de génération spatiale sera traitée différemment car le type étudié de demande convient mal au formalisme qui entoure les trois autres phases. En effet, l'évaluation des départs et arrivées se fait de préférence au moyen de modèles linéaires décrivant la relation numérique entre voyageurs et caractéristiques de la zone. Pour les autres phases, on favorisera une démarche classique; énoncé du problème d'optimisation et dérivation des conditions régissant la solution optimale. Deux des sous-chapitres se consacreront à l'intégration de modèles dans le but évident de fournir une formulation unifiée d'un modèle de demande de transport. Au terme du chapitre, on abordera les méthodes d'estimation et de calibration ainsi que la procédure algorithmique requise pour la solution du modèle complètement intégré.

3.2 GENERATION DES DEPLACEMENTS

3.2.1 Définition du déplacement

La phase initiale de l'analyse de la demande de transport reste sans contredit la «génération des déplacements». Dans un environnement urbain multi-zonal, il s'agit d'évaluer les nombres approximatifs de départs et d'arrivées par zone. On parle ici de variables de «production» et d'«attraction» qui sont exprimées respectivement par O_i (pour Origine) et D_j (pour Destination). Cependant, il y a deux définitions distinctes des déplacements. L'une ne spécifie que le mouvement «aller» et, par conséquent, la production est confinée uniquement à l'origine et la destination reste le seul lieu d'attraction. Par ailleurs, certains analystes préfèrent considérer l'aller et le retour comme deux voyages; à l'aller la production est donnée à l'origine et l'attraction, à la destination tandis qu'au retour, cette même destination devient le lieu de production et l'origine est logiquement l'attraction [Stopher, Meyburg, 1975 chap.4]. Cette dernière reste plus apte à l'appréhension des déplacements multiples (i.e. en chaîne) mais tend à mêler les facteurs productifs et attractifs. Ici, on s'en tiendra à la première définition car les déplacements de retour sont implicites. C'est aussi la définition privilégiée dans les enquêtes O-D effectuées par la Société de Transport de la C.U.M. . Ordinairement, les quantités réfèrent à une période de temps particulière, soit en général, une semaine, une journée de semaine ou une heure d'affluence. En dépit des intervalles et définitions différents, la modélisation de la génération de voyageurs reste quasi-identique, mis à part l'intervention de variables spécifiques.

Cette section présentera les deux approches principales rattachées aux modèles d'évaluation du trafic généré par zone. L'une d'elle concerne le caractère générateur des activités localisées dans la zone. L'autre regarde la mobilité individuelle différentielle comme conséquence des caractéristiques particulières de la population voyageuse. Selon Hutchinson [1974 chap.2], celle-ci semble plus s'accorder au volet «production» de la modélisation alors que la première est davantage assortie au volet «attraction». Ici, toutefois, on tentera de concilier les deux approches et de profiter plutôt de leurs compétences spécifiques.

3.2.2 Approche par activité urbaine

Parmi les facteurs affectant les volumes de sorties et d'entrées, il y a sans doute les types et l'ampleur des activités situées dans une zone. Mitchell et Rapkin [1954] soulignèrent la liaison quantitative entre les catégories d'occupation du sol et le trafic généré. La catégorie d'activité implique des genres de déplacements différents; il peut s'agir de migration de travail, d'étude, d'achat, d'affaires, sociale ou récréative. La fréquence migratoire n'en reste pas moins variable. Certaines classes impliquent surtout des départs comme l'activité résidentielle tandis que d'autres suscitent plutôt des arrivées comme les activités industrielle, de bureau, commerciale (vente) et éducative. Il faut incidemment tenir compte de l'intensité de l'activité dans la zone. La densité résidentielle, le nombre d'employés, la taille des établissements commerciaux et institutionnels conditionnent substantiellement les volumes d'interaction produits et attirés.

On peut dresser une liste sommaire des principales catégories d'activités rencontrées dans une région urbaine en soulignant les voyages typiques qu'elles occasionnent. Il faut signaler que chaque catégorie comprend une grande variété d'occupation du sol dont chacune conditionne les motifs et les fréquences de déplacement.

ACTIVITE	MOTIF
1-Terminaux de transport (gares)	d'emploi, voyages interurbains
2-Industrielle (manufacturière et d'entrepôt)	d'emploi
3-Résidentielle	tous les motifs (orig.), de visite (dest.)
4-Hôtelière	d'affaire, de tourisme, d'emploi
5-Récréative	sportif, de spectacle, d'emploi
6-Administrative et financière	d'emploi, d'affaires
7-Commerciale (vente et service)	de consommation, d'emploi
8-Institutionnelle (éducative et religieuse)	d'étude, d'emploi, religieux
9-Médicale	d'emploi, de soins, de visite

Dans les manuels de génie du transport, on trouve usuellement des mesures de fréquences de déplacement par unité de surface pour chaque catégorie d'utilisation du sol ventilée selon leurs variétés respectives et certaines classes de densité [voir Levinson, 1976]. Pour calculer le trafic généré dans une zone, il suffit alors de sommer les produits de ces taux avec la superficie qu'occupe chacune des sous-catégories. Il peut s'agir d'un travail exhaustif car il importe de repérer et de mesurer adéquatement l'ensemble des activités.

Lorsque des observations de O_i et D_j sont disponibles pour un ensemble de zone, l'«analyse de régression» reste l'approche la plus commode et la plus répandue. Les coefficients de régression traduisent autant la capacité génératrice des activités que la corrélation statistique existant entre la présence de l'activité et la mobilité totale. Le modèle requiert certaines variables indépendantes relatives à la taille des activités avec lesquelles il établit une équation linéaire capable de reproduire la variable de transport. L'approche par régression ne nécessite donc pas une ventilation très précise des occupations du sol mais plutôt des indicateurs pertinents de l'intensité de l'activité. Voici une représentation générale de modèles voués à l'évaluation des origines et des destinations:

$$O_i = \alpha_0 + \sum_r \alpha_r X_{r,i} \quad (3.1)$$

$$D_j = \beta_0 + \sum_r \beta_r Y_{r,j} \quad (3.2)$$

où α_r , β_r : coefficients de régression; $X_{r,i}$: variable indépendante indexée r de la zone d'origine i ; $Y_{r,j}$: variable indépendante indexée r de la zone de destination j .

Parmi les variables susceptibles de composer l'équation d'origine, citons la population résidente reflétant le nombre potentiel de voyageurs et l'emploi de bureau symbolisant la production des voyages d'affaire. Quant à l'équation de destination, on compte plusieurs candidats au titre de régresseur. L'emploi total désigne l'attraction à motif de travail, l'emploi de bureau, celle à motif d'affaire, l'emploi dans le commerce de détail ou la superficie de plancher de vente, celle à motif de consommation, les effectifs estudiantins, celle à motif d'étude et la population résidente comme représentant des interactions sociales. D'autres indicateurs du volume de destinataires demeurent valables comme le nombre de lits

d'hôpital, les capacités assises des amphithéâtres, l'achalandage quotidien des gares mais ces mesures devront être proprement calibrées.

L'avantage certain de la démarche régressive reste, d'une part, la souplesse de la méthode, laquelle n'exige que les variables les plus pertinentes et, d'autre part, la possibilité de pratiquer une procédure d'inclusion - exclusion des variables selon leur validité («stepwise regression») [voir Hutchinson, 1974 chap.2]. Dans un contexte d'évaluation de la demande zonale, il reste néanmoins maints inconvénients. D'abord, la méthode de régression n'enregistre pas la mobilité générée par les variables mais bien la covariance des séries. Si les observations sont des zones, on détecte davantage la corrélation évidente des tailles plutôt que la relation numérique vraie entre trafic et variables d'activité. On peut remédier à ce problème en procédant à l'estimation des coefficients avec des variables standardisées par la taille de la zone. Ensuite, la fonction est présumée linéaire en paramètre (i.e. additive) malgré le fait que chacune des variables puisse être transformée individuellement; d'autres formes sont plausibles même si la forme additive se révèle habituellement appropriée [O'Sullivan et al., 1979].

La recherche de coefficients valables pour toute zone de la région urbaine fait oublier la génération différentielle des zones. Certaines caractéristiques externes comme l'accessibilité et la localisation relative de la zone occasionnent des variations de mobilité. Celles-ci sont saisissables à l'aide de variables auxiliaires affectées à chacune des catégories de zones présentant les caractéristiques semblables [Ou, Yu, 1982]. Dans une analyse intra-urbaine, les zones peuvent facilement être groupées en secteurs du centre-ville, quartiers centraux, secteurs suburbains et secteurs périphériques. Une variable auxiliaire est une variable dichotomique (0 ou 1) selon l'appartenance au groupe en question. En appliquant la loi des moindres carrés, les coefficients associés à ces termes binaires deviendront des constantes (ordonnées à l'origine) compensant pour la moyenne de chaque catégorie zonale.

3.2.3 Mobilité individuelle et caractéristique de population

L'analyse de la mobilité individuelle reste un autre domaine d'investigation de la génération des déplacements. Il s'agit d'évaluer la fréquence de transport chez les individus ayant des attributs semblables. Habituellement, on estime cette fréquence comme une fonction des variables mesurant des caractéristiques communes au groupe. La dimension spatiale apparaît facultative car les modèles s'intéressent surtout aux régularités d'un groupe socio-économique. Il y a pourtant diverses formulations de relation entre caractéristiques et mobilité.

La plus intuitive est la tabulation croisée ou analyse de catégorie dans laquelle chaque combinaison de classe est assignée une fréquence moyenne calculée à l'aide d'un échantillon. La fréquence totale de déplacements par zone est obtenue en sommant le produit du taux avec l'effectif de la zone appartenant à la sous-catégorie en question. Aucune présomption quant aux formes fonctionnelles unissant les facteurs à la mobilité n'est nécessaire. On reconnaît usuellement des catégorisations typiques. La structure d'âge, la structure de revenu, les groupes professionnels, la taille du ménage, la possession d'une auto reste les structures les plus classiques [Prendergast, Williams, 1981]. L'approche par catégorie convient peu à une analyse multi-zonale car elle implique une classification exhaustive de la population de chaque zone. Avec même les structures les plus simples, le nombre de sous-groupes augmente rapidement et excède facilement la centaine; pour des petites unités spatiales, de telles données restent indisponibles.

L'approche proposée par Wootton et Pick [1967] semble remédier au caractère exhaustif des classements croisés et discrets. En effet, on préfère plutôt représenter la distribution d'une structure par une fonction de densité de probabilité. Les auteurs présument, par exemple, que les revenus sont distribués dans une population selon une loi «gamma». L'information requise est dès lors réduite car deux paramètres suffisent à la définition de la distribution. Il semble toutefois avantageux d'adopter une fonction de densité r muni de trois paramètres [voir Wellar, LaCava, 1974] car ils raffinent substantiellement la distribution. Considérons le revenu x distribué selon une loi $r(\alpha, \mu, s)$:

$$f(x) = \frac{(x-\mu)^{\alpha-1} \exp(-x/s)}{\Gamma(\alpha) s^{\alpha}} \quad (3.3)$$

$$\text{ou } \Gamma(\alpha) = \int_0^{\infty} z^{\alpha-1} \exp(-z) dz$$

Le paramètre μ situe le maximum de la fonction, s mesure la dispersion et α détermine la forme la courbe de probabilité. Avec ces paramètres, la densité gamma est une généralisation de plusieurs distributions standards comme l'exponentielle-négative, la normale et la χ^2 (chi-carré). Plusieurs distributions socio-économiques peuvent donc être représentées synthétiquement par une loi paramétrisée par trois nombres seulement.

Lorsque la structure socio-économique est connue ou à estimer, elle devra être décrite par une distribution de probabilité connue. Cela nécessite une forte présomption et représente la principale faiblesse de cette approche. Par exemple, Wootton et Pick considèrent que la proportion de ménages possédant n auto(s), $P(n)$, est conditionnée par le revenu x :

$$P(n) = \int_0^{\infty} P(n|x) \cdot f(x) dx \quad (3.4)$$

La probabilité conditionnelle $P(n|x)$ est présumée une fonction parente de la fonction gamma:

$$P(n|x) = a_n \cdot x^{b_n} \exp(-c_n x) \quad (3.5)$$

où a_n , b_n et c_n sont des paramètres relatifs à chaque catégorie n

La fonction (3.5) a l'avantage de décrire la structure socio-économique avec un nombre réduit d'information. Quant à la forme conditionnelle, elle permet de reproduire la répartition croisée des groupes. Par ailleurs, les auteurs appliquent un taux de mobilité propre à chaque catégorie de possession d'auto pour évaluer la fréquence de déplacement.

L'analyse de régression s'applique aussi pour la modélisation de la mobilité individuelle. Il suffit de construire une équation comportant certaines variables d'attributs de la population et d'évaluer par moindres carrés les coefficients associés. De cette façon, le découpage de la population par classe de caractéristiques n'est plus nécessaire; seuls des pourcentages, des moyennes et des médianes de zone suffisent. Stopher et Meyburg [1975 chap.4] rappellent qu'une homogénéité intra-zonale des attri-

buts doit être assurée et préviennent des interprétations trompeuses liées au problème de l'«ecological fallacy». Ce problème se rapporte à l'association erronée que font les chercheurs entre une caractéristique socio-économique d'une zone et les individus s'identifiant à cette caractéristique dans l'interprétation d'une relation. Dans un contexte spatial, il faut conférer aux variables un rôle strictement indicateur plutôt qu'explicatif au sens véritable du mot.

Malgré tout, l'analyse de la mobilité par personne porte surtout sur les déplacements produits au lieu de résidence (les attributs étant enregistrés au domicile des gens). L'approche par activité apparaît alors plus universelle car elle traite aussi bien le trafic à l'origine comme à la destination. Toutefois, comme elle néglige la propension différentielle à la mobilité de la population, il serait bienvenu de préconiser un type de modélisation conciliant les deux conceptions.

3.2.4 Approche mixte

Une approche moins restrictive est recherchée ici. On favorisera l'analyse de régression car elle convient autant au traitement des variables d'activité qu'à celles qualifiant la population. Il s'agit d'abord de spécifier des équations de production et d'attraction comme étant fonctions des activités localisées dans la zones. Parmi les variables d'activités, se retrouvent la «population zonale» sur laquelle une pondération proportionnelle au taux de mobilité sera appliquée. Un sous-modèle régressif sera spécialement formulé pour évaluer le taux de mobilité de la zone. Voici le cadre proposé:

$$D_i = \alpha_0 + \alpha_1 (\bar{Q}_i P_i) + \sum_{r=2}^R \alpha_r X_{r,i} \quad (3.7)$$

$$D_j = \beta_0 + \beta_1 (\bar{Q}_j P_j) + \sum_{r=2}^R \beta_r Y_{r,j} \quad (3.8)$$

$$\text{où } \bar{Q}_i = \psi_0 + \sum_r \psi_r f_r(Z_{r,i}) \quad (3.9)$$

$\alpha_r, \beta_r, \psi_r$: coefficients de régression; P_i : population de la zone i ; $\bar{Q}_i$: taux de mobilité de la zone i ; $X_{r,i}, Y_{r,j}$: variables d'activité liées à la production et à l'attraction de i ou j ; $Z_{r,i}$: variables d'attribut de la population de i ; $f_r(\cdot)$: fonction spécifique à la variable Z_r .

Les modèles de génération ci-dessus s'apparentent évidemment à ceux présentés en (3.1) et (3.2) mis à part l'isolement de la variable population. Dans le sous-modèle de mobilité (3.9), les régresseurs Z_r sont des mesures globales d'une caractéristique de la population: cela peut être une moyenne ou une médiane de zone tout comme un pourcentage d'appartenance à un groupe particulier. L'usage de sous-fonctions $f_r(\cdot)$ est motivé par les relations non-linéaires qui joignent la mobilité aux variables socio-économiques. Par exemple, selon Prendergast et Williams [1981], le revenu est lié logarithmiquement à la fréquence de déplacement: au bas de l'échelle, la croissance de la mobilité est très rapide tandis qu'en haut, le taux marginal tend à être nul. Ainsi, pour le revenu médian de la zone, la fonction s'écrit comme:

$$f_1(Z_{11}) = \ln(Z_{11} - a_1)$$

Pour la taille moyenne des ménages, il semble que la mobilité individuelle tend à décroître car les ménages plus nombreux se composent de plus d'enfants qui, selon la théorie, effectuent moins de déplacements que les adultes. La fonction puissance-inverse devrait alors convenir:

$$f_2(Z_{21}) = (Z_{21})^{-a_2}$$

La structure d'âge d'une zone demeure un critère de mobilité fondamental. Toutefois ici, on considère que l'âge moyen de la population zonale renseigne sur cette structure. Si une telle hypothèse se vérifie, on pourra spécifier une fonction imitant la relation mobilité - âge, laquelle montre une augmentation rapide de la mobilité de l'enfance à l'âge adulte et une diminution lente après la trentaine. Une fonction de type gamma serait appropriée:

$$f_3(Z_{31}) = (Z_{31})^{a_3} \exp(-b_3 Z_{31})$$

Quant à la possession d'un véhicule-moteur, son statut de variable pertinente semble mitigé. Wootton et Pick [1967] observèrent une relation positive plus ou moins linéaire avec le nombre d'autos par ménage. Downes et Morrell [1981] notèrent plutôt que les ménages motorisés ne se déplacent pas plus souvent mais parcourent de plus longues distances. Dans les études de transport urbain nord-américaines, on admet habituellement une augmentation de la mobilité moins que proportionnelle. Dans ce cas, la transformation logarithmique suivante serait convenable:

$$f_4(Z_{41}) = \ln(Z_{41} - a_4)$$

Les transformations proposées se fondent sur l'hypothèse que les relations observées sur les groupes d'individus ont les mêmes formes qu'au niveau des moyennes de zone. Chacune d'elles nécessite une calibration de ses paramètres qui ne pourra se faire qu'en conjonction de l'estimation du modèle de mobilité zonale (3.9). Une procédure sera définie dans le dernier sous-chapitre, consacré aux problèmes d'estimation des modèles de demande de transport.

3.3 DISTRIBUTION SPATIALE DES DEPLACEMENTS

3.3.1 Modèle analogique

Dans le processus de planification des transports, la distribution spatiale apparaît comme l'étape qui consiste à évaluer les volumes de voyageurs, pour une période de temps définie, pour chacune des paires «origine-destination» (O-D). Ces quantités sont usuellement unidirectionnelles et ne réfèrent pas aux réseaux de transport sous-jacents.

Les premiers modèles de distribution spatiale s'appuyèrent sur une hypothèse «gravitaire» par analogie à la loi physique de Newton [voir Reilly, 1931; Stewart, 1941; Carrothers, 1956]. Il n'y avait pourtant aucun fondement théorique soutenant une telle hypothèse si ce n'est qu'une régularité frappante des patrons d'interaction spatiale. A cause de sa commodité et de son universalité, les géographes et autres chercheurs en sciences sociales ont d'abord abondamment exploité la formulation gravitaire de manière déterministe (voir Dillson [1965] pour une revue de la littérature). Malgré tout, la proportionnalité des flux envers la taille des lieux de départ et d'arrivée, souvent appelés "termes de masse" par les géographes, et leur atténuation progressive à l'égard de la distance à parcourir s'avèrent plutôt des relations tautologiques [Sheppard, 1979] que des lois proprement dites. Mais il n'en reste pas moins qu'elles rendent étonnement bien compte des variations quantitatives des interactions spatiales de toutes sortes.

Dans un contexte de transport urbain, les termes de taille peuvent s'identifier aux volumes de voyageurs générés par zones (soient O_i et D_j). Supposant la disponibilité d'une mesure adéquate du coût de transport inter-zonal, un modèle de gravité classique prend la forme suivante:

$$t_{ij} = k O_i D_j f(c_{ij}) \quad \text{où } \partial f / \partial c_{ij} < 0 \quad (3.10)$$

$$k = T / \sum_i \sum_j O_i D_j f(c_{ij}), \quad T = \sum_i O_i = \sum_j D_j$$

t_{ij} : nombre de voyageurs effectuant le trajet $i \rightarrow j$

c_{ij} : coût de transport pour un trajet $i \rightarrow j$

Selon cette approche, le type de fonction de transport $f(c_{ij})$ est fixé exogènement bien que le ou les paramètres puissent être estimés à l'aide d'observations. De plus, le modèle semble insatisfaisant parce que, mis à part le respect d'une contrainte $\sum_j t_{ij} = T$ due à la constante k , les quantités estimées de départs et d'arrivées ne coïncident pas avec les variables d'origine et de destination.

3.3.2 Modèles issus de la maximisation d'entropie

On doit à Murchland [1966] et à Wilson [1967], les premières dérivations d'un modèle d'interaction spatiale à l'aide d'un problème d'optimisation équivalent. Murchland définit un problème de maximisation dont les conditions d'optimalité donnerent un modèle doté d'une fonction exponentielle-négative et respectant les contraintes O-D. Toutefois, le programme mathématique ne recourait à aucune rationalité statistique, comportementale ou économique mais référait analogiquement à la loi thermo-dynamique des gaz. Énonçons le problème en question:

$$\max_{\{T\}} \sum_{i,j} t_{ij} [1 - \ln t_{ij} - f(c_{ij})] \quad (3.11)$$

$$\text{sujet à } \sum_j t_{ij} = O_i \quad \sum_i t_{ij} = D_j \quad t_{ij} \geq 0 \quad (3.12-3.14)$$

La solution est donnée par l'équation d'interaction suivante:

$$t_{ij} = r_i s_j \exp\{-f(c_{ij})\} \quad (3.15)$$

où les paramètres r_i et s_j sont tels que les contraintes (3.12-3.13) sont satisfaites; la restriction de non-négativité (3.14) n'a plus d'incidence car les t_{ij} restent toujours positifs. Bien que Murchland discuta des formes possibles de $f(c_{ij})$, il n'apporta aucune justification quant à l'option d'une fonction quelconque.

L'approche «wilsonienne», très répandue, révolutionna quant à elle le domaine de la modélisation du transport en y introduisant le concept d'entropie statistique. En géographie, l'entropie fut surtout considérée comme une mesure inverse de l'information [voir Berry, Schwind, 1969] car on s'inspira d'abord de la définition fournie par Shannon [1948]. La méthode de maximisation d'entropie que proposa Wilson [1967, 1970] se voulait plutôt une technique de modélisation faisant les plus faibles

présomptions sur une distribution statistique, considérant les renseignements fournis par les contraintes. En fait, toute autre formulation mathématique du modèle reste biaisée car elle contient des informations n'apparaissant pas dans le problème [Gould, 1972]. La méthode sert à la dérivation explicite d'une variété de modèles d'interaction et de localisation spatiale [Wilson, 1970; 1974].

Un problème d'entropie maximale cherche donc la distribution la plus probable parmi celles, innombrables, qui conviennent aux contraintes imposées a priori. Par exemple, dans le cas de la matrice de déplacement $T = \{t_{ij}\}$ $M \cdot N$, le nombre de combinaison parmi le nombre total de flux T est donnée par la fonction multinômiale combinatoire suivante:

$$H = \frac{T!}{\prod_{ij} t_{ij}!} \quad (3.16)$$

Il suffit de maximiser H ou toute autre fonction monotone de H pour obtenir la distribution de flux ayant le maximum d'occurrence. Usuellement, la forme logarithmique de H apparaît plus convenable dans un problème d'optimisation à cause de la grande magnitude des nombres que supposent (3.16). L'approximation de Stirling [voir Wilson, 1970 appendice 1] fournit la forme courante du terme d'entropie:

$$S = \ln H \approx -\sum_{ij} t_{ij} (\ln t_{ij} - 1) \quad (3.17)$$

La maximisation de S sujet à (3.12-3.13) et à une contrainte globale du coût de transport:

$$\sum_{ij} t_{ij} c_{ij} = C \quad \text{où } C \text{ est une constante exogène} \quad (3.18)$$

est solutionnée par:

$$t_{ij} = \exp(-\beta c_{ij} - \eta_i - \nu_j) \quad (3.19)$$

où η_i , ν_j et β sont les multiplicateurs de Lagrange associés respectivement aux contraintes (3.12), (3.13) et (3.18). Wilson [1967] proposa les transformations:

$$a_i = \frac{\exp(-\eta_i)}{O_i} \quad b_j = \frac{\exp(-\nu_j)}{D_j}$$

de manière à obtenir la formulation d'allure gravitaire:

$$t_{ij} = a_i b_j D_j \exp(-\beta c_{ij}) \quad (3.20)$$

$$\text{ou } a_i = 1/(\sum_j b_j D_j \exp(-\beta c_{ij})) \quad b_j = 1/(\sum_i a_i D_i \exp(-\beta c_{ij}))$$

Les paramètres a_i et b_j sont appelés «facteurs de balancement» car ils servent à calibrer le modèle par une méthode itérative de balancement. La non-linéarité explicite de ceux-ci permet de converger vers une distribution $\{t_{ij}\}$ satisfaisant les contraintes marginales. Le β approprié est celui qui s'accorde avec l'égalité (3.18).

Wilson [1971a] fit remarquer que le modèle d'interaction (3.20) est un cas particulier parmi une famille de modèle qu'on différencie par le type de contrainte. Celui-ci est dit «doublement-contraint» car les totaux marginaux de $\{t_{ij}\}$, à l'origine comme à la destination, correspondent aux informations fournies préalablement. Le modèle déterministe (3.10) peut être aussi dérivé par maximisation de l'entropie lorsque des contraintes $\sum_i \sum_j t_{ij} = T$ et (3.18) sont actives; $f(c_{ij})$ est alors de forme $\exp(-\beta c_{ij})$. Des modèles contraints seulement à l'origine ou à la destination font partie de cette famille. Par exemple, le programme non-linéaire:

$$\max_{\{T\}} -\sum_{ij} t_{ij} \left[\ln \frac{t_{ij}}{H_j} - 1 \right] \quad (3.21)$$

$$\text{sujet à } \sum_j t_{ij} = O_i \quad \sum_{ij} t_{ij} c_{ij} = C \quad (3.22-3.23)$$

possède la condition d'optimalité suivante pour t_{ij} :

$$-\ln t_{ij} + \ln H_j - \eta_i - \beta c_{ij} = 0 \quad (3.24)$$

$$\text{soit } t_{ij} = a_i O_i H_j \exp(-\beta c_{ij}) \quad \text{ou } a_i = 1/(\sum_j H_j \exp(-\beta c_{ij})) \quad (3.25)$$

Le terme d'attraction H_j est une variable exprimant une quelconque proportionnalité avec $\sum_i t_{ij}$. La fonction objective (3.21) s'apparente ainsi avec la mesure d'entropie conditionnelle de Kullback [1959]. Dans ce cas, elle équivaut à:

$$-\sum_{ij} t_{ij} (\ln t_{ij} - 1) + \sum_j (\sum_i t_{ij}) \ln H_j \quad (3.26)$$

L'entropie de la distribution est alors maximisée au même titre que l'information contenue dans la série H_j . Ce second terme apparaît comme une présomption du chercheur qui soupçonne une certaine liaison numérique entre le nombre de destinataires et la variable d'attraction. En ce sens,

le programme (3.21-3.23) ne peut plus se considérer comme un problème pur d'entropie maximale.

3.3.3 Modèles issus d'un problème d'optimisation

En modélisation du transport, le terme d'entropie n'est pas utilisé dans le but d'obtenir des modèles sans biais mais bien pour introduire un effet dispersif dans une distribution. En effet, le modèle d'interaction spatiale peut être vu comme la solution d'un problème de minimisation de coût de transport doté d'une fonction de dispersion [Beckmann, Golob, 1971; Evans, 1973]. Ici, il est présumé que les voyageurs cherchent habituellement à réduire leurs dépenses de déplacement mais exhibent néanmoins des comportements spatiaux variables. Le programme non-linéaire correspondant à cette hypothèse est:

$$\max_{\{T\}} -\frac{1}{\beta} \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) - \sum_{i,j} t_{i,j} c_{i,j} \quad (3.27)$$

$$\text{sujet à } \sum_j t_{i,j} = O_i \quad \sum_i t_{i,j} = D_j \quad t_{i,j} \geq 0 \quad (3.28-3.30)$$

Le premier terme de (3.27) est strictement concave [Evans, 1973] tandis que le second est linéaire. La fonction objective reste donc strictement concave (un seul maximum) et les conditions nécessaires d'optimalité de Kuhn-Tucker deviennent ainsi suffisantes:

$$-\frac{1}{\beta} \ln t_{i,j} - c_{i,j} - \eta_i - v_j = 0; \quad t_{i,j} > 0 \quad (3.31)$$

$$\eta_i (O_i - \sum_j t_{i,j}) = 0; \quad \eta_i \text{ sans restriction} \quad (3.32)$$

$$v_j (D_j - \sum_i t_{i,j}) = 0; \quad v_j \text{ sans restriction} \quad (3.33)$$

La contrainte de non-négativité devient ineffective car la solution est intérieure et se rapporte au modèle:

$$t_{i,j} = \exp(-\beta c_{i,j} + \beta \eta_i + \beta v_j) \quad (3.34)$$

La composition de ce programme s'inspire beaucoup du problème émis par Murchland [1966], bien qu'aucune fonction de transport doive être spécifiée préalablement. Par ailleurs, le terme de minimisation du coût remplace la contrainte (3.18) du problème de maximisation d'entropie qui, selon Erlander [1977], demeure toujours difficile à spécifier; toutefois, un coefficient $(1/\beta)$ relativisant le terme d'entropie avec celui du coût doit

être connu. Par ailleurs, la solution de (3.27-3.30) étant mathématiquement équivalente à la maximisation d'entropie de Wilson, β interprète encore le paramètre de transport.

Empiriquement, la taille de β traduit l'atténuation générale du volume de flux causée par le coût de transport. S'il est faible, les voyageurs tiennent peu compte de ce coût et se déplacent facilement dans l'espace (forte variabilité des comportements); s'il est élevé, par contre, les individus deviennent très conscients du prix de voyage et privilégient les parcours O-D les plus économiques. Dans le programme (3.27-3.28), lorsque $\beta \rightarrow 0$, la fonction d'entropie supplante celle du coût total si bien qu'une solution triviale (ou biproportionnelle) indépendante de c_{ij} survient:

$$t_{ij} = \frac{O_i D_j}{T} \quad (3.35)$$

Par contre, lorsque $\beta \rightarrow \infty$, le poids de l'entropie dans la fonction objective s'efface et le programme se réduit au fameux «problème de transport de programmation linéaire» [Evans, 1973]:

$$\min_{\{T\}} \sum_{ij} t_{ij} c_{ij} \quad (3.36)$$

sujet à (3.28-3.30)

La linéarité de la fonction objective suppose pour un certain nombre d'entrées t_{ij} une solution frontière: la contrainte de non-négativité (3.30) redevient ainsi effective. Les M contraintes d'origine et N de destinations impliquent $M+N-1$ entrées indépendantes; la solution contiendra alors tout au plus $M+N-1$ interactions non-nulles. Les éléments du problème conviennent bien à l'efficace algorithme du Simplexe [voir Dantzig, 1963] pour une recherche directe de la solution. Les conditions K-T d'optimalité sont les suivantes:

$$\begin{aligned} c_{ij} - \eta_i - v_j &\geq 0; & t_{ij} &\geq 0 \\ \eta_i &= 0; & t_{ij} &\geq 0 \\ \eta_i &> 0; & t_{ij} &= 0 \end{aligned} \quad (3.37)$$

Les variables duales η_i et v_j symbolisent des prix-fantômes («shadow prices») rattachés à chacune des origines et des destinations. Selon l'interprétation économique classique, ils représentent des coûts cachés

occasionnées par la non-utilisation d'un trajet (ou non-consommation d'un bien dans un sens plus général). Les conditions indiquent que les paires O-D dont le coût (c_{ij}) dépasse la somme des prix-fantômes ($n_i + v_j$) deviennent prohibitives et, conséquemment, n'obtiennent aucun trafic. La distribution est alors (économiquement) optimale car la variabilité des comportements n'existe plus. Ce programme linéaire s'assimile davantage à un problème de l'aménagiste ou du contrôleur de réseau centralisé qu'à la description d'une configuration observable de flux. Il fut pourtant fréquemment sollicité dans l'analyse du transport urbain au cours des années soixante [voir LaCava, Wellar, 1970].

Du point de vue théorique, la solution du problème linéaire (3.28-3.30, 3.36) et la solution triviale (3.35) s'avèrent les cas limites du modèle de distribution spatiale doublement-contraint. La formulation pure d'entropie maximale de Wilson, dotée d'une contrainte de coût total, occulte la présence de ces bornes bien qu'il existe un C minimal en deça duquel une solution est infaisable et un C maximal au-delà duquel la fonction de transport devient invraisemblablement exponentielle-positive.

Erlander [1977] soutint que, dans la plupart des applications courantes, β et C sont tout autant indisponibles. Il préconisa donc l'usage de l'entropie comme contrainte d'un problème de minimisation du coût de transport. Erlander nota que l'entropie figure comme la mesure de variabilité spatiale des voyageurs et représente alors un indicateur de la liberté des voyageurs dans leurs choix de transport. En ce sens, le niveau d'entropie S signifie l'accessibilité générale du système ou, selon le terme employé par l'auteur, l'interactivité.

La mesure S, si standardisée en terme de probabilité, se prête aisément à des comparaisons entre distribution car son domaine est bien connu. Considérons une matrice O-D ($\{t_{ij}\} = \{t_{ij}/T\}$) de M·N entrées dépourvues de contraintes marginales. La distribution uniforme $t_{ij} = 1/MN$ coïncide avec le maximum d'entropie:

$$-\sum_{i,j} (1/MN) \{\ln(1/MN) - 1\} = \ln(MN) + 1 \quad (3.38)$$

Par ailleurs, si la distribution est très inégale i.e. tous les déplacements se concentrent dans un même trajet $i \rightarrow j$, l'entropie égale "un". Si

notre distribution est assujettie à des contraintes marginales du type:

$$\sum_j \tau_{1,j} = o_1 \quad \sum_i \tau_{i,j} = d_j \quad \text{où} \quad o_1 = O_1/T, \quad d_j = D_j/T \quad (3.39-3.40)$$

la distribution la moins présomptueuse reste la solution triviale:

$$\tau_{1,j} = o_1 d_j, \quad \text{alors} \quad (3.41)$$

$$-\sum_{i,j} o_1 d_j \{ \ln(o_1 d_j) - 1 \} = \sum_i o_i \ln o_i - \sum_j d_j \ln d_j + 1 = S^{\max} \quad (3.42)$$

Il est néanmoins plus difficile d'identifier avec exactitude un seuil minimal lorsque les contraintes marginales sont actives. Le niveau d'entropie reste au moins supérieur ou égal à la série marginale la plus dispersée, soit:

$$S^{\min} \geq \max \{ S(o), S(d) \} \quad (3.43)$$

$$\text{où } S(o) = -\sum_i o_i (\ln o_i - 1), \quad S(d) = -\sum_j d_j (\ln d_j - 1)$$

La connaissance des minima et maxima de l'entropie permet à l'analyste de spécifier la variabilité désirée ou déjà observée. Reformulons un problème d'optimisation équivalent au modèle d'interaction spatiale:

$$\min_{\{T\}} \sum_{i,j} t_{i,j} c_{i,j} \quad (3.44)$$

$$\text{sujet à (3.28-3.30) et à } -\sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \geq S^0 \quad (3.45)$$

Les conditions K-T d'optimalité dévoilent des caractéristiques intéressantes du modèle:

$$c_{i,j} - \eta_i - v_j + \gamma \ln t_{i,j} \begin{cases} = 0; & t_{i,j} > 0 & (\gamma \neq 0) \\ \geq 0; & t_{i,j} \geq 0 & (\gamma = 0) \end{cases} \quad (3.46)$$

$$S^0 + \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \begin{cases} = 0; & \gamma > 0 \\ < 0; & \gamma = 0 \end{cases} \quad (3.47)$$

Les conditions précédentes montrent que si la contrainte d'entropie est respectée par égalité, la solution correspond au modèle d'interaction:

$$t_{i,j} = \exp(-c_{i,j}/\gamma + \eta_i/\gamma + v_j/\gamma) \quad (3.48)$$

Si, par contre, elle n'est satisfaite qu'en inégalité, le problème se réduit en celui de programmation linéaire. Conséquemment, le niveau minimal de variabilité est systématiquement atteint en résolvant le problème linéaire; le chercheur peut alors fixer un niveau S^0 arbitraire ($S^0 = 1$) dans le but d'évaluer la valeur d'entropie minimale.

3.3.4 Signification géographique des multiplicateurs de Lagrange

Les multiplicateurs de Lagrange liés aux contraintes d'origine et de destination possède une forte signification géographique s'ils sont convertis en terme de «coût composé de transport». Une mesure dite composée est une agrégation de caractéristiques d'un certain nombre d'alternatives [voir Williams, 1977]. Soit la solution du problème non-linéaire de maximisation (3.27-3.30):

$$t_{i,j} = \exp(-\beta c_{i,j} + \beta \eta_i + \beta v_j) \quad (3.49)$$

Pour que la somme des interactions provenant de la zone i puisse égaler O_i et que celle dirigée vers la zone j puisse équivaloir D_j , il faut que:

$$\eta_i = \frac{1}{-\beta} \ln O_i + \frac{1}{\beta} \ln \sum_j \exp(-\beta c_{i,j} - \beta v_j) \quad (3.50)$$

$$v_j = \frac{1}{-\beta} \ln D_j + \frac{1}{\beta} \ln \sum_i \exp(-\beta c_{i,j} - \beta \eta_i) \quad (3.51)$$

Fixons les coûts composés de la manière suivante [voir Boyce, 1984]:

$$\bar{c}_i = \frac{1}{-\beta} \ln \sum_j D_j \exp(-\beta c_{i,j} + \beta \bar{c}_j) \quad (3.52)$$

$$\bar{c}_j = \frac{1}{-\beta} \ln \sum_i O_i \exp(-\beta c_{i,j} + \beta \bar{c}_i) \quad (3.53)$$

On obtient, par conséquent, la formule de gravité:

$$t_{i,j} = O_i D_j \exp(-\beta c_{i,j} + \beta \bar{c}_i + \beta \bar{c}_j) \quad (3.54)$$

Le coût $\bar{c}_i$ exprime l'accessibilité générale de l'origine i à l'égard de l'ensemble des destinations tandis que $\bar{c}_j$ désigne une mesure analogue pour une destination j à l'égard des origines. La formulation de (3.52) et de (3.53) reste conforme à la notion d'utilité composée en théorie de l'utilité aléatoire [Williams, 1977] dans le sens où la probabilité de choix n'est pas modifiée par l'agrégation. Elle s'accorde aussi avec les mesures d'accessibilité décrites par Hansen [1959]; chacun des coûts consiste en une somme d'un attribut (O_i ou D_j) pondéré par la distance (ici, le coût de transport). Le modèle gravitaire traduit donc une relation numérique entre le flux et les volumes d'arrivées et de départs ainsi qu'un déclin causé par le prix de déplacement corrigé par l'accessibilité relative de l'origine et de la destination concernée.

3.4 MODELISATION DE LA REPARTITION MODALE

3.4.1 Aspects fondamentaux

Dans l'analyse de demande de transport, la répartition modale se définit comme la proportion de voyageurs utilisant un mode parmi un ensemble disponible. Divers éléments sont à considérer dans l'évaluation de cette demande (ou plutôt de la part de marché).

De façon générale, le choix du mode dépend des caractéristiques décrivant les performances comparatives des modes concurrents et de certains attributs des voyageurs. Selon cette conceptualisation essentiellement économique, il semble y avoir cinq aspects fondamentaux à identifier dans un contexte de modélisation [voir Ben-Akiva, Lerman 1985 chap.3]. Le premier est la détermination du marché étudié. Il ne suffit seulement pas de cerner une région d'analyse mais de préciser dans quel environnement, l'individu effectue son choix de transport. Par exemple, selon la présomption la plus courante en planification, les voyageurs décident pour un trajet origine-destination donné. La considération d'un tel «marché» laisse supposer une séquence «choix spatial - choix modal» du processus de décision des voyageurs.

Dans chacun des marchés, existe un ensemble d'alternatives non-vide dans lequel une option est possible. Chacune des alternatives doit être faisable et connue lors du processus de choix. La faisabilité d'un choix modal est définie par une variété de contraintes d'ordre physique (existence du service dans la zone), monétaire (capacité d'usage du mode) et informative (connaissance du service). Il y a aussi des restrictions quant à la distinction des alternatives.

L'attrait d'un mode de transport est évalué d'après un certain nombre d'attributs pertinents. Ceux-ci mesurent, outre les qualités générales du mode (confort, commodité, sécurité, fiabilité) les performances relatives pour un parcours particulier (temps d'accès et de transport, coûts monétaires sur trajet et terminaux). La définition de l'unité de décision consiste à spécifier le sujet du choix. La probabilité de choix peut référer à un individu, un ménage, une catégorie de population ou bien

l'ensemble des voyageurs effectuant un même trajet $i \rightarrow j$. Selon le sujet considéré, des caractéristiques appropriées intervenant dans la perception des attributs modaux doivent être sélectionnées.

La pierre angulaire de toutes les approches économiques modélisant le processus de choix demeure la spécification de la «fonction d'utilité» évaluant les alternatives d'après leurs attributs respectifs et des individus décideurs. La dite fonction résume les caractéristiques en leur accordant un poids ou une élasticité relative selon leurs pertinences dans l'utilité modale. Le net avantage de cette approche s'avère la réduction du vecteur d'attribut à un scalaire plus facile à manipuler.

3.4.2 Détermination de la fonction d'utilité

La composition d'une fonction d'utilité pour un mode pour un trajet $i \rightarrow j$ peut prendre la forme générale suivante (ici, on omet toute catégorisation des voyageurs):

$$u_{T_j} = \alpha_w f_w(W^m) + \alpha_x f_x(X_T) + \alpha_y f_y(Y_j) - \alpha_c f_c(c_{T_j}) \quad (3.55)$$

La variable u_{T_j} est l'utilité du mode m dans un parcours $i \rightarrow j$. W^m est une caractéristique propre au mode m , X_T et Y_j sont des attributs à l'origine et à la destination (accès, stationnement) tandis que c_{T_j} reste le coût de transport modal (d'où le signe négatif interprétant la désutilité). La linéarité paramétrique pour α . n'astreint pas chacune des fonctions $f(\cdot)$ à la forme linéaire; par exemple, si toutes les fonctions sont logarithmiques, alors (3.55) devient multiplicatif et les α . sont des coefficients d'élasticité:

$$u_{T_j} = (W^m)^{\alpha_w} \cdot (X_T)^{\alpha_x} \cdot (Y_j)^{\alpha_y} \cdot (c_{T_j})^{-\alpha_c} \quad (3.56)$$

La fixation préalable de la forme fonctionnelle de l'utilité doit s'appuyer sur une solide connaissance du processus de choix. Malheureusement, cette connaissance apparaît souvent déficiente et l'option d'une fonction classique s'avère pour le moins arbitraire.

Gaudry et Wills [1978] proposèrent l'estimation de la forme fonctionnelle en même temps que celle du modèle de demande en recourant à des transformations paramétrisant la forme comme les transformations «Box-

Tukey» et «Box-Cox» [Box, Cox, 1964]. La transformation Box-Tukey d'une variable positive Z^m s'écrit comme suit:

$$Z^m \langle \psi_z, \mu_z \rangle = \begin{cases} \frac{(Z^m - \mu_z) \psi_z - 1}{\psi_z} & \mu_z \geq 0; \psi_z \neq 0 \\ \ln(Z^m - \mu_z) & \mu_z \geq 0; \psi_z = 0 \end{cases} \quad (3.57)$$

Le paramètre μ_z sert à déplacer la position de la série (Z^m) tandis que la paramètre ψ_z détermine la forme de la variable modifiée. La transformation Box-Cox est obtenue en spécifiant $\mu_z = 0$. Ces transformations possèdent la propriété de continuité au voisinage de $\psi_z = 0$ [Box, Cox, 1964]. Toutes deux sont estimables par vraisemblance maximale [voir Gaudry, Wills, 1978].

Prenons la fonction d'utilité (3.55), remplaçons u_{T_j} par $f_u(u_{T_j})$ et considérons les fonctions f_u, f_w, f_x, f_v et f_c comme des formes Box-Cox dotées respectivement de paramètres $\psi_u, \psi_w, \psi_x, \psi_v$ et ψ_c . Les combinaisons de valeurs que peuvent prendre de tels paramètres font ressortir comme cas spéciaux certaines formes fonctionnelles standards:

$(\psi_u = \dots = \psi_c = 1) \rightarrow$ linéaire

$(\psi_u = \dots = \psi_c = 0) \rightarrow$ log-linéaire ou multiplicative (comme 3.56)

$(\psi_u = 0; \psi_w = \dots = \psi_c = 1) \rightarrow$ exponentielle

La procédure d'estimation ne donnera habituellement pas des valeurs paramétriques aussi homogènes et entières mais plutôt des valeurs fractionnaires auxquelles sont associées des formes fonctionnelles intermédiaires. Néanmoins, des tests statistiques conventionnels émettant des hypothèses du type $\psi = 0, \psi = 1$ permettront d'opter pour une forme standard appropriée.

3.4.3 Modèles économétriques

La plupart des modèles économétriques de choix sont basés sur la maximisation de l'utilité pour le décideur. De ce principe de rationalité, sont générés plusieurs types de formulation dont l'une s'inspire de la théorie du consommateur tandis que d'autres, qui sont probabilistes, ont des présomptions différentes de l'utilité. Entre la présomption de constance et de stochasticité de l'utilité, s'opposent les dérivations par optimisation de l'entropie qui se soustraient à un bon nombre d'hypothèses économiques.

L'objectif de base de la théorie micro-économique du consommateur est de fournir une fonction de demande exprimant l'action du consommateur selon certaines présomptions sur ses préférences. Il s'agit essentiellement d'une théorie individualiste qui s'applique à des groupes par agrégation. Considérons les biens/services $k \in K$ caractérisés par des quantités q_k et des prix r_k . La fonction d'utilité:

$$U(q_k; k \in K) = \prod_k (q_k)^{\gamma_k} \quad (3.58)$$

représente l'utilité de se procurer un bien ou service $k \in K$ en quantité q_k . Les exposants (élasticités) γ_k sont fixés exogènement et traduisent les préférences du consommateur. Soit le problème avec argument $Q = \{q_k\}$:

$$\max_{\{Q\}} U = \prod_k (q_k)^{\gamma_k} \quad (3.59)$$

$$\text{sujet à } \sum_k q_k r_k = R \quad (\text{Revenu disponible}) \quad (3.60)$$

De la solution, sont obtenues les fonctions de demande selon un vecteur-prix $\{r_k\}$ et un revenu R :

$$q_k = \frac{\gamma_k}{\sum_k \gamma_k} \cdot \frac{R}{r_k} \quad (3.61)$$

La calibration des exposants γ_k se fait habituellement par catégorie de consommateur selon des demandes observées. Il ne s'agit pas pourtant d'un modèle de choix discret car le consommateur peut sélectionner toutes les alternatives mais dans des quantités variables. Dans un contexte agrégé, ces répartitions de quantité peuvent toutefois s'assimiler à des proportions de choix. Il reste toujours le problème de définition de l'utilité et du revenu disponible.

Les théories de choix probabiliste apparurent en psychologie pour rendre compte des inconsistances des préférences chez les individus exposés à des alternatives qui leur sont présentées répétitivement. Dans une analyse économique, la probabilité de choix peut cependant être assignée à une proportion d'un groupe faisant des selections discrètes. Ici, cette dernière définition sera adoptée.

Dans un premier temps, regardons les modèles munis d'une fonction d'utilité strictement déterministe. Luce [1959] dérivait un modèle résultant de la présomption suivante qu'il nomma «axiome de choix» mais que d'autres

appelleront «axiome de Luce». Celle-ci présume que les probabilités de choix dépendent seulement des alternatives incluses dans un ensemble et restent indépendantes de toute autre alternative qui pourrait exister. Autrement dit, si $K' \subseteq K$:

$$P(k|K) = P(k|K') \cdot P(K'|K) \quad (3.62)$$

La probabilité d'opter pour l'alternative k tiré d'un ensemble K ou K' n'est pas altérée par l'omission ou l'addition des éléments disjoints de K et K' . Cette propriété est largement connue sous son appellation anglaise de «independence from irrelevant alternatives» (IIA). Cette propriété implique que peu importe le sous-ensemble $K' \subseteq K$, le rapport entre deux probabilités demeure invariable:

$$\frac{P(k|K')}{P(l|K')} = \frac{P(k|K)}{P(l|K)} \equiv \frac{p_k}{p_l} \quad (3.63)$$

Luce [1959] démontra que si l'axiome est valide et que si $p_k \propto f(u_k)$ où u_k est l'utilité de k , on a une fonction d'«utilité stricte» comme équation de partage:

$$p_k = \frac{f(u_k)}{\sum_k f(u_k)} \quad (3.64)$$

L'avantageuse invariance des rapports émanant de la propriété IIA est souvent exploitée dans la formulation de modèle de partage sollicitant une relation d'attributs des options compétitives. Cette simplicité est cependant contre-balançée par un inconvénient si les alternatives possèdent des attributs communs. Mayberry [1970] exposa le «red bus/ blue bus problem» comme un cas extrême de modes concurrents qui se chevauchent. Supposons un partage modal auto - autobus où les deux modes exhibent des caractéristiques d'utilité équivalentes; un modèle comme (3.64) concédera une part de marché égale (1/2, 1/2). Peinturons la moitié de la flotte d'autobus en rouge et l'autre moitié en bleu et considérons les comme deux alternatives rivales; le modèle accordera alors irréalistement des proportions égales (1/3, 1/3, 1/3). Cette anomalie peut toutefois être corrigée en ne définissant que des alternatives rigoureusement distinctes ou en tenant compte de la structure de covariance des attributs [Williams, 1977].

Parmi les modèles économétriques exploitant l'invariance des rapports de proportion, l'un des plus intuitifs est celui inspiré de Berkson [1944] et repris par Theil [1969]:

$$\frac{p^m}{1-p^m} = \exp(\gamma_0) \cdot \prod_r \left[\frac{X_r^m}{\sum_m X_r^m} \right]^{\gamma_r} \quad (3.65a)$$

ou sous sa forme logarithmique:

$$\ln\{p^m/(1-p^m)\} = \gamma_0 + \sum \gamma_r \ln\{X_r^m/\sum_m X_r^m\} \quad (3.65b)$$

Le modèle de régression multiple ci-dessus présente une transformation qui permet de contenir les probabilités dans un intervalle [0,1]. Il s'agit d'une formulation linéarisée d'un modèle «logit» appelé ainsi pour la courbe logistique qu'il dessine dans un graphique probabilité - attributs. Le modèle (3.65a-b) offre très peu de fondement théorique mise à part l'assimilation des paramètres à des coefficients d'élasticité consistants avec la théorie du consommateur [Theil, 1969].

L'approche préconisée dans la «théorie de l'utilité aléatoire» présente un potentiel explicatif beaucoup plus riche. Ben-Akiva [1974], Domencich et McFadden [1975], Stopher et Meyburg [1976] et Manski [1977] décrivent et précisent le formalisme de l'utilité aléatoire dans un contexte de décision de transport et, tout particulièrement, pour le choix modal. Par ailleurs, les monographies de Ben-Akiva et Lerman [1985] (demande de transport) et de Van Lierop [1986] (choix résidentiel) sont des synthèses récentes des développements ayant eu cours depuis les débuts de la théorie. L'analyse s'appuie sur la «stochasticité de l'utilité» dont les sources d'incertitude proviennent (1) des attributs inobservés (et inobservables), (2) des variations perceptuelles et préférentielles, (3) des erreurs de mesure et (4) de l'usage de variables instrumentales ("proxy variables") se substituant aux vraies attributs [Manski, 1977].

Brièvement, la dérivation d'un modèle à partir de l'hypothèse d'une variation aléatoire de l'utilité s'appuie sur la distribution de probabilité du terme d'erreur (ε^m) associé à l'utilité déterministe u^m . Soit l'utilité réelle:

$$\tilde{u}^m = u^m + \varepsilon^m \quad (3.66)$$

où les perturbations ε^m sont considérées «indépendamment et identiquement distribuées» (IID) selon une loi Gumbel (ou Weibull) munie d'un paramètre d'échelle θ . Soit la fonction cumulative de distribution:

$$F(\varepsilon^*) = \Pr\{\varepsilon^* > \varepsilon^m; \forall m\} = \exp\{-\sum_m \exp(-\theta \varepsilon^m)\} \quad (3.67)$$

et la fonction de densité:

$$f(\varepsilon^*) = \partial F / \partial \varepsilon^* = \theta \exp(-\theta \varepsilon^*) \cdot \exp\{-\sum_m \exp(-\theta \varepsilon^m)\} \quad (3.68)$$

La probabilité que les voyageurs perçoivent l'utilité du mode m supérieure à tous les autres est (pour une dérivation semblable, voir Stopher, Meyburg, [1976] ou Kanafani [1983]):

$$p^k = \Pr\{u^k + \varepsilon^k > u^m + \varepsilon^m; \forall m\} \quad (3.69a)$$

$$= \Pr\{\varepsilon^k + u^k - u^m > \varepsilon^m; \forall m\}$$

$$= F(\varepsilon^k + u^k - u^m) = \int_{-\infty}^{+\infty} f(\varepsilon^k + u^k - u^m) d\varepsilon^k$$

$$= \int_{-\infty}^{+\infty} \theta \exp(-\theta \varepsilon^k) \cdot \exp\{-\sum_m \exp(-\theta(\varepsilon^k + u^k - u^m))\} d\varepsilon^k \quad (3.69b)$$

Pour l'intégration, adoptons la règle en chaîne, $v^k = \exp(-\theta \varepsilon^k)$; donc, $\varepsilon^k = -\ln v^k / \theta$ et $d\varepsilon^k = (1/\theta v^k) dv^k$:

$$p^k = \int_0^1 \theta v^k \cdot \exp\{-v^k \cdot \sum_m \exp(-\theta(u^k - u^m))\} / \theta v^k dv^k$$

$$p^k = \frac{-\exp(\theta u^k)}{\sum_m \exp(\theta u^k)} \cdot \exp\left[-v^k \frac{\exp(-\theta u^k)}{\sum_m \exp(-\theta u^k)}\right] \Bigg|_0^1 = 0 + \frac{\exp(\theta u^k)}{\sum_m \exp(\theta u^k)} \quad (3.69c)$$

La proportion de choix est alors obtenue avec un modèle «logit multinomial» (MNL) doté d'un paramètre θ . Celui-ci est intimement lié à la variance d'une variable de loi Gumbel car $\text{Var}(\varepsilon) = \pi^2/6\theta^2$. Une variabilité accrue du terme de perturbation signifiera une fonction logit plus lâche à l'égard de l'utilité.

L'aspect le plus discuté du modèle logit multinomial reste certes la propriété IIA. Il est facile, d'une part, de noter la similitude de (3.69c) avec (3.64) et, d'autre part, d'y reconnaître l'invariance des rapports de probabilité tel que:

$$\frac{p^k}{p^m} = \frac{\exp(\theta u^k)}{\exp(\theta u^m)} \quad (3.70)$$

démontrant l'inaffection du quotient vis-a-vis l'utilité des autres modes.

L'une des difficultés théoriques de cette approche reste la présomption concernant le type de distribution de l'erreur. On peut facilement émettre des réserves sur l'indépendance des erreurs et sur leurs variabilités identiques. La distribution IID-Gumbel est souvent préférée car c'est la seule qui conserve sa forme fonctionnelle à travers un processus de maximisation [Domencich, McFadden, 1975]. Ainsi, si une série $\{u^m\}$ est de loi IID-Gumbel(θ), alors le maximum:

$$\bar{u} = \max\{u^m; \forall m\} \quad (3.71)$$

demeure de type Gumbel avec une variance identique à celle des éléments u^m :

$$\text{Var}(\bar{u}) = \pi^2/6\theta^2 \quad (3.72)$$

et avec une espérance mathématique:

$$E(\bar{u}) = \frac{1}{\theta} \ln \sum_m \exp(\theta u^m) \quad (3.73)$$

Cette moyenne est une forme appropriée de «mesure composée» possédant un certain nombre de propriétés compatibles avec un schème hiérarchisé de choix [Williams, 1977].

Les perturbations $\{e^m\}$ peuvent être aussi considérées comme distribuées selon loi Normale-Multivariée de moyenne nulle et avec une matrice variance-covariance Σ_e . Ainsi, des effets de corrélation peuvent être insérés parmi les termes d'erreurs. De cette loi, on dérive le modèle «probit multinomial» dont l'usage représente toutefois certaines difficultés d'ordre estimatif. Initialement développé dans le cas de choix binaire où la solution du probit binomial reste relativement direct, l'emploi de sa version multivariée sollicite, par contre, un nombre d'intégrations équivalent au nombre d'alternatives moins un [Daganzo, 1980]. En dépit des problèmes inhérents de calcul, l'un des avantages du probit reste l'incorporation de la variation de l'erreur dans une fonction linéaire de l'utilité [voir Ben-Akiva, Lerman, 1985 chap.5] du fait de la compatibilité de la loi Normale avec toute combinaison linéaire de variables issues de même loi.

3.4.4 Modèle dérivant d'un problème de maximisation d'utilité

Dans un contexte de modèle de demande de transport, la maximisation de l'utilité pour un groupe de voyageurs via la programmation mathématique reste la plus simple, la moins exigeante en terme de présomption et la plus

consistante avec les autres modèles de transport. Considérons ici la répartition modale par trajet $i \rightarrow j$, $p_{T_{ij}}$. Il suffit d'un terme d'entropie pour introduire une variabilité aux comportements rationels des voyageurs. Le problème de maximisation suivant permet alors de définir une distribution de proportion:

$$\max_{\{p_{T_{ij}}\}} = \frac{1}{-\theta} \sum_i \sum_j \sum_m p_{T_{ij}} \ln p_{T_{ij}} + \sum_i \sum_j \sum_m p_{T_{ij}} u_{T_{ij}} \quad (3.74)$$

$$\text{sujet à } \sum_m p_{T_{ij}} = 1 \quad \forall i, j \quad (3.75)$$

Les conditions d'optimalité révèle que:

$$-\frac{1}{\theta} \ln p_{T_{ij}} + u_{T_{ij}} - \lambda_{i,j} = 0 \quad (3.76)$$

$$\text{soit } p_{T_{ij}} = \exp(\theta u_{T_{ij}} - \theta \lambda_{i,j}) = \frac{\exp(\theta u_{T_{ij}})}{\sum_m \exp(\theta u_{T_{ij}})} \quad (3.77)$$

On obtient encore une fonction de partage du type logit multinomial qui n'impose toutefois pas l'hypothèse stochastique de l'utilité. La compatibilité avec l'approche de l'utilité aléatoire n'en reste pas la. Il faut noter l'isomorphie du multiplicateur lagrangien $\lambda_{i,j}$ avec la valeur espérée du maximum d'une série IID-Gumbel(θ) comme (3.73):

$$\lambda_{i,j} = \frac{1}{\theta} \ln \sum_m \exp(\theta u_{T_{ij}}) \quad (3.78)$$

Le multiplicateur de Lagrange peut alors s'interpréter comme une mesure d'utilité composée synthétisant les bénéfices encourus sur un trajet en terme du service de transport.

3.5 INTEGRATION DES DISTRIBUTIONS SPATIALE ET MODALE

3.5.1 Distributions simultanées

La section précédente exposa les éléments d'une modélisation de la répartition modale en tant que processus distinct de la demande de transport. Or, le choix du moyen de transport reste étroitement lié au trajet origine-destination sélectionné. D'une part, les voyageurs à leurs lieux d'origine possèdent certaines caractéristiques individuelles influençant le mode approprié à leurs besoins et leurs capacités. D'autre part, la qualité du service de chaque mode diffère d'un trajet à l'autre; ces différences apparaissent à l'origine (accès, desserte), à la destination (coûts terminaux) et sur le parcours proprement dit. Une modélisation intégrant les distributions spatiale et modale s'impose surtout si l'on considère que les voyageurs, dans l'évaluation de l'accessibilité O-D, apprécient l'ensemble des performances des modes disponibles sur un trajet.

Wilson [1970], dans sa présentation de la méthode de maximisation d'entropie, proposa quelques formulations combinant les deux demandes. Selon le format de problème d'optimisation employé ici, la formulation la plus simple est celle où le coût modal de transport est minimisé et l'entropie maximisée, soit:

$$\max_{\{T^m\}} = -\frac{1}{\beta} \sum_m \sum_{i,j} t_{ij}^m (\ln t_{ij}^m - 1) - \sum_m \sum_{i,j} t_{ij}^m c_{ij}^m \quad (3.79)$$

$$\text{sujet à} \quad \sum_m \sum_{i,j} t_{ij}^m = O_i \quad \sum_m \sum_{i,j} t_{ij}^m = D_j \quad (3.80-3.81)$$

où t_{ij}^m : nombre de voyageurs $i \rightarrow j$ usant du mode m et $T^m = \{t_{ij}^m\}$.

La solution du programme ci-dessus est:

$$t_{ij}^m = a_i O_i b_j D_j \exp(-\beta c_{ij}^m) \quad (3.82)$$

$$\text{où } a_i = 1 / \left(\sum_j \sum_m b_j D_j \exp(-\beta c_{ij}^m) \right) \quad b_j = 1 / \left(\sum_i \sum_m a_i O_i \exp(-\beta c_{ij}^m) \right)$$

La faiblesse de ce modèle réside dans son appréhension indifférente des alternatives modales et spatiales; c'est comme si les voyageurs comparaient simultanément et selon les mêmes critères leurs destinations et leurs moyens de transport. Une des formulations alternatives fournies par Wilson est celle où les voyageurs sont catégorisés dès leurs lieux d'origine. Pour chaque catégorie est présumée une perception distincte des coûts de

transport. Cette perception se rattache autant aux préférences et besoins du voyageurs qu'à sa capacité d'user d'un mode. Elle est exprimée par un paramètre de transport distinct par catégorie n (soit β^n). En formulant un problème similaire à (3.79-3.81) où la fonction d'entropie est pondérée par une série $1/\beta^n$ et où les origines sont ventilées par n (O_i^n), on obtient le modèle désagrégé suivant:

$$t_{ij}^m = a_i^n b_j^n \exp(-\beta^n c_{ij}^m) \quad (3.83)$$

où t_{ij}^m est le nombre de voyageurs $i \rightarrow j$ de type n usant du mode m
et où a_i^n et b_j^n sont définis de façon à respecter les contraintes O-D.

3.5.2 Partage modal prioritaire

La désagregation des voyageurs à l'origine suggère une conceptualisation différente du partage modal. Par exemple, si les caractéristiques individuelles pèsent davantage sur le choix du moyen de transport que les attributs de celui-ci, alors une répartition modale précédant le choix du trajet O-D devient fort envisageable. Ainsi, les volumes d'origine sont préalablement partagés en fonction de variables discriminantes. Par exemple:

$$O_i^m = \sum_n p_{in}^m O_i^n \quad (3.84)$$

$$\text{où } p_{in}^m = f(u_{in}^m) \quad (3.85)$$

O_i^m : origines i utilisant le mode m ; p_{in}^m : probabilité de choisir m si on provient de i et on appartient au type n ; u_{in}^m : utilité du mode m pour un voyageur de type n provenant de la zone i .

Etant donné la nature probabiliste du terme p_{in}^m , un format strictement linéaire (voire additif) de la fonction $f(\cdot)$ reste inconvenable. Plutôt, un modèle logit comme celui-ci permettrait de scinder la demande selon m :

$$p_{in}^m = \frac{\exp(\theta u_{in}^m)}{\sum_m \exp(\theta u_{in}^m)} \quad (3.86)$$

L'usage de la répartition logit suppose, d'une part, que les voyageurs peuvent se permettre toutes les alternatives et, d'autre part, que la probabilité de choix dépend essentiellement des attributs modaux et individuels (propriété IIA). Dans un contexte de choix modal fondé sur des

groupes différenciés de voyageurs, on peut difficilement prétendre au libre accès des individus à tous les modes. Le cas le plus évident (et le plus classique) de caractéristique limitant l'usage d'un mode reste la possession d'un véhicule automobile. Ainsi, il est juste de concevoir un phénomène de captivité modale chez une catégorie de population dont la majorité ne dispose pas d'auto. En surcroît, la propriété IIA (voir section 3.4.3) réitère que l'admission d'un nouveau mode ou la modification d'un attribut d'un mode concurrent n'affecte pas le rapport de probabilité de deux options initiales.

Gaudry et Dagenais [1979] introduisirent un modèle encore peu répandu, le «dogit» qui esquive les restrictions de la propriété IIA tout en gardant la possibilité de préserver les attraits du logit pour certaines paires. La forme particulière du dogit ne contraint pourtant pas à écrire des sous-modèles distincts pour certaines alternatives. Elle contient explicitement des variables relatives au «niveau de captivité» de chaque alternative. Écrivons un modèle de distribution de $\{p_{jn}^m\}$:

$$p_{jn}^m = \frac{\exp(\theta u_{jn}^m) + \zeta^{mn} \sum_w \exp(\theta u_{jn}^w)}{(1 + \sum_w \zeta^{wn}) \sum_w \exp(\theta u_{jn}^w)} \quad (3.87a)$$

ou dans une forme équivalente et plus compréhensible:

$$p_{jn}^m = \frac{1}{(1 + \sum_w \zeta^{wn})} \cdot \frac{\exp(\theta u_{jn}^m)}{\sum_w \exp(\theta u_{jn}^w)} + \frac{\zeta^{mn}}{(1 + \sum_w \zeta^{wn})} \quad (3.87b)$$

Le dernier terme de (3.87b) denote la probabilité qu'un voyageur catégorisé n soit captif de l'option m (cette interprétation n'est valide que si les paramètres ζ^{mn} sont non-négatifs; voir Ben-Akiva [1977]). Le premier terme reste la probabilité complémentaire auquel est joint un sous-modèle logit de répartition des voyageurs non-captifs. Le terme de captivité spécifique à chaque mode interprète un seuil minimal correspondant à une part de marché acquise pour le mode peu importe la performance des rivaux. De même que si un tel minimum existe pour une alternative, existent aussi pour ses concurrents des seuils maximaux i.e. des parts de marché impénétrables. Gaudry et Dagenais [1979] démontrèrent l'insubordination en général du dogit à l'égard de l'axiome de Luce [1959]. Considérons le rapport de probabilité de deux options:

$$\frac{p_{jn}^1}{p_{jn}^2} = \frac{\exp(\theta u_{jn}^1) + \zeta^{1n} \sum_w \exp(\theta u_{jn}^w)}{\exp(\theta u_{jn}^2) + \zeta^{2n} \sum_w \exp(\theta u_{jn}^w)} \quad (3.88)$$

Ce quotient ne depend pas seulement des attributs des modes 1 et 2 mais de toutes celles qui sont en compétition. L'indépendance vis-à-vis les autres alternatives disponibles n'apparaît que dans le cas particulier:

$$\zeta^{1n} = \zeta^{2n} = 0$$

où le dogit redevient logit pour la paire considérée et lorsque:

$$\zeta^{1n}/\zeta^{2n} = \exp(\theta u_1^n)/\exp(\theta u_2^n)$$

où le modèle reste de type dogit mais preserve la propriété IIA pour la paire 1-2. Selon la première condition, le logit demeure un cas spécial du dogit dans lequel les paramètres de captivité sont nuls [Gaudry, 1984]. Selon Gunn et Bates [1980], le dogit est une formalisation d'une pratique commune en analyse du transport qui consiste à partager à priori la partie des voyageurs reconnue comme captive et distribuer la partie restante selon une formule logistique. Cette pratique equivaut à fixer exogenement les paramètres ζ^{mn} . Gaudry et Wills [1979] présentèrent toutefois une méthode de vraisemblance maximale permettant l'estimation simultanée des paramètres de captivité et de la fonction de partage logit. Toutefois, cette méthode requiert des proportions observées et impose une linéarisation du dogit à l'aide d'un rearrangement des termes et d'une transformation logarithmique.

La rationalité du dogit reste intuitive bien qu'elle soit issue de la théorie économique de la demande. Il y manque une dérivation mathématique formelle justifiant sa formulation spécifique. Il peut être démontré toutefois qu'un problème d'optimisation adéquatement formulé permet d'obtenir le dogit. Pour cela, il importe de préserver la propriété du dogit de distinguer la probabilité de captivité de celle du mode choisi librement. Dans l'appendice 3.A, nous présentons une dérivation du modèle dogit en recourant aux méthodes de programmation non-linéaire.

La répartition modale prioritaire à la distribution modale comporte un inconvénient majeur, celui de ne pouvoir tenir compte des différentes performances des modes sur un trajet. Ainsi, la connaissance préalable des choix modaux par origine est incompatible avec la notion de coût de transport modal. La seule distribution spatiale possible devient celle où un coût de trajet agrégé est évaluable; par exemple:

$$t_{T,j} = a_T O T b_{T,j} D_j \exp(-\beta \bar{C}_{T,j}) \quad (3.89)$$

avec $\bar{c}_{1,j} = \min_{\{m\}} (cT_{1,j})$ (minimum)

ou $= \frac{1}{M} \sum_m cT_{1,j}$ (moyenne arithmétique)

ou $= -\frac{1}{\xi} \ln \sum \exp(-\xi cT_{1,j})$ (mesure composée)

Bien qu'il existe d'autres façons de déterminer une mesure multimodale, les trois ci-dessus désignent les formules générales les plus communes. Le coût minimal a l'avantage d'être simple à calculer mais méconnaît le choix de la clientèle. La moyenne arithmétique considère tous les coûts modaux à poids égaux, même ceux des moyens quasi-inutilisés. Dans la mesure composée, la contribution de chaque coût modal est relative à la demande; la fonction exponentielle-négative coïncide normalement à celle inscrite dans le logit associé à la répartition modale. Le coût composite reste ainsi la seule mesure consistante avec la théorie de l'utilité aléatoire [Williams, 1977] et les problèmes d'optimisation.

3.5.3 Structure hiérarchisée

Afin de trouver une formulation consistante de distribution spatiale - répartition modale, il convient d'écrire un modèle véritablement intégré plutôt que simplement accolé et d'adopter une structure hiérarchisée (ou nichée) de choix. Une telle structure implique une séquence des processus de distribution telle que l'un est conditionnel au précédent. Ici, le partage modal ne peut être que successif aux interactions O-D car les attributs modaux diffèrent d'un trajet à l'autre.

La programmation mathématique offre, encore une fois, un cadre approprié de modélisation intégrée. Boyce et al. [1983] préconiserent une approche unifiant la distribution spatiale, la répartition modale et l'affectation routière en utilisant une structure nichée de contraintes d'entropie. Le concept d'hiérarchisation de la dispersion permet la création d'une famille de modèles distinguant des sous-processus distributifs. Ici, une version privilégiant notre notation et omettant l'affectation routière (qui sera traitée dans les sous-chapitres 3.6 et 3.7) sera montrée dans le but de convenir aux ambitions de la section présente. Le problème consiste en une simple minimisation du total des coûts par trajet

et par mode $\{c_{T_j}\}$ munie de deux contraintes d'entropie et des contraintes origines-destinations:

$$\min_{\{t_{T_j}\}} = \sum_{i,j,m} t_{T_j} c_{T_j} \quad (3.90)$$

$$\text{sujet à} \quad -\sum_{i,j,m} t_{T_j} (\ln t_{T_j} - 1) \geq S^1 \quad (3.91)$$

$$-\sum_{i,j} (\sum_m t_{T_j}) (\ln (\sum_m t_{T_j}) - 1) \geq S^2 \quad (3.92)$$

$$\sum_m t_{T_j} = O_i \quad \sum_m t_{T_j} = D_j \quad (3.93-3.94)$$

La différentiation de la forme lagrangienne de (3.90-3.94) en respect de t_{T_j} donne l'équation d'optimalité suivante:

$$c_{T_j} + \frac{1}{\theta} \ln t_{T_j} + \frac{1}{\phi} \ln (\sum_m t_{T_j}) - \eta_i - \nu_j = 0 \quad (3.95)$$

soit [voir Boyce et al., 1983]:

$$t_{T_j} = \exp(-\theta c_{T_j} + \theta \eta_i + \theta \nu_j) \cdot \exp\left\{(-\theta/\phi) \ln \sum_m \exp(-\theta c_{T_j})\right\} \quad (3.96a)$$

$$= \frac{\exp(-\theta c_{T_j})}{\sum_m \exp(-\theta c_{T_j})} \cdot a_i O_i b_j D_j \exp(-\beta \chi_{i,j}) \quad (3.96b)$$

$$\text{où} \quad \chi_{i,j} = -\frac{1}{\theta} \ln \sum_m \exp(-\theta c_{T_j}) \quad \beta = \theta - \frac{\theta^2}{\phi} \quad (3.97-3.98)$$

Il s'agit d'un modèle d'interaction spatiale doté d'une fonction d'atténuation du coût composé multi-modale (3.97) et couplé avec une fonction de partage logit multinômiale. L'emploi de la valeur composée $\chi_{i,j}$ reste tout de même implicite. En effet, c'est par une substitution au niveau de (3.96a) que l'on parvient expressement à définir un tel coût de manière à produire une formulation claire. On notera d'après (3.98) que le paramètre β reste toujours inférieur à θ , impliquant que le choix modal est davantage sensible au coût de déplacement que l'interaction spatiale. Cette propriété désirable chez un tel modèle est consistante avec un cadre d'utilité aléatoire comme note Williams [1977].

En sollicitant un format conditionnel du modèle intégré tel que:

$$t_{T_j} = t_{i,j} \cdot p_{T_j} \quad (3.99)$$

une dérivation davantage évidente en résulte [voir Brice, 1989a]. Dans un modèle hiérarchique, p_{T_j} est vu comme la probabilité conditionnelle de choisir un mode m connaissant l'origine et la destination du déplacement i.e. $\Pr(m|ij)$. Si nous substituons t_{T_j} par $t_{i,j} \cdot p_{T_j}$ dans la fonction

objective (3.7^o) et donnons des paramètres différents β and θ , deux fonctions d'entropie pondérées et une fonction de coût total sont obtenues:

$$\begin{aligned} \max_{(T, P^m)} Z = & -\frac{1}{\beta} \sum_{i,j} t_{i,j} \ln t_{i,j} - \frac{1}{\theta} \sum_{i,j} t_{i,j} \sum_m p_{T,j} \ln p_{T,j} \\ & - \sum_{i,j} t_{i,j} \sum_m p_{T,j} c_{T,j} \end{aligned} \quad (3.100)$$

$$\text{soit } a \quad \sum_j t_{i,j} = O_i \quad \sum_i t_{i,j} = D_j \quad \sum_m p_{T,j} = 1 \quad \forall i, j \quad (3.101-3.103)$$

Alors, les conditions d'optimalité deviennent:

$$\frac{\partial Z}{\partial p_{T,j}} = -\frac{1}{\theta} t_{i,j} \ln p_{T,j} - t_{i,j} c_{T,j} + \gamma_{i,j} = 0 \quad (3.104)$$

$$\frac{\partial Z}{\partial t_{i,j}} = -\frac{1}{\beta} \ln t_{i,j} - \frac{1}{\theta} \sum_m p_{T,j} \ln p_{T,j} - \sum_m p_{T,j} c_{T,j} - \eta_i - \nu_j = 0 \quad (3.105)$$

Ainsi,

$$p_{T,j} = \exp\{-\theta c_{T,j} + \theta(\gamma_{i,j}/t_{i,j})\} \equiv \exp(-\theta c_{T,j} + \theta \gamma_{i,j}) \quad (3.106)$$

ou forcement, la variable $\gamma_{i,j}$ formée de $\gamma_{i,j}/t_{i,j}$ est définie de la façon suivante:

$$\gamma_{i,j} = -\frac{1}{\theta} \ln \sum_m \exp(-\theta c_{T,j}) \quad (3.107)$$

Par conséquent, la répartition modale conditionnelle est donnée par une formule logistique:

$$p_{T,j} = \frac{\exp(-\theta c_{T,j})}{\sum_m \exp(-\theta c_{T,j})} \quad (3.108)$$

A partir de (3.103) et (3.106) on trouve une équation pour $t_{i,j}$:

$$\begin{aligned} -\frac{1}{\beta} \ln t_{i,j} - \frac{1}{\theta} \sum_m p_{T,j} (-\theta c_{T,j} + \theta \gamma_{i,j}) - \sum_m p_{T,j} c_{T,j} - \eta_i - \nu_j &= 0 \\ -\frac{1}{\beta} \ln t_{i,j} - \gamma_{i,j} - \eta_i - \nu_j &= 0 \end{aligned} \quad (3.109)$$

$$t_{i,j} = a_i O_i b_j D_j \exp(-\beta \gamma_{i,j}) \quad (3.110)$$

avec lequel on obtient le modèle intégré:

$$t_{T,j} = t_{i,j} \cdot p_{T,j} = a_i O_i b_j D_j \exp(-\beta \gamma_{i,j}) \cdot \frac{\exp(-\theta c_{T,j})}{\sum_m \exp(-\theta c_{T,j})} \quad (3.111)$$

Selon la relation conditionnelle définie en (3.99), les modèles de transport (3.108) et (3.110) sont parfaitement réunis et consistants. L'avantage net de cette dérivation sur la précédente reste, d'une part,

la relation claire entre le coût composé (3.107) et le multiplicateur de Lagrange associé à la contrainte unitaire (3.103). D'autre part, les paramètres pondérant les fonctions d'entropie (soient β et θ) correspondent directement aux paramètres de transport du modèle solution.

La pondération inégale donnée aux fonctions d'entropie séparées, indique que la variabilité statistique du comportement est présumée différente selon le processus de la demande. Pourtant, les paramètres β et θ n'ont pas à être évalués selon cette présomption. Ils doivent avant tout être considérés comme des paramètres de coût de transport et peuvent être fixés a priori.

Intuitivement, les valeurs paramétriques sont sûrement sujettes à des restrictions. Dans un cadre d'utilité aléatoire, le rapport β/θ est lié à la corrélation des composantes d'erreur et doit, par conséquent, rester en deça de l'unité [Ben-Akiva, Lerman, 1985 chap.10]. L'approche d'optimisation présentée ici ne montre malheureusement aucune restriction pour β et θ . Si l'inégalité $\beta \leq \theta$ est respectée, la variabilité accordée à la distribution des déplacements est alors supérieure ou égale à celle du partage modal. Néanmoins, l'interprétation économique des modèles de demande de transport offre une justification raisonnable à l'inégalité [Brice, 1989a]. Un examen des élasticités croisées signale un comportement contre-intuitif quand la condition en question est violée. En effet, en regardant l'intensité de réponse associée à la demande t_{ij} , en respect de c_{ij} , où $w \neq m$, on remarque une inconsistance notable:

$$\varepsilon_{c_{ij}}^{t_{ij}} = \frac{\partial t_{ij}}{\partial c_{ij}} \cdot \frac{c_{ij}}{t_{ij}} = (\theta - \beta) p_{ij} c_{ij} + \beta t_{ij} c_{ij} (O_i^{-1} + D_j^{-1}) \quad (3.112)$$

$$\begin{cases} \geq 0 & \text{si } \beta \leq \theta \\ \geq 0 \text{ ou } < 0 & \text{si } \beta > \theta \end{cases}$$

La dernière condition signifie qu'une amélioration du service d'un mode particulier w sur un trajet $i \rightarrow j$ peut augmenter la demande d'un mode concurrent m . Cette réponse anormale paraît évidemment causée par une plus grande sensibilité de t_{ij} que de p_{ij} , lorsque β est trop grand. Conséquemment, la dispersion consentie à la distribution spatiale doit être relativement plus grande de façon à réduire sa sensibilité au coût de

transport. Senior and Williams [1977] notèrent que la restriction de la taille paramétrique est fréquemment violée dans les études de transport qui emploient le «logit niché» avec une séquence conditionnelle distribution spatiale - répartition modale. Cela dénote vraisemblablement un schème de modélisation inapproprié pour certains systèmes urbains.

3.6 MODELES D'AFFECTATION ROUTIERE

3.6.1 Approches déterministes

Dans l'analyse du transport urbain, l'affectation routière est l'étape consacrée à la distribution des voyageurs, pour qui l'option modale est déjà connue, sur des itinéraires. Pour chaque combinaison spatiale-modale $\langle i, j, m \rangle$, un ensemble de chemins représentant des alternatives plus ou moins distinctes doit être formé. Il ne faut certes pas oublier que chaque itinéraire est composé de «liens» ou «arcs» unidirectionels tirés d'un réseau; par conséquent, un lien l peut appartenir à plus d'un itinéraire et, naturellement, à plus d'un trajet O-D. Les chemins k sont donc plus ou moins "corrélés" entre eux.

La modèle d'affectation routière le plus simple et le plus évident reste «l'affectation tout-ou-rien», c'est-à-dire allouer l'ensemble des voyageurs d'un trajet $i \rightarrow j$ à l'itinéraire dont le coût est minimal. Alébroquement, cela se représente ainsi:

$$h_{Tij}^k = \begin{cases} t_{Tij} & \text{si } c_{Tij}^k = \min_{\langle k \rangle} (c_{Tij}^k) \\ 0 & \text{sinon} \end{cases} \quad k \in K(i, j, m) \quad (3.113)$$

où h_{Tij}^k : nombre de voyageurs faisant le trajet $i \rightarrow j$ sur le mode m et empruntant l'itinéraire k . Le coût modal de transport est alors égal au coût du parcours optimal:

$$c_{Tij} = \min_{\langle k \rangle} (c_{Tij}^k) \quad (3.114)$$

La simplicité attrayante de cette approche paraît contre-balancée par les lacunes suivantes. D'une part, il n'y a pas de variabilité de comportement i.e. tous les usagers connaissent et empruntent systématiquement le plus court chemin. D'autre part, les caractéristiques du réseau n'interviennent pas d'où l'éventualité de segments congestionnés par une repartition par trop abusive et insouciante de la juxtaposition des itinéraires dans un réseau.

Une approche plus élaborée consiste à considérer les itinéraires comme des alternatives (routières) pouvant être sélectionnées selon un critère d'utilité. Dial [1971] proposa une allocation qu'il jugea "probabiliste"

bien qu'aucune dérivation stochastique soit à la source du modèle [Florian, Fox, 1976]. Malgré tout, il s'agissait ni plus ni moins un modèle de partage de type «logit» comportant une fonction du coût par chemin:

$$h_{ij}^k = \frac{\exp(-\rho c_{ij}^k)}{\sum_{k \in K(i,j)} \exp(-\rho c_{ij}^k)} \quad k \in K(i,j) \quad (3.115)$$

Evidemment, selon certaines présomptions, la formulation peut aussi bien originer d'un cadre d'utilité aléatoire que d'un problème d'optimisation isomorphe à ceux posés dans les sections précédentes.

En évitant l'énumération exhaustive des liens composant les itinéraires, le modèle de Dial simplifie beaucoup le problème d'affectation mais néglige en conséquence la configuration topologique du réseau et les capacités limitées des liens. Divers autres inconvénients sont inhérents à ce type de modèle dans des applications de réseau comme soulignèrent Florian et Fox [1976], Daganzo et Sheffi [1977] ainsi que Florian [1981]. L'un des plus notables reste sans doute la sensibilité des proportions estimées à l'égard du nombre de chemins admissibles (inscrits dans les sous-ensembles $K(i,j)$). D'une part, l'ajout d'un chemin dans un sous-ensemble donné perturbe la totalité des proportions initiales et ce, peu importe l'efficacité du nouveau venu. D'autre part, la détermination des itinéraires demeure indubitablement arbitraire et certaines routes sont plus ou moins dissemblables; la variabilité du nombre admis affecte la qualité de l'allocation à la mesure de la similitude entre les parcours [Daganzo, Sheffi, 1977]. Cela tient à la propriété IIA qui dote toute formulation logit multinômiale comme (3.115). En effet, les chemins quasi-identiques (par exemple, tous les arcs communs excepté un) possèdent des attributs tellement semblables que le voyageur ne peut considérer ces chemins comme des alternatives.

Pour surmonter partiellement ces obstacles, une structure nichée des alternatives routières paraît convenable. Il s'agit de dresser un schéma arborescent des chemins en fonction des arcs communs. Selon Sheffi [1979], cette structure doit correspondre aux options qui s'offrent consécutivement au voyageur dans son choix de transport. Il n'y a alors plus d'itinéraire proprement dit mais des sous-itinéraires «consécutifs» dont l'option de l'un deux implique l'option de toutes alternatives préalables. Cette structure peut se visualiser tel un réseau imaginaire et se représenter par

une matrice de covariance identifiant le chevauchement des itinéraires. Dans un autre contexte de transport, Sheffi [1979] y définit un modèle logit hiérarchisé basé sur l'utilité marginale qu'offrent les choix successifs. Cette modélisation s'appuie toutefois sur une séquentialité rigide des options. Il est difficile d'appréhender une structure complexe comme des chemins qui partagent des liens tant au début, au milieu qu'à la fin du parcours. Dans ce cas, une structure de corrélation-croisée apparaît davantage appropriée mais aucun modèle de choix n'a encore été dérivé selon une telle structure [Wilson et al., 1981 chap.3].

3.6.2 Approche par trafic d'équilibre

A propos de l'affectation routière, les modèles d'équilibre de trafic sont à peu près dépourvus des inconvénients occasionnés chez les modèles mentionnés précédemment. L'une des particularités de ces modèles est leur exigence d'une fonction de congestion imitant l'augmentation du temps (coût) de transport causée par les conditions prévalantes de circulation (voir Branston [1976] pour un survol de ces fonctions). On tient compte, par conséquent, des capacités distinctes des liens. Voici la formulation la plus répandue d'une fonction de congestion (celle du F.H.W.A) qui possède les propriétés de continuité, de différentiabilité et de monotonie:

$$s_\ell(v_\ell) = s_\ell^0 \cdot [1 + \alpha (v_\ell/q_\ell)^\tau] \quad (3.116)$$

où $s_\ell(\cdot)$: fonction de congestion pour le lien ℓ ; v_ℓ : volume de trafic sur le lien ℓ ; s_ℓ^0 : coût minimal sur ℓ lorsque le trafic est nul; q_ℓ : capacité technique du lien ℓ ; α, τ : paramètres.

Les modèles d'équilibre sont habituellement fondés sur l'un des deux principes de Wardrop [1952]. Le premier stipule que les voyageurs se répartissent de manière à minimiser le total des coûts de transport du système. Si la demande par trajet $i \rightarrow j$ est exogène s'il n'y a qu'un mode, le problème d'optimisation associé à une affectation «optimale du système» s'écrit comme suit:

$$\min_{\{v, h^k\}} \sum_{\ell} s_{\ell}(v_{\ell}) \quad (3.117)$$

$$\text{sujet à } v_{\ell} = \sum_{i,j,k} \delta_{\ell}^{ij} h_{ij}^k \quad (3.118)$$

$$\sum_k h_{ij}^k = t_{ij} \quad h_{ij}^k \geq 0 \quad (3.119-3.120)$$

h_{ij}^k : voyageurs du trajet $i \rightarrow j$ empruntant l'itinéraire k ; δ_{ij}^k : indice binaire des chemins (=1 si arc i appartient au chemin k , $k \in K(ij)$; = 0 sinon; $v = \{v_i; i=1,2,\dots\}$; $H^k = \{h_{ij}^k\}$ M-N.

Selon la théorie de Kuhn et Tucker [1951], les distributions solutions sont exprimées par les conditions suivantes:

$$\sum_k \left[\frac{\partial s_i}{\partial h_{ij}^k} \right] \cdot \delta_{ij}^k - v_{ij} \begin{cases} = 0; & h_{ij}^k \geq 0 \\ > 0; & h_{ij}^k = 0 \end{cases} \quad (3.121)$$

La solution d'une affectation optimale de système satisfait la condition où les coûts marginaux de déplacement sont égaux pour tous chemins utilisés d'une zone i vers une zone j et aucun chemin inutilisé montre un coût marginal inférieur [LeBlanc, Morlok, 1976; Stewart, 1980]. Ce principe d'allocation ne présume pas que les voyageurs sélectionnent leurs itinéraires de façon à minimiser leurs coûts marginaux [Boyce, Janson, 1980]. Le second principe de Wardrop semble plus accepté car il implique une rationalité individuelle, beaucoup plus réaliste dans un contexte de transport. Le principe déclare que les temps de parcours (ou coût) sur les routes actuellement utilisées dans un trajet O-D sont inférieurs ou égaux que ceux que pourrait connaître un voyageur sur tous les autres chemins inutilisés. Ainsi, selon certaines présomptions, les chemins présentement empruntés sont actuellement les plus courts d'après les conditions circulatoires prévalantes et les perceptions de l'utilisateur [Florian, 1984]. Wardrop nota que le principe suscitait une condition d'équilibre stable où le voyageur ne peut diminuer son coût de transport en changeant de parcours. Beckmann, McGuire et Winsten [1956] furent les premiers à formuler un problème de programmation convexe équivalent au trafic d'équilibre de Wardrop. Voici une formulation supposant une distribution spatiale unimodale $\{t_{ij}\}$ fixe:

$$\min_{\{v, H^k\}} \sum_i \int_0^{v_i} s_i(v) dv \quad (3.122)$$

sujet à (3.118-3.120)

Les conditions d'optimalité de Kuhn-Tucker réitérent la correspondance du problème avec les flux d'équilibre optimaux pour l'utilisateur:

$$\sum_k \delta_{ij}^k s_i(v_i) - g_{ij} \begin{cases} = 0; & h_{ij}^k \geq 0 \\ > 0; & h_{ij}^k = 0 \end{cases} \quad (3.123)$$

Le premier terme de (3.123) désigne le coût de transport réel sur le chemin k du trajet $i \rightarrow j$ en considérant le trafic sur tous les liens composant ce chemin. Le multiplicateur lagrangien $\varphi_{1,j}$ associé à la contrainte de distribution spatiale (3.119) interprète, quant à lui, le «coût d'équilibre» du trajet $i \rightarrow j$. En effet, la variable $h_{k,j}$ ne peut être positive si le coût réel dépasse celui de l'équilibre, égal pour tout chemin $k \in K(i,j)$ emprunté. Evidemment, la validité du modèle s'appuie sur une condition d'«état soutenu» établie sur une période de temps donnée [Florian, 1984].

En exploitant les relations entre les divers types de programmation mathématique, le problème du trafic d'équilibre peut être réécrit comme un problème «non-linéaire complémentaire» (NCP) ou un problème d'«inégalité variationnelle» (VI) [Boyce, 1984]. Reprenons les conditions K-T sous une forme autre que (3.123):

$$v_{k,j} = \sum_l \delta_{k,l} s_l(v_l) - \varphi_{1,j} \geq 0; \quad h_{k,j} \geq 0 \quad (3.124)$$

$$\lambda_{1,j} = t_{1,j} - \sum_k h_{k,j} = 0; \quad \varphi_{1,j} > 0 \quad (3.125)$$

soit, la forme NCP:

$$\sum_{i,j} \left[\sum_k v_{k,j} h_{k,j} + \lambda_{1,j} \varphi_{1,j} \right] = 0; \quad v_{k,j}, h_{k,j} \geq 0, \quad \lambda_{1,j} = 0, \quad \varphi_{1,j} > 0$$

ou dans une formule abrégée:

$$\sum_{i,j,k} h_{k,j} \cdot [\sum_l \delta_{k,l} s_l(v_l) - \varphi_{1,j}] = 0 \quad (3.126)$$

où $v_{k,j}$: multiplicateur de Lagrange liée à la contrainte de non-négativité de $h_{k,j}$ (3.120); $\lambda_{1,j}$: multiplicateur implicite de la positivité du coût d'équilibre $\varphi_{1,j}$.

Dans la forme VI équivalente à la forme NCP, on définit un vecteur de flux $v = \{v_l; l=1,2,\dots\}$, un vecteur fonctionnel $s(v) = \{s_l(v_l); l=1,2,\dots\}$ et un ensemble des vecteurs de flux $F(T)$ satisfaisant (3.118-3.120) pour la distribution $T = \{t_{1,j}\}$ donnée. $F(T)$ est un ensemble convexe compacte non-vide [Nguyen, 1982]. Smith [1979] définit le flot d'équilibre $\bar{v}$ comme suit:

"Un flot $\bar{v}$ est un flot d'équilibre si et seulement si $\bar{v}$ résout le problème d'inégalité variationnelle suivant":

$$s(v) \cdot (v - \tilde{v}) \geq 0 \quad \forall v \in F(T) \quad (3.127)$$

où $v, \tilde{v} \in F(T)$ et " $\cdot$ " dénote un produit scalaire.

Les relations entre les diverses formulations du trafic d'équilibre sont clairement expliquées dans Boyce [1984]. Les algorithmes de solution diffèrent mais permettent d'obtenir la même distribution d'équilibre. La version classique (3.118-3.120, 3.122) expose davantage la minimisation du coût de transport par arc bien que les conditions de solution rappellent, pour chaque trajet $i \rightarrow j$, l'égalité des coûts des chemins empruntés. Le problème NCP traite les conditions de Kuhn-Tucker comme une équation non-linéaire à résoudre. La forme VI montre, quant à elle, la solution d'équilibre comme le flot d'équilibre sur le réseau plutôt que la répartition par itinéraire. Incidemment, ce flot ($\tilde{v}$) est unique alors que le trafic d'équilibre par itinéraire ne l'est pas [Erlander, Nguyen, Stewart, 1979].

Dans toutes les disciplines intéressées par le transport urbain, l'approche par trafic d'équilibre avec optimalité pour l'utilisateur est considérée comme la plus apte à décrire les flux de circulation [Florian, 1984]. Aussi, une attention considérable a été accordée à ce type de modèle, à ses propriétés et à sa solution (pour une couverture complète, voir Florian [1976]). Malgré tout, l'affectation par équilibre comporte quelques défauts qu'il convient de noter. D'abord, il peut y avoir un comportement contre-intuitif des flux et des temps de parcours après l'ajout d'un lien dans le réseau ou une modification de sa longueur. Ce vice est connu sous le nom de «paradoxe de Braess», nommé en l'honneur de son découvreur, Braess [1968]. L'exemple le plus classique est offert par Murchland [1970] dans lequel l'addition d'un arc fait accroître le coût total de transport alors que sa suppression le fait diminuer. L'occurrence du paradoxe dépend beaucoup de la fonction de congestion utilisée et s'avère d'une importance significative dans un contexte de planification d'un réseau. En effet, en adoptant le principe du trafic d'équilibre, il peut ironiquement devenir avantageux de restreindre des portions d'un réseau. Néanmoins, en privilégiant le principe d'optimalité du système, ce problème est évité car le total du coût de transport y est strictement monotone [LeBlanc, 1975]. Par ailleurs, la formulation associée (3.117-3.120) y apparaît nettement moins descriptive mais plutôt normative.

Les deux problèmes d'affectation routière par trafic d'équilibre qui ont été présentes sollicitent une codification complète du réseau de transport. Dans leur forme originale, ils ne s'accommodent pas plus d'un sous-réseau extrait d'un plus grand que d'un sous-réseau dit "grossier" i.e. constitué des axes principaux. Nous proposons dans l'appendice 3.B une version du problème d'affectation qui permet d'isoler les éléments primaires des éléments secondaires du réseau et, du même coup, d'appliquer le modèle à des régions où les informations du réseau secondaire sont indisponibles.

3.7 INTEGRATION DE LA DISTRIBUTION SPATIALE, DE LA REPARTITION MODALE ET DE L'AFFECTATION ROUTIERE

3.7.1 Inadéquation de l'approche par interface

Dans la section 3.5.3, a été discuté le problème de conciliation de la distribution spatiale avec la répartition modale. Une formulation intégrée hiérarchisant les deux processus a été décrite. Dans ce sous-chapitre, il sera question, dans un premier temps, de dévoiler des formulations assimilant les stades de distribution et d'affectation du trafic. Ensuite, pour la spécification d'un modèle de demande de transport parfaitement intégré, il suffira d'adopter une version multi-modale.

Dans les analyses conventionnelles de demande du transport urbain, la distribution origine-destination et l'affectation des chemins sont traitées séparément et consécutivement tels des problèmes indépendants. Cette approche fut jadis désignée comme une approche «intégrée» mais elle apparût plutôt comme un «interface» de sous-modèle. On connaît pourtant bien leur interdépendance qui se manifeste, au niveau de l'interaction spatiale, par une demande fonction du coût de déplacement et, au niveau de l'affectation routière, par la fonction de congestion. Le problème de l'interdépendance fut habituellement surmonté par l'emploi d'un schème itératif (méthode «ad hoc») visant la convergence des résultats par l'un et l'autre modèles de demande. L'itération consiste à procéder séquentiellement au calcul des flux origines-destinations selon un coût fixe et des volumes sur lien selon la demande O-D courante. Cette approche n'assure malheureusement pas une convergence des résultats mais risque plutôt, selon Boyce [1984], de produire des oscillations. En surcroît, elle ne garantit pas une solution satisfaisante du problème d'affectation. C'est pour vaincre cette difficulté inhérente que nombre de chercheurs se sont penchés sur la formulation d'un modèle intégré.

3.7.2 Le problème combinant distribution et affectation

Dans le problème original de trafic d'équilibre de Beckmann et al. [1956], deux termes distincts y apparaissaient, l'un associé à l'affectation, l'autre à la demande spatiale. Le second terme comporte une fonction

dite inverse de demande et traduit l'atténuation de la fréquence des déplacements à l'égard du coût de transport. Voici une présentation du problème utilisant notre notation:

$$\min_{\{T, H^k\}} \sum_i \int_0^{v_i} s_i(v) dv - \sum_{i,j} \int_0^{t_{i,j}} d_{ij}(t) dt \quad (3.128)$$

$$\text{sujet à } v_i = \sum_{j,k} \delta_{ij} h_{jk} \quad (3.129)$$

$$\sum_k h_{kj} - t_{i,j} = 0 \quad h_{kj} \geq 0 \quad (3.130-3.131)$$

La fonction de la demande O-D, $d_{i,j}$, est telle que $t_{i,j} = d_{i,j}(c_{i,j})$. Une contrainte de non-négativité pour $t_{i,j}$ reste superflue car les contraintes (3.130) et (3.131) suffisent pour le contenir dans son domaine. Les conditions d'optimalité du programme convexe sont les suivantes:

$$\sum_i \delta_{ij} s_i(v_i) - g_{i,j} \geq 0; \quad h_{kj} \geq 0 \quad (3.132)$$

$$-d_{ij}(t_{i,j}) + g_{i,j} \geq 0; \quad t_{i,j} \geq 0 \quad (3.133)$$

Du fait que $t_{i,j}$ ne peut être nul que si $g_{i,j} \rightarrow \infty$, l'inégalité de gauche de (3.133) devient généralement une équation égale à zéro [Boyce, 1984]. Implicitement, le programme de Beckmann et al. [1956] donna la première formulation intégrée de distribution - affectation. Cependant, le modèle accentue peu l'aspect spatial, ne référant pas aux départs et arrivées comme contraintes linéaires et ne spécifiant pas la fonction de demande. Evans [1976] nota que si la fonction inverse de la demande est spécifiée telle que:

$$d_{ij}(t) = -\frac{1}{\beta} \ln(t) \quad (3.134)$$

alors le second terme de (3.128) devient un terme de dispersion:

$$- \sum_{i,j} \int_0^{t_{i,j}} -\frac{1}{\beta} \ln(t) dt = \frac{1}{\beta} \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \quad (3.135)$$

Par conséquent, la solution pour la demande spatiale correspond à une équation exponentielle-négative du coût d'équilibre $g_{i,j}$. Cette version s'apparente au modèle intégré proposé par Bruynooghe [1969] qui, toutefois, prit soin d'insérer les contraintes marginales d'origine et de destination.

On doit à Florian, Nguyen et Ferland [1975] et à Evans [1976] les formulations les plus claires desquelles s'inspire la plupart des modèles combinés contemporains. Le modèle de Florian et al. s'appuie sur un pro-

gramme non-linéaire composé de deux termes, l'un relatif aux coûts sur le réseau, l'autre octroyant la dispersion voulue pour la demande O-D. Alors:

$$\min_{\langle T, H^* \rangle} \sum_c \int_0^{v_c} s_c(v) dv + \frac{1}{\beta} \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \quad (3.136)$$

sujet à (3.129-3.131) et à

$$\sum_j t_{i,j} = O_i \quad \sum_i t_{i,j} = D_j \quad (3.137-3.138)$$

Les conditions pour un équilibre des flux par itinéraires sont respectées:

$$\sum_c \delta h_{i,j}^c s_c(v_c) - g_{i,j} \geq 0; \quad h_{i,j}^c \geq 0 \quad (3.139)$$

et la demande spatiale est une fonction gravitaire du coût d'équilibre:

$$\frac{1}{\beta} \ln t_{i,j} + g_{i,j} + \eta_i + v_j = 0; \quad t_{i,j} > 0 \quad (3.140)$$

$$t_{i,j} = \exp(-\beta g_{i,j} - \beta \eta_i - \beta v_j) \quad (3.141)$$

Les conditions d'optimalité (3.139) et (3.140) représentent les conditions nécessaires de Kuhn-Tucker; du fait que la fonction à minimiser est strictement convexe et dotée de contraintes linéaires, ces conditions sont suffisantes [Florian, Nguyen, 1978]. Evans [1976] démontra que la distribution spatiale $\{t_{i,j}\}$ de même les volumes sur arc $\{v_c\}$ sont uniques pour chaque valeur de β . En général, cependant, la distribution des flux par itinéraire $\{h_{i,j}^c\}$ n'est pas unique. Ce constat signifie que pour une distribution O-D et un débit sur un réseau, il peut y avoir diverses affectations acceptables et équivalentes en terme d'optimalité.

Le modèle d'optimisation contient un paramètre libre qui doit être spécifié. Le problème peut être reformulé comme un simple problème d'affectation (minimisation des coûts sur réseau comme (3.125-3.127, 3.129)) muni d'une demande élastique avec une contrainte d'entropie [Erlander, 1977]. Le multiplicateur lagrangien de celle-ci équivaut au coefficient $1/\beta$ de la fonction objective précédente. Il est également possible de calibrer β en ne retenant que le terme d'entropie à maximiser et imposant une contrainte de la fonction de coût [Erlander, Nguyen, Stewart, 1979]:

$$\sum_c \int_0^{v_c} s_c v(c) dv = F \quad (3.142)$$

Par ailleurs, Erlander et al. préviennent que le véritable coût total du système:

$$\sum_e s_e(v_e) \cdot v_e = C \quad (3.143)$$

ne peut être retenu comme contrainte linéaire car la solution de flux selon un coût total donné n'est pas unique. Le recours à cette fonction semble, à première vue, légitime parce que le problème inhérent s'assimilerait à celui de maximisation d'entropie de Wilson [1970] (voir section 3.3.2) dans lequel le coût de transport est variable. Ils démontrèrent que sa monotonie en respect de β n'est aucunement assurée alors que l'entropie totale et la somme des intégrales des fonctions de délais sont monotones.

3.7.3 Version multi-modale du problème de distribution - affectation

Florian [1977] développa un modèle générant un trafic d'équilibre pour deux modes, l'un étant le véhicule privé, l'autre, le transport collectif. Il ne s'agissait pas de définir deux problèmes de minimisation simplement couplés ensemble. L'auteur y inscrivit l'effet d'interaction entre les deux modes qui empruntent les mêmes voies. Néanmoins, la demande spatiale y était fixe. Le modèle combinant les trois stades de la demande de Florian et Nguyen [1978] possède la même structure qu'ils présentèrent dans Florian et al. [1975]. Cependant, le trafic d'équilibre concerne uniquement l'automobile car, pour les usagers du transport public, le coût par itinéraire y est présumé constant, la vitesse et la fréquence des véhicules étant réglées par l'organisme de transport. Il en résulte une affectation «tout-ou-rien» dans le cas du mode collectif. Florian et Nguyen [1978] ne spécifièrent qu'une fonction d'entropie pour la demande spatiale-modale t_{ij} ; par conséquent, les alternatives modales et spatiales sont traitées au même niveau comme le modèle (3.82) suggéré par Wilson [1970]. Boyce et al. [1983] favorisèrent plutôt une approche hiérarchique en imposant des contraintes d'entropie nichées comme (3.91-3.92) à un problème de trafic d'équilibre comme (3.118-3.120, 3.122).

Il est reconnu que le trafic d'équilibre convient d'avantage au comportement d'un automobiliste qu'à un usager des transports de masse. En effet, le ralentissement de la vitesse de conduite causé par le volume excessif de véhicules reste un phénomène inhérent à la circulation automobile. De plus, le comportement optimisant de l'automobiliste est rendu possible par la flexibilité du mode (il peut rectifier son parcours en tout

temps en fonction des conditions prévalantes). Quant aux choix de l'itinéraire en transport en commun, il se fait d'une manière plus restrictive. Ainsi, l'utilisateur opte pour le chemin le plus efficace en fonction des lignes desservant le mieux le site d'origine et le site de destination. Un parcours "physiquement optimal" est écarté s'il implique un temps d'accès trop long ou si le nombre de correspondance paraît abusif. Par surcroît, l'accroissement du volume d'utilisateurs ne suscite généralement pas d'importants délais lorsque la capacité des équipements est suffisante; la présomption de coût invariable demeure alors justifiable.

L'inflexibilité relative du choix d'itinéraires pour le mode collectif n'implique nécessairement pas une affectation radicale «tout-ou-rien». La distribution intra-zonale des sites de départ et d'arrivée fait qu'une variabilité des comportements existe pour les voyageurs effectuant le même trajet inter-zonal. Cette variabilité peut incidemment s'incorporer via un terme d'entropie compatibles avec les autres de même nature. En reprenant le cadre de modélisation nichée décrit dans Brice [1989a] et en lui accolant un terme lié au trafic d'équilibre, on retrouve le programme intégré d'optimisation suivant:

$$\begin{aligned} \min_{(T, H^k, P, G^n)} \quad & \sum_l \int_0^{v_l} s_l(v) dv + \frac{1}{\beta} \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) + \frac{1}{\theta} \sum_{i,j} t_{i,j} \sum_m p_{T,j} (\ln p_{T,j} - 1) \\ & + \frac{1}{\rho} \sum_{i,j,n} g_{T,j} \left[\ln \frac{g_{T,j}}{t_{i,j} p_{T,j}} + \ln N_{i,j} \right] + \sum_{i,j,n} g_{T,j} z_{T,j} \end{aligned} \quad (3.144)$$

$$\text{sujet à} \quad v_l = \sum_{i,j,k} \delta_{i,j}^k h_{i,j}^k \quad (3.145)$$

$$\sum_k h_{i,j}^k - t_{i,j} p_{T,j} = 0 \quad h_{i,j}^k \geq 0 \quad (3.146-3.147)$$

$$\sum_n g_{T,j} - t_{i,j} p_{T,j} = 0 \quad \sum_m p_{T,j} = 1 \quad (3.148-3.149)$$

$$\sum_j t_{i,j} = O_i \quad \sum_i t_{i,j} = D_j \quad (3.150-3.151)$$

où $p_{T,j}$: proportion choisissant le mode m ($m=1$: auto/ $m=2$: public); $g_{T,j}$: clientèle choisissant l'itinéraire n du transport public; $z_{T,j}$: coût de transport $i \rightarrow j$ sur chemin n ; v_l : trafic sur segment l du réseau routier ($m=1$); $K(i,j)$: ensemble des chemins k admis pour $i \rightarrow j$ selon le mode auto; $N(i,j)$: ensemble des chemins n admis pour $i \rightarrow j$ selon le mode public; $N_{i,j}$: cardinalité de l'ensemble $N(i,j)$.

Les deux premiers termes d'entropie concernent la distribution spatiale et la répartition modale. Le troisième se rattache à l'affectation des itinéraires du transport collectif; le terme $\ln N_{1j}$ y apparaît pour compenser le nombre inégal de chemins admissibles par trajet 0-0. La somme des intégrales des fonctions de congestion relève exclusivement de la circulation automobile alors que le total des coûts de transport avec argument zT_j s'applique à la distribution des usagers. Les conditions d'optimalité sont nombreuses car il y a plusieurs variables décisionnelles. Pour l'affectation sur le réseau routier:

$$\sum_c \delta x_j s_c(v_c) - c f_j \geq 0; \quad h f_j \geq 0 \quad (3.152)$$

Pour la répartition des itinéraires du transport en commun:

$$\frac{1}{\rho} \ln g T_j - \frac{1}{\rho} \ln(t_{1j} p f_j) - \frac{1}{\rho} \ln N_{1j} + z T_j - c f_j = 0; \quad g T_j > 0 \quad (3.153)$$

$$g T_j = t_{1j} p f_j N_{1j} \exp(-\rho z T_j + \rho c f_j) \quad (3.154)$$

$$\text{où} \quad c f_j = \frac{1}{-\rho} \ln \sum_m \exp(-\rho z T_j) + \frac{1}{\rho} \ln N_{1j} \quad (3.155)$$

Pour le partage modal, on a pour deux modes;

$$\frac{1}{\theta} t_{1j} \ln p T_j + t_{1j} c T_j - x_{1j}^* = 0; \quad p T_j > 0 \quad (3.156)$$

$$p T_j = \exp\{-\theta c T_j + \theta(x_{1j}^*/t_{1j})\} \quad (3.157)$$

$$\text{où} \quad x_{1j}^* = \frac{x_{1j}}{t_{1j}} = \frac{1}{-\theta} \ln \sum_m \exp(-\theta c T_j) \quad (3.158)$$

Enfin, pour les interactions spatiales, la condition d'optimalité devient:

$$\frac{1}{\beta} \ln t_{1j} + \frac{1}{\theta} \sum_m p T_j \ln p T_j + \frac{1}{\rho} p T_j c T_j + \eta_1 + v_j = 0; \quad t_{1j} = 0 \quad (3.159a)$$

se simplifie considérablement avec les équations solutions trouvées précédemment:

$$\frac{1}{\beta} \ln t_{1j} + x_{1j}^* - \eta_1 - v_j = 0 \quad (3.159b)$$

$$t_{1j} = \exp(-\beta x_{1j}^* + \beta \eta_1 + \beta v_j) \quad (3.160)$$

On retrouve pour chacun des deux modes des équations générales consistantes entre elles. Pour l'auto ($m=1$):

$$\sum_c \delta x_j s_c(v_c) - c f_j \begin{cases} = 0; & h f_j \geq 0 \\ > 0 & h f_j = 0 \end{cases} \quad (3.161)$$

$$\sum_n h_{ij}^n = t_{ij} \cdot p_{ij} = a_i O_i b_j D_j \exp(-\beta \chi_{ij}) \cdot \frac{\exp(-\theta c_{ij})}{\sum_n \exp(-\theta c_{ij}^n)} \quad (3.162)$$

Pour le transport collectif (m=2):

$$g_{ij}^n = a_i O_i b_j D_j \exp(-\beta \chi_{ij}) \cdot \frac{\exp(-\theta c_{ij}^n)}{\sum_n \exp(-\theta c_{ij}^n)} \cdot \frac{\exp(-\rho z_{ij}^n)}{\sum_n \exp(-\rho z_{ij}^n)} \quad (3.163)$$

où $a_i = 1/[\sum_j b_j D_j \exp(-\beta \chi_{ij})]$ $b_j = 1/[\sum_i a_i O_i \exp(-\beta \chi_{ij})]$

Le modèle spatial - modal - routier (3.161-3.163) s'apparente à celui présenté dans Florian et Nguyen [1978] à l'exception d'une structure hiérarchisée pour le partage modal et de la présomption réaliste que le choix d'itinéraire de transport public comporte une composante de variabilité. La version du mode collectif (3.163) s'inspire particulièrement du cadre de modélisation nichée décrit par Williams [1977]. Dans ce cadre, les processus sont dits conditionnels et sont reliés entre eux par des mesures composites appropriées résumant l'utilité du niveau inférieur. Ici, dans le cas de m=2, la variable c_{ij}^n synthétise les coûts de transport connus sur les différents itinéraires n admis dans un trajet $i \rightarrow j$. Dans sa définition (3.155), il tient compte du nombre de chemins admissibles N_{ij} et, de ce fait, il se compare à la mesure rivale c_{ij} . Quant à χ_{ij} , il évalue une mesure d'accessibilité entre i et j en considérant le service offert par les deux modes. Son évaluation tient compte évidemment des options de la clientèle (voir 3.187). Pour m=1, c_{ij} reste le coût d'équilibre, égal pour tous chemins utilisés par les automobilistes pour effectuer un trajet $i \rightarrow j$. L'intérêt théorique de ces mesures, composées ou d'équilibre, demeure leurs définitions en terme de multiplicateur lagrangien des contraintes émises sur le problème principal (3.144-3.151).

Le modèle de trafic d'équilibre capable de partager préalablement les voyageurs entre les parcours principaux et un parcours auxiliaire (présenté dans la section 3.6.3) reste compatible avec le modèle intégré précédent. Il suffit de substituer les termes appropriés dans la fonction objective (3.144). Dans ce cas, il y aurait pour le mode privé un niveau supplémentaire s'interposant entre l'affectation d'équilibre et le partage modal. La section suivante traitera des procédures d'estimation liés à la recherche des solutions numériques des problèmes d'optimisation de demande de transport. On reverra certains algorithmes permettant d'effectuer une recherche efficace des distributions-solutions et/ou des paramètres.

3.8 METHODES D'ESTIMATION POUR LES MODELES DE DEMANDE DE TRANSPORT

3.8.1 Introduction

Les modèles de demande de transport ont été présentés dans ce chapitre sous un aspect essentiellement théorique. A l'exception de la phase de génération de déplacements, nous nous sommes intéressés à la formulation d'un problème d'optimisation et à la détermination des conditions régissant la solution optimale. Dans certains cas, les conditions d'optimalité équivalent à des équations de distribution. Cependant, celles-ci sont habituellement paramétrisées, nécessitant alors une calibration du modèle. Dans d'autres cas, les conditions d'optimalité ne spécifie que les propriétés nécessaires et/ou suffisantes de la solution désirée. Il faut alors recourir à un algorithme capable de chercher efficacement la distribution unique coïncidant avec ces propriétés.

Dans la section 3.8.2, on fera un bref survol de la méthode de régression linéaire multiple. La méthode, d'ailleurs fort connue, convient à l'estimation des modèles (régressifs) de production et d'attraction qui ont été exposés dans la sous-chapitre 3.2. On discutera aussi des équations intrinsèquement non-linéaires et leurs incidences dans la procédure d'estimation. Dans 3.8.3, il s'agit de la calibration des modèles de distribution spatiale et de répartition modale. Ceux-ci ont des structures gravitaire et logistique et peuvent être calibrés selon soit une contrainte de coût total (ou d'utilité totale), soit une contrainte d'entropie. On montrera aussi une méthode appropriée à la calibration d'un modèle hiérarchisé. Enfin, la dernière section (3.8.4) traitera d'un algorithme permettant d'obtenir la solution optimale d'un problème combinant affectation routière (avec trafic d'équilibre) et distribution O-D. Le problème étant convexe et pourvu de contraintes linéaires, il peut être résolu par une procédure algorithmique de la famille des «directions faisables». On étendra l'algorithme au cas conciliant la répartition modale comme le problème décrit par (3.144-3.151).

3.8.2 Analyse de régression multiple

La régression linéaire multiple (deux régresseurs et plus) peut être traitée comme un procédé de calcul purement descriptif ou comme un modèle d'inférence statistique. On s'en tiendra surtout à la première vocation car une méthode d'évaluation de la demande plutôt qu'une de vérification d'hypothèse est recherchée. La méthode régressive vise donc à déterminer une fonction linéaire dotée de coefficients dont les valeurs respectives permettent de minimiser l'écart entre les valeurs observées de la variable régressée et les valeurs prédites par le modèle. Ces écarts sont compris dans un terme d'erreur. Voici une fonction linéaire générale où la variable observée Y dépend des variables indépendantes X_r ; $r=1,2,\dots,P$:

$$Y_1 = \alpha_0 + \alpha_1 X_{11} + \alpha_2 X_{21} + \dots + \alpha_P X_{P1} + \epsilon_1 \quad (3.164a)$$

La notation matricielle permet un format plus compact:

$$y = X\alpha + \epsilon \quad (3.164b)$$

$$y' = (Y_1, Y_2, \dots, Y_N) \quad \epsilon' = (\epsilon_1, \epsilon_2, \dots, \epsilon_N)$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{P1} \\ 1 & X_{12} & X_{22} & \dots & X_{P2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \dots & X_{PN} \end{bmatrix} \quad \alpha' = (\alpha_0, \alpha_1, \dots, \alpha_P)$$

La méthode des moindres carrés consiste à minimiser la somme des erreurs au carré:

$$\min_{(\alpha)} \sum_1 (\epsilon_i)^2 = \epsilon' \epsilon \quad (3.165)$$

On peut substituer à ϵ sa valeur donnée par la relation (3.164b), de sorte que:

$$\epsilon' \epsilon = (y - X\alpha)' (y - X\alpha) \quad (3.166)$$

Le résultat de la minimisation en respect du vecteur α est:

$$X'X\alpha = X'y \quad (3.167)$$

Le vecteur de paramètres est alors obtenu simultanément en résolvant un système formé d'équations normales:

$$\alpha = (X'X)^{-1}X'y \quad (3.168)$$

La matrice $(X'X)$ de taille $(P+1) \times (P+1)$ est communément appelée matrice de variance-covariance et traduit la structure de covariation existant chez les variables indépendantes. Sa forme inversée multipliée par le scalaire $\epsilon' \epsilon / N$ (variance des erreurs ou σ_ϵ^2) est dite matrice variance-covariance

des coefficients; la diagonale se compose donc des erreurs-types des estimateurs. La théorie statistique rappelle en outre que si les variables X_i sont distribuées normalement, alors le vecteur α , combinaison linéaire de celles-ci, possède un comportement distributionnel normal. L'estimateur non-biaisé de la variance d' α :

$$S(\alpha) = \{e'e / (N-P-1)\} \cdot \text{diag}[(X'X)^{-1}] \quad (3.169)$$

suit alors une loi t-Student à $N-P-1$ degrés de liberté. Des tests de validité des coefficients (et, implicitement, des régresseurs) deviennent alors possibles.

Il y a divers concepts rattachés à l'analyse de régression, notamment, celui du coefficient de corrélation simple. Celui-ci mesure l'intensité de la relation linéaire entre deux variables. Il s'agit d'une mesure standardisée i.e. non-influencée par les magnitudes des variables couplées. Par exemple:

$$r_{Y_1} = \text{cov}(Y_1, X_{11}) / [\text{var}(Y_1) \cdot \text{var}(X_{11})]^{1/2} \quad (3.170)$$

où $\text{cov}(Y_1, X_{11}) = \sum_1 Y_1 X_{11} / N - \sum_1 Y_1 \cdot \sum_1 X_{11} / N^2$

$$\text{var}(Y_1) = \sum_1 Y_1^2 - (\sum_1 Y_1 / N)^2 \quad \text{var}(X_{11}) = \sum_1 X_{11}^2 - (\sum_1 X_{11} / N)^2$$

Le coefficient de corrélation permet d'apprécier le degré de relation que partagent la variable dépendante et un régresseur mais aussi deux régresseurs ensemble. Une corrélation excessive entre deux régresseurs paraît nuisible à la fiabilité des estimateurs car elle augmente l'erreur-type des coefficients [voir Draper, Smith, 1966]. En effet, la diagonale de l'inverse de la matrice $(X'X)$ s'accroît davantage que la somme des carrés $e'e$ dans (3.166). Ce problème, connu sous le nom de multicollinéarité, concerne l'importance de la variance commune des régresseurs. Une trop grande proportion de variance commune contrevient non seulement à la présomption d'indépendance des variables mais signifie aussi que l'addition d'une variable déjà corrélée avec une autre n'ajoute rien à la performance prédictive du modèle.

La validité globale d'un modèle de régression se vérifie par la variance "expliquée" ou rendue compte par l'équation linéaire. La variance de la variable observée, ici le vecteur y , est décomposable en variance commune avec les régresseurs et en variance résiduelle ($\text{var}(e)$). La partie

due aux régresseurs équivaut celle du vecteur estimé $\hat{y}$. Le coefficient de détermination multiple mesure donc d'une façon standard la proportion accordée au modèle:

$$R^2_x = \text{var}(\hat{y})/\text{var}(y) = \text{var}(X\alpha)/\text{var}(y) = 1 - \{\text{var}(e)/\text{var}(y)\} \quad (3.171)$$

Ordinairement, pour évaluer la signification statistique de ces rapports, on procède à une «analyse de variance». Il s'agit de comparer la taille de la variance régressive par rapport à la variance résiduelle. On sait que le rapport de deux sommes des carrés de séries (de moyenne nulle et distribuées normalement), chacun pondéré par leurs degrés de liberté respectifs, suit une loi de probabilité de F-Snedecor. Ainsi:

$$F = \frac{\{\hat{y}'\hat{y} - \bar{y}'\bar{y}\}/P}{\{e'e\}/(N-P-1)} \quad \text{suit une loi } F_{P, (N-P-1)} \quad (3.172)$$

où $\hat{y} = X\alpha$: série estimée; $\bar{y}$: vecteur de moyenne de y ou $\hat{y}$; N : nombre d'observations; P : nombre de régresseurs admis.

L'analyse de régression classique s'entoure d'un certain nombre de présomptions à l'égard des séries étudiées [voir Draper, Smith, 1966]. La première est la linéarité des relations entre régresse et régresseur. Si une telle exigence n'est pas rencontrée, des transformations fonctionnelles sont possibles. Deuxièmement, les estimateurs sont sans biais et de variance minimale si la série e conditionnelle à la matrice X est:

- de moyenne nulle $E(e_i) = 0; \forall i$
- de variance constante (homoscédastique) $E(e_i^2) = \sigma_e^2; \forall i$
- non autocorrélée $E(e_i e_j) = 0; \forall i \neq j$
- distribuée normalement $\{e_i\}$ est $N(0, \sigma_e^2)$

Ces présomptions qui confirment les propriétés des estimateurs ne sont en général pas respectées totalement. Dans un contexte spatial, il est difficile d'obtenir un terme d'erreur vraiment aléatoire. L'hétérogénéité des tailles zonales fait que pour les valeurs élevées des X_{i-1} et Y_i correspondent des valeurs fortes de e_i et, par conséquent, des variances plus importantes. Aussi, les tendances régionales qui échappent au modèle induisent chez le terme résiduel une autocorrélation spatiale positive nuisible. Néanmoins, il est possible dans l'étude d'une région urbaine d'employer un système spatial homogène et, par le fait même, de préserver une constance des moyenne et variance de e_i . Ensuite, l'autocorrélation spatiale paraît contrôlable avec l'introduction d'un terme «autoregressif» d'erreur; Cliff

et Ord [1973, 1980] s'employèrent à formuler des modèles d'autorégression pour le terme résiduel en s'inspirant des formes connues dans les analyses typiquement temporelles [voir Johnston, 1972]. Il faut bien reconnaître, toutefois, l'embarrassante complexité et la performance inférieure des modèles autorégressifs conçus pour le domaine spatial.

Dans la section 3.2, on a indiqué que l'introduction de variables auxiliaires permettait de corriger les différences spatiales dans les variables de générations. Elle peut à l'occasion diminuer l'inconstance de la moyenne résiduelle. En effet, la variable auxiliaire absorbe une partie de la variation non-expliquée de y et l'associera à l'hétérogénéité spatiale plutôt que d'accorder aux régresseurs une influence qu'ils ne possèdent pas en réalité. Supposons trois groupes de zones, la nouvelle matrice X aura la forme suivante:

$$X = \begin{bmatrix} 1 & 0 & 0 & X_{11} & X_{21} & \dots & X_{P1} \\ 1 & 0 & 0 & X_{12} & X_{22} & \dots & X_{P2} \\ 0 & 1 & 0 & X_{13} & X_{23} & \dots & X_{P3} \\ 0 & 1 & 0 & X_{14} & X_{24} & \dots & X_{P4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & X_{1N} & X_{2N} & \dots & X_{PN} \end{bmatrix} \quad (3.173)$$

Ici, les deux premiers cas appartiennent au premier groupe tandis que les deux suivants sont issus du deuxième. Le vecteur paramétrique α est évidemment allongé pour contenir les nouveaux coefficients associés aux colonnes binaires de X .

Dans le modèle de mobilité individuelle (3.9), les régresseurs apparaissent comme des fonctions elles-mêmes munies de paramètres. La calibration de ceux-ci ne peut se faire directement en appliquant strictement le critère des moindres carrés. La démarche la plus simple consiste à calibrer séparément chacune des fonctions avec la variable de mobilité $\tilde{z}_i$ observée. La méthode des moindres carrés convient à des fonctions non-linéaires mais les estimateurs ne proviennent plus d'une équation normale comme (3.162) mais d'une procédure itérative. Par exemple, prenons le cas de la taille moyenne des ménages (Z_{21}):

$$\tilde{z}_i = b \cdot (Z_{21})^{-a} + e_i \quad (3.174)$$

Soit minimiser la somme des carrés de la série $\{e_i\}$:

$$\min_{\{a,b\}} \sum_i (e_i)^2 = \sum_i \left[\bar{q}_i - b \cdot (Z_{2i})^{-a} \right]^2 \quad (3.175)$$

L'équation d'estimation pour l'exposant a devient:

$$\sum_i \bar{q}_i (Z_{2i})^{-a} \ln(Z_{2i}) = \sum_i (Z_{2i})^{-2a} \ln(Z_{2i}) \quad (3.176)$$

Tandis que le coefficient b est déterminé directement par:

$$b = \frac{\sum_i \bar{q}_i (Z_{2i})^{-a}}{\sum_i (Z_{2i})^{-2a}} \quad (3.177)$$

L'équation (3.177) est clairement non-linéaire, obligeant ainsi, à suivre une procédure itérative en respect de b . Draper et Smith [1966, chap.10] rappellent que la méthode de régression non-linéaire est équivalente à celle de vraisemblance maximale où la série résiduelle est normale.

En procédant consécutivement à la calibration des sous-fonctions de cette manière, on obtiendra des régresseurs proprement transformés prêts à s'insérer dans le modèle principal (3.9). Une démarche alternative mais plus compliquée serait d'estimer le modèle principal en simultanéité avec les sous-fonctions. Il faut alors recourir à la régression non-linéaire puisqu'au moins un terme n'est pas linéaire. Cette approche est davantage consistante mais son emploi se heurte à de très nombreuses équations d'estimation hautement non-linéaires.

3.8.3 Calibration des modèles de distributions spatiale et modale

Dans un premier temps, on présentera des méthodes conventionnelles de calibration du modèle d'interaction spatiale. Ensuite, il s'agira de désigner une procédure apte à l'estimation simultanée des paramètres spatiaux et modaux. Les méthodes présentées ne conviennent plus si on intègre la phase d'affectation routière selon une présomption d'équilibre. Ce problème sera discuté dans la section 3.8.4.

Si le paramètre de transport β est connu à priori dans le programme non-linéaire (3.27-3.30), la matrice $\{t_{ij}\}$ de solution est facilement trouvée avec une méthode de balancement usuelle. Celle-ci consiste à ajuster les variables duales (ici, a_i et b_j), faisant ainsi respecter alternativement les contraintes d'origine et de destination jusqu'à convergence. La procédure s'appuie sur l'interdépendance des variables:

$$a_i = 1/[\sum_j b_j D_j \exp(-\beta c_{ij})] \quad (3.178)$$

$$b_j = 1/[\sum_i a_i D_i \exp(-\beta c_{ij})] \quad (3.179)$$

Puisqu'il s'agit de contraintes d'égalité, les multiplicateurs lagrangiens associés, η_i et ν_j , sont sans restriction; a_i et b_j , étant des exponentielles de ceux-ci, restent positifs. Leurs valeurs sont déterminées par une valeur initiale arbitraire telle que si le problème est solutionné par les vecteurs $\{a_i\}$ et $\{b_j\}$, alors il est aussi résolu par $\{a_i\} \cdot \sigma$ et $\{b_j\}/\sigma$ car (pour comparer, voir (3.31)):

$$-\frac{1}{\beta} \ln t_{ij} - c_{ij} - (\eta_i + \ln \sigma / \beta) - (\nu_j - \ln \sigma / \beta) = 0 \quad (3.180)$$

Dans une nouvelle analyse de la demande de transport, il semble fort possible que le paramètre β soit inconnu. Lorsque l'entropie maximale est recherchée, celui-ci dépend strictement du total (ou de la moyenne) des coûts de transport dans le système spatiale étudié [voir Erlander, Stewart, 1978]. Si ce total est connu, on peut exploiter les conditions d'optimalité du problème (3.27-3.30) ou, d'une façon équivalente, considérer le problème comme une stricte maximisation d'entropie avec contrainte de coût comme formula Wilson [1967; 1970] (voir 3.17-3.18). En fait, dans la méthode proposée, la fonction de dispersion n'est pas sollicitée explicitement car le modèle gravitaire (3.20) assure l'optimalité de la dite fonction. La méthode d'estimation consiste uniquement à faire respecter la contrainte suivante:

$$C' = \sum_{i,j} t_{ij} c_{ij} = C(\beta) \quad (3.181)$$

où $t_{ij} = a_i D_i b_j D_j \exp(-\beta c_{ij})$ et a_i, b_j définis selon (3.178-3.179)

La fonction $C(\beta)$ contient plusieurs fois le terme β , dans la fonction frictionnelle et dans les variables de balancement. Cela implique une procédure itérative de recherche du paramètre. Pour l'estimation de β , on peut utiliser une approximation de $C(\beta)$ avec l'expansion de Taylor de première ordre correspondant à $\beta = \beta'$:

$$C(\beta' + i) \approx C(\beta') + i \cdot \frac{\partial C(\beta)}{\partial \beta} \Big|_{\beta=\beta'} \quad (3.182)$$

Le terme i est un incrément infinitésimal. Soit β^* satisfaisant (3.181), le remplacement de i par l'écart $(\beta^* - \beta')$ revient à projeter linéairement la

tangente au point β' . Ainsi, l'expansion:

$$C(\beta^*) \approx C(\beta') + (\beta^* - \beta') \cdot \frac{\partial C(\beta)}{\partial \beta} \Big|_{\beta=\beta'} \quad (3.183)$$

permet de dériver la procédure incrémentielle:

$$\beta^{r+1} = \beta^r + \frac{C^* - C(\beta^r)}{\partial C(\beta^r) / \partial \beta^r} \quad (3.184)$$

Les suscrits r indiquent le numéro de l'iteration. Cette procédure de recherche est nommée méthode de Newton-Raphson et est réputée pour sa convergence rapide vers le ou les paramètres optimaux [Batty, Mackie, 1972]. L'inconvénient principal tient au fait qu'il faille évaluer numériquement la dérivée première à chaque itération. Pour un terme t_{1j} issu d'un modèle gravitaire doublement-contraint, il n'y a pas de dérivée exacte car les coefficients a_i et b_j sont infiniment interdépendants. Une bonne approximation est obtenue si l'on se restreint à la première apparition de a_i et b_j sous leurs formes fonctionnelles:

$$t_{1j} = \frac{O_i D_j \exp(-\beta c_{1j})}{[\sum_j b_j D_j \exp(-\beta c_{1j})] \cdot [\sum_i a_i O_i \exp(-\beta c_{1j})]} \quad (3.185)$$

Ainsi:

$$\frac{\partial C(\beta)}{\partial \beta} \approx \sum_{i,j} t_{1j} c_{1j} \cdot [-c_{1j} + \frac{1}{2}(\bar{c}_i + \bar{c}_j)] \quad (3.186)$$

où $\bar{c}_i = \sum_j t_{1j} c_{1j} / O_i$ et $\bar{c}_j = \sum_i t_{1j} c_{1j} / D_j$

Erlander [1977] souligna la difficulté d'obtenir le total des coûts de transport dans un système donné. La moyenne individuelle des coûts occasionnés n'est pas exclusivement une question comportementale mais tient au patron spatial des origines et des destinations. De plus, lorsque ce total est une variable à minimiser comme c'est le cas dans un problème incluant une affectation d'équilibre, il ne peut servir de facteur de calibration. Conséquemment, il faut se rabattre sur un autre terme influençant la valeur de β , la fonction d'entropie. En effet, dans la fonction objective du problème d'optimisation (3.27), le coefficient de la fonction entropique est la réciproque de β . Dans la version minimisant uniquement le total des coûts, la contrainte d'entropie (programme 3.28-3.30, 3.44-3.45) est dotée d'un multiplicateur de Lagrange équivalent à $1/\beta$. Une procédure de recherche semblable à celle décrite précédemment convient à l'égalisation de la contrainte selon un niveau S' prédéterminé. Comme il a été démontré dans 3.3.3, le niveau d'entropie est une mesure dont on connaît les valeurs

extrêmes: la variabilité des comportements plutôt que la géographie intra-urbaine affecte la valeur de l'entropie.

La stricte concavité de la fonction d'entropie autorise une simple recherche paramétrique conduite par la dérivée première:

$$\beta^{r+1} = \beta^r + \frac{S' - S(\beta^r)}{\partial S(\beta^r) / \partial \beta^r} \quad (3.187)$$

où $S(\beta) = - \sum_{i,j} t_{i,j} \ln t_{i,j}$

L'approximation de la dérivée se fait selon la définition (3.20) de $t_{i,j}$:

$$\frac{\partial S(\beta)}{\partial \beta} \approx \sum_{i,j} t_{i,j} \ln t_{i,j} \cdot [c_{i,j} - \frac{1}{2}(\bar{c}_i + \bar{c}_j)] \quad (3.188)$$

A chaque iteration, la matrice $\{t_{i,j}\}$ doit être réévaluée selon la nouvelle valeur de β ; une procédure sub-itérative correspondant à la méthode de balancement bi-proportionnel est exécutée afin de mettre à jour $\{t_{i,j}\}$.

Le modèle combinant la distribution spatiale et la répartition modale a une structure similaire à celle du modèle gravitaire. Le problème d'optimisation équivalent (3.79-3.81) ressemble à (3.100-3.104), mis à part la scission de la fonction d'entropie. Cette séparation donne lieu à une structure hiérarchique des processus de distribution. Le programme mathématique contient deux paramètres qui permettent, lors de la solution, de produire deux fonctions distinctes de partage. La structure du modèle combiné est telle qu'une mesure composée du coût $i \rightarrow j$ à travers les modes est utilisée par le modèle gravitaire.

Dans une analyse où les paramètres de transport β et de partage modal θ sont fixés préalablement, la solution du problème d'optimisation spatial-modal est donnée par les conditions d'optimalité. Il suffit de calculer $t_{i,j}$ et pT_j de la façon suivante:

$$t_{i,j} = a_i O_i b_j D_j \exp(-\beta \gamma_{i,j}) \quad (3.190)$$

$$pT_j = \exp(-\theta cT_j + \theta \gamma_{i,j}) \quad (3.191)$$

$$\text{où } \gamma_{i,j} = -\frac{1}{\theta} \ln \sum_m \exp(-\theta cT_j) \quad (3.192)$$

$$a_i = 1 / [\sum_j b_j D_j \exp(-\beta \gamma_{i,j})] \quad b_j = 1 / [\sum_i a_i O_i \exp(-\beta \gamma_{i,j})]$$

Encore une fois, l'indisponibilité courante des deux paramètres oblige à procéder à une recherche paramétrique. Si les modèles spatial et modal étaient distincts, ils pourraient se calibrer avec une contrainte de coût de transport total. Ce type de contrainte apparaît en tant qu'équation d'estimation pour les modèles gravitaires et logistiques selon les approches de vraisemblance maximale et de maximisation d'entropie [Wilson, 1970]. La contrainte linéaire suivante semble à première vue appropriée à la calibration simultanée de β et θ :

$$\sum_{i,j} t_{i,j} \sum_m p_{T_j} c_{T_j} = C' \quad (3.193)$$

Toutefois, cette équation n'apparaît pas suffisante pour un modèle niché car une seule contrainte ne nécessite qu'un seul multiplicateur. En fait, un seul paramètre peut être estimé selon ce schéma (soit en fixant l'autre ou en présupposant l'égalité des deux).

Les modèles nichés sont calibrés en pratique d'une manière séquentielle; chaque niveau est calibré séparément, commençant par le niveau inférieur jusqu'au plus agrégé et, entre eux, calculant la mesure composée requise par la formulation suivante [voir Ben-Akiva, Lerman, 1985, chap.10]. Cependant, cette procédure heuristique à multi-étape ne s'accorde pas avec le cadre d'optimisation décrit précédemment où une approche unifiée est cherchée. De plus, la présence multiple de θ cause sans aucun doute certains problèmes d'interdépendance qui sont ignorés dans une approche séquentielle.

L'utilisation des contraintes d'entropie autorise par contre une approche simultanée et consistante avec la programmation mathématique. En effet, le problème de maximisation (3.100-3.104) équivaut à:

$$\min_{\{T,P\}} Y = \sum_{i,j} t_{i,j} \sum_m p_{T_j} c_{T_j} \quad (3.194)$$

$$\text{soit à } \sum_j t_{i,j} = D_i \quad \sum_i t_{i,j} = D_j \quad \sum_m p_{T_j} = 1 \quad \forall i,j \quad (3.195-3.197)$$

$$- \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \geq S^T \quad - \sum_{i,j} t_{i,j} \sum_m p_{T_j} (\ln p_{T_j} - 1) \geq S^P \quad (3.198-3.199)$$

Les multiplicateurs lagrangiens des contraintes (3.198) et (3.199) sont respectivement symbolisés par $(1/\beta)$ et $(1/\theta)$. Les conditions d'optimalité des deux paramètres correspondent alors au respect des inégalités (les contraintes sont usuellement satisfaites en égalité, sauf dans des cas

extrêmes: voir 3.3.3). Etant donné que les deux fonctions d'entropie, appelons-les $S^T(\beta)$ et $S^P(\theta)$, sont toutes deux paramétrisées par β et θ , la procédure de recherche Newton-Raphson doit considérer les effets croisés des deux paramètres. En fait, l'itération principale concernera l'évaluation d'un vecteur incrémentiel:

$$\begin{bmatrix} \beta^{r+1} \\ \theta^{r+1} \end{bmatrix} = \begin{bmatrix} \beta^r \\ \theta^r \end{bmatrix} + \begin{bmatrix} S^T - S^T(\beta) \\ S^P - S^P(\theta) \end{bmatrix} \begin{bmatrix} \partial S^T(\beta)/\partial\beta & \partial S^T(\beta)/\partial\theta \\ \partial S^P(\theta)/\partial\beta & \partial S^P(\theta)/\partial\theta \end{bmatrix}^{-1} \quad (3.200)$$

Les dérivées sont approximées comme précédemment:

$$\frac{\partial S^T(\beta)}{\partial\beta} \approx \sum_{i,j} t_{i,j} \ln t_{i,j} \cdot (\gamma_{i,j} - \frac{1}{2}(\gamma_i + \gamma_j)) \quad (3.201)$$

$$\frac{\partial S^T(\beta)}{\partial\theta} \approx - \sum_{i,j} t_{i,j} \ln t_{i,j} \cdot (z_{i,j} - \frac{1}{2}(\bar{z}_i + \bar{z}_j)) \quad (3.202)$$

$$\frac{\partial S^P(\theta)}{\partial\beta} \approx \sum_{i,j} t_{i,j} \sum_m p_{T,j} \ln p_{T,j} \cdot (\gamma_{i,j} - \frac{1}{2}(\gamma_i + \gamma_j)) \quad (3.203)$$

$$\frac{\partial S^P(\theta)}{\partial\theta} \approx - \sum_{i,j} t_{i,j} \sum_m p_{T,j} \ln p_{T,j} \cdot (\gamma_{i,j} - \frac{1}{2}(\gamma_i + \gamma_j) - c_{T,j} + \bar{c}_{i,j}) \quad (3.204)$$

$$\begin{aligned} \text{où } z_{i,j} &= (\beta/\theta) \cdot (\gamma_{i,j} - \bar{c}_{i,j}) & \bar{c}_{i,j} &= \sum_m p_{T,j} c_{T,j} \\ \gamma_i &= \sum_j t_{i,j} \gamma_{i,j} / O_i & \gamma_j &= \sum_i t_{i,j} \gamma_{i,j} / D_j \\ \bar{z}_i &= \sum_j t_{i,j} z_{i,j} / O_i & \bar{z}_j &= \sum_i t_{i,j} z_{i,j} / D_j \end{aligned}$$

La solution du programme de minimisation (3.194-3.199) dépend essentiellement des multiplicateurs de Lagrange. Si on use de la forme lagrangienne de Y , soit L_V , on s'aperçoit que les conditions d'optimalité du 1^{er} degré des multiplicateurs $(1/\beta)$ et $(1/\theta)$ correspondent aux contraintes de dispersion et, incidemment, à l'approximation linéaire requise par la recherche paramétrique:

$$\frac{\partial L_V}{\partial(1/\beta)} = S^T + \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) = S_T - S^T(\beta) \quad (3.205)$$

$$\frac{\partial L_V}{\partial(1/\theta)} = S^P + \sum_{i,j} t_{i,j} \sum p_{T,j} (\ln p_{T,j} - 1) = S_P - S^P(\theta) \quad (3.206)$$

De même, la matrice de dérivées dans (3.200) coïncide avec la matrice Hessienne de dérivées secondes de L_V :

$$H = \begin{bmatrix} \partial^2 L_V / \partial(1/\beta)^2 & \partial^2 L_V / \partial(1/\beta) \partial(1/\theta) \\ \partial^2 L_V / \partial(1/\theta) \partial(1/\beta) & \partial^2 L_V / \partial(1/\theta)^2 \end{bmatrix} \quad (3.207)$$

Il est bien connu dans la théorie de l'optimisation que les conditions suffisantes pour un minimum local auprès de $(1/\beta)$ et de $(1/\theta)$ sont que la matrice Hessienne H soit négative semi-définie à $((1/\beta), (1/\theta))$. Cette condition rappelle que la fonction L_V doit être convexe i.e. que la magnitude des dérivées premières doit s'estomper à mesure de la proximité du point optimal.

3.8.4 Algorithme de solution du modèle intégré

Dans cette section, il sera question d'algorithme fournissant une solution au problème du trafic d'équilibre combiné avec une distribution émergeant de l'entropie maximale. D'abord, on traitera du cas générique conciliant l'affectation routière et la distribution spatiale pour ensuite l'étendre au cas comportant la répartition modale avec un des modes n'obéissant au principe de Wardrop.

Il y a deux algorithmes largement répandus aptes à trouver la solution du problème de distribution-affectation (3.129-3.131, 3.136-3.138). Tous deux appartiennent à la famille des «algorithmes de directions faisables». Ces méthodes conviennent à la solution de programmes non-linéaires dotés de contraintes linéaires. Leurs stratégies générales est de déterminer successivement la direction de la plus grande descente (ou ascension) à l'aide du gradient et, à chaque étape, de procéder à une recherche linéaire pour évaluer la longueur appropriée des incréments. L'algorithme qui sera d'abord présenté est dû à Florian, Nguyen et Ferland [1975] et est considéré comme une adaptation de la «méthode de Frank-Wolfe» [voir Frank, Wolfe, 1956]. La procédure alternative fut publiée par Evans [1976] et peut aussi être vue comme une généralisation de l'approche Frank-Wolfe [Boyce, 1984].

Sommairement, la méthode de Frank-Wolfe consiste à linéariser la fonction objective avec son gradient. La technique de génération de la direction repose sur la résolution du programme linéaire qui en est issu. Ayant déterminé la direction de l'incrément, il reste le calcul de la longueur de la progression. La méthode s'appuie sur la propriété de séparabilité d'un programme non-linéaire qui implique que la fonction objective peut être fragmentée en section linéaire. Ainsi, peut-on contruire un problème

approximant qui conserve les propriétés de convexité du problème initial [Simmons, 1975 chap.8].

Afin de permettre une présentation plus claire, il convient de réécrire le problème de minimisation (3.129-3.131, 3.136-3.138) qui sera étudié:

$$\min_{\{T, H^k\}} W = \sum_c \int_0^{v_c} s_c(v) dv + \frac{1}{\beta} \sum_{i,j} t_{i,j} (\ln t_{i,j} - 1) \quad (3.208)$$

$$\text{sujet à} \quad v_c = \sum_{i,j,k} \delta_{ik}^j s_k h_{i,j} \quad \sum_k h_{i,j} - t_{i,j} = 0 \quad (3.209-3.210)$$

$$h_{i,j} \geq 0 \quad \sum_j t_{i,j} = D_i \quad \sum_i t_{i,j} = D_j \quad \forall i,j \quad (3.211-3.213)$$

Les conditions de Kuhn-Tucker régissant les propriétés de la solution sont définies par (3.149-3.150). La première étape de l'algorithme de Florian et al. [1975] consiste à sélectionner une distribution initiale $\{t_{i,j}\}$ et une affectation initiale $\{h_{i,j}^k\}$ qui respectent l'ensemble des contraintes émises. Lorsque β est fixé, il s'agit, par exemple, d'évaluer la distribution spatiale selon le modèle gravitaire correspondant en utilisant des estimés raisonnables des coûts O-D. Quant à l'affectation routière, on peut soit allouer le trafic au plus court chemin (sans l'effet de congestion) ou soit octroyer une part égale à tous les itinéraires d'un même trajet $i \rightarrow j$.

La seconde étape s'avère fondamentale car elle concerne la linéarisation de la fonction objective. Considérons les solutions non-optimales courantes $\{t_{i,j}\}$ et $\{h_{i,j}^k\}$ de même que les coûts $\{s_k^j\}$ associés non en équilibre:

$$s_k^j = \sum_c \delta_{ik}^j s_c(v_c) \quad \text{où } v_c \text{ défini par (3.209)} \quad (3.214)$$

La fonction objective (3.208) linéarisée devient:

$$\min_{\{y_{i,j}, y_k^j\}} = \sum_{i,j,k} y_k^j s_k^j + \frac{1}{\beta} \sum_{i,j} y_{i,j} \ln t_{i,j} \quad (3.215)$$

$$\text{sujet à} \quad \sum_j y_{i,j} = D_i \quad \sum_i y_{i,j} = D_j \quad (3.216-3.217)$$

$$\sum_k y_k^j - y_{i,j} = 0 \quad y_k^j \geq 0 \quad (3.218-3.219)$$

Les termes $y_{i,j}$ et y_k^j sont les composantes du gradient de Y en respect de $t_{i,j}$ et de $h_{i,j}^k$ à leurs valeurs courantes. Ce programme linéaire est considérablement simplifié si on remplace s_k^j par la longueur du plus court chemin, soit $c_{i,j}$. Dans une solution quasi-optimale, tous les chemins

dont $s_{ij}^* > c_{ij}$ n'auront aucun trafic. Dans ce cas, $c_{ij} = c_{ij}^*$, celui-ci étant le coût à l'équilibre. Cette substitution affecte peu la fonction qu'est (3.215) car les y_{ij}^* correspondant aux s_{ij}^* largement différents de c_{ij} seront presque nuls. Cela permet de diminuer les h_{ij} dont le coût dépasse c_{ij} et d'augmenter celui lié au plus court chemin. En absorbant la contrainte de couplage (3.218), la direction de descente est alors donnée par la minimisation du problème linéaire suivant:

$$\min_{\{y_{ij}\}} \Omega = \sum_{i,j} y_{ij} (\ln t_{ij} / \beta + c_{ij}) \quad (3.220)$$

$$\text{sujet à (3.216-3.218) et à } y_{ij} \geq 0 \quad (3.221)$$

Ce programme linéaire est équivalent au problème de transport de Hitchcock (voir section 3.3.3). Une procédure très efficace et largement connue pour l'obtention d'une solution à ce problème reste la fameuse «méthode du Simplexe» [voir Taha, 1976]. Cette méthode exploite la dépendance algébrique d'un sous-ensemble de variables (appelées «basiques») sur les variables restantes («non-basiques»). La solution optimale d'un programme linéaire étant toujours une solution frontière, l'algorithme du Simplexe effectue des déplacements longeant les limites du plan-solution dressées par les contraintes.

La matrice $\{y_{ij}\}$ résolvant (3.216-3.217, 3.220-3.221) sert ensuite à définir $\{y_{ij}^*\}$ en assignant le flux O-D au parcours le plus court chemin. La direction de descente est alors obtenue de la façon suivante:

$$d_{ij} = y_{ij} - t_{ij} \quad \text{et} \quad d_{ij}^* = y_{ij}^* - h_{ij} \quad (3.222)$$

La troisième étape consiste à vérifier l'optimalité de la solution courante à l'aide d'un paramètre de convergence, ζ :

$$\sum_{i,j} (t_{ij} - y_{ij}) \cdot (\ln t_{ij} / \beta + c_{ij}) \begin{cases} \leq \zeta; \text{ stopper, la solution est optimale} \\ > \zeta; \text{ continuer} \end{cases} \quad (3.223)$$

Si la solution courante n'a pas atteint le critère d'optimalité, il faut chercher une solution améliorée. A ce stade-ci, la direction ayant le gradient maximal étant connue grâce au problème linéaire de transport (3.216-3.217, 3.220-3.221) il reste à trouver la longueur appropriée de l'incrément. Celui-ci doit être suffisant afin de progresser rapidement vers l'optimum mais pas trop grand pour prévenir les oscillations. La longueur du «pas» π est obtenue en solutionnant le problème d'optimisation

unidimensionnel:

$$\min_{\{\pi\}} \frac{1}{\beta} \sum_{i,j} \left[(1-\pi) t_{i,j} \ln[(1-\pi) t_{i,j}] + \pi y_{i,j} \ln(\pi y_{i,j}) \right] + \sum_c s_c \{ (1-\pi) v_c + \pi w_c \} \quad (0 \leq \pi \leq 1) \quad (3.224)$$

$$\text{où } v_c = \sum_{i,j,k} \delta x_i \delta y_j h_{i,j} k_{i,j} \quad w_c = \sum_{i,j,k} \delta x_i \delta y_j k_{i,j} \quad (3.225-3.226)$$

Une recherche lineaire univariee est seulement requise pour trouver π . La solution revisee devient:

$$\begin{aligned} t_{i,j}^{r+1} &= (1-\pi) t_{i,j}^r + \pi y_{i,j}^r \\ v_c^{r+1} &= (1-\pi) v_c^r + \pi w_c^r \end{aligned} \quad (3.227)$$

Une iteration est alors terminée et il faut poursuivre en retournant à la deuxième étape. La première validation empirique fut accomplie par Florian et Nguyen [1976] sur le réseau routier de Winnipeg de 1971. L'utilisation de cet algorithme adapté de la méthode de Frank-Wolfe paraît aujourd'hui plus répandue, étant installée dans le logiciel de transport EMME/2 [voir Babin et al., 1982] et la version corrigée de U.T.P.S. du "U.S. Department of Transportation".

En fait, l'algorithme d'Evans [1976] diffère peu de celui de Florian et al. [1975]. Il conserve une structure de type Frank-Wolfe à l'exception de l'étape déterminant la direction de descente. On ne linéarise que le terme référant aux fonctions d'impédance des liens plutôt que le problème complet. Par conséquent, la matrice gradient $\{y_{i,j}\}$ ne provient plus de la solution du problème linéaire de transport selon la valeur courante de $\{t_{i,j}\}$. Elle est déterminée par la solution du problème d'optimisation initial selon la matrice courante de coût minimal $\{c_{i,j}\}$. En fait, le programme (3.220-3.221, 3.216-3.218) est remplacé par celui-ci:

$$\min_{\{y_{i,j}\}} \Omega = \sum_{i,j} y_{i,j} (\ln y_{i,j} / \beta + c_{i,j}) \quad (3.228)$$

$$\text{sujet à (3.216-3.218) et à } y_{i,j} \geq 0 \quad (3.229)$$

Les increments $y_{i,j}$ sont donc calculés avec le modèle de distribution. LeBlanc et Farhangian [1981] affirment que la technique d'Evans donne des directions de descente plus appropriées, la méthode de calcul permettant de moduler davantage les incréments selon la paire O-D. Quant aux incréments issus du problème linéaire, les fréquentes valeurs nulles côtoyant des valeurs élevées provoquent certaines oscillations. En effet, un incrément

nul ne signifie pas la fixité de la valeur courante puisque les contraintes marginales impliquent des réajustements de toutes les valeurs en fonction de ceux qui augmentent.

Il convient d'adopter l'une des deux procédures algorithmiques à un problème multi-modal tel que le programme (3.144-3.151). L'ampleur du problème considéré est telle qu'il s'impose de le réduire. Il faut aussi tenir compte que les paramètres sont d'abord inconnus; le schéma général de solution comprendra alors une phase de calibration. Boyce et al. [1983] démontrèrent qu'il est possible d'insérer entre chaque solution algorithmique, une étape comportant une procédure de Newton-Raphson; celle-ci met à jour les paramètres courants mais détruit le caractère optimal de la solution actuelle. Il faut par conséquent reobtenir une nouvelle solution avec les paramètres corrigés jusqu'à ce que la solution, ni les paramètres se modifient. En pratique, il convient de donner des valeurs paramétriques initiales vraisemblables ou proches des valeurs vraies afin de diminuer les recherches de la solution et des paramètres.

Dans un premier temps, on réduira le problème intégral en écartant l'affectation des itinéraires pour le mode collectif. On suppose alors que le paramètre ρ est connu; si tel est le cas, les résultats de l'affectation n'ont aucune incidence sur le problème principal car il ne concerne que des proportions et il n'y a pas de relation coût-traffic pour le transport public. Les deux derniers termes de la fonction objective (3.144) disparaissent de même que la contrainte (3.148). Les conditions d'optimalité devaient toutefois une mesure de coût composé pour le mode 2 qu'emploie la fonction de partage modal (3.157). En définissant:

$$c_{ij}^2 = \frac{1}{-\rho} \ln \sum_n \exp(-\rho z_{ij}^n) + \frac{1}{\rho} \ln N_{ij} \quad (3.230)$$

où z_{ij}^n : coût de transport $i \rightarrow j$ sur itinéraires n et N_{ij} : nombre de chemins admissibles. On peut reformuler le problème (3.144-3.151) ainsi:

$$\begin{aligned} \min_{\{T, H^k, P\}} \quad & \sum_i \int_0^{v_i} s_i(v) dv + \frac{1}{\beta} \sum_{i,j} t_{ij} \ln(t_{ij}-1) + \\ & \frac{1}{\theta} \sum_{i,j} t_{ij} \sum_m \rho T_{ij}^m (\ln \rho T_{ij}^m - 1) + \sum_{i,j} t_{ij} \rho \bar{f}_{ij} c_{ij}^2 \end{aligned} \quad (3.231)$$

sujet à (3.145-3.147, 3.149-3.151)

La linéarisation de la fonction objective demande la définition de la fonction gradient en respect des vecteurs implicites. Tout comme l'ont fait Florian et al. [1975], on écartera les flux sur chemins $\{h_{ij}^k\}$ en recourant à la mesure du plus court chemin. La fonction objective linéarisée devient:

$$\sum_{i,j} \left[\sum_m p_{T_j} c_{T_j} + \ln t_{i,j} / \beta + \left(\sum_m p_{T_j} (\ln p_{T_j} - 1) \right) / \theta \right] \cdot y_{i,j} + \sum_{i,j,m} \left[t_{i,j} c_{T_j} + (t_{i,j} \ln p_{T_j}) / \theta \right] x_{T_j} \quad (3.232)$$

En prenant l'équation de partage modal (3.157), on s'aperçoit que le terme du bas se réduit à:

$$\sum_{i,j,m} [\gamma_{i,j} \cdot x_{T_j}] \quad \text{où } \gamma_{i,j} \text{ défini par (3.158)} \quad (3.233)$$

Puisque x_{T_j} est contraint par $\sum_m x_{T_j} = 1, \forall i,j$, ce terme devient sans pertinence. Le nouveau programme linéaire à résoudre est:

$$\min_{\{y_{i,j}\}} \sum_{i,j} \Omega_{i,j} y_{i,j} \quad (3.234)$$

$$\text{où } \Omega_{i,j} = \sum_m p_{T_j} c_{T_j} + \frac{1}{\beta} \ln t_{i,j} + \frac{1}{\theta} \sum_m p_{T_j} (\ln p_{T_j} - 1)$$

sujet à (3.216-3.217, 3.221) et où

$c_{i,j}$: plus court chemin courant parmi $k \in K(i,j)$ et c_{T_j} défini par (3.230)

A l'exception de la variable $y_{i,j}$, les variables ont les valeurs de la solution courante. Le problème devient comme prévu un programme linéaire de transport soluble par l'algorithme du Simplexe. Sa solution permet d'obtenir la direction de descente pour $t_{i,j}$. La répartition modale a pu être évincée du problème linéarisé parce qu'elle est contenue implicitement dans le terme $\Omega_{i,j}$. La direction de descente de $t_{i,j}$ tient alors compte des volumes accordés à chacun des modes. Le partage est effectué selon la solution courante (c_{T_j} fixe et $c_{i,j}$, coût du plus court chemin $i \rightarrow j$) et à l'aide du modèle logit défini par (3.157). La variable $y_{i,j}^k$ est construite telle que $y_{i,j}^k = \gamma_{ij} p_{i,j}^k$, si $s_{i,j}^k = c_{i,j}$; $y_{i,j}^k = 0$, sinon.

La longueur optimale de l'incrément reste encore déterminée par la solution d'un problème de minimisation unidimensionnelle:

$$\min_{\{\pi\}} \frac{1}{\beta} \sum_{i,j} \left[(1-\pi) t_{i,j} \ln[(1-\pi) t_{i,j}] + \pi y_{i,j} \ln(\pi y_{i,j}) \right] + \quad (3.235)$$

$$\sum_{i,j} G_{i,j} \{ (1-\pi) t_{i,j} + \pi y_{i,j} \} + \sum_k s_k \{ (1-\pi) v_k + \pi w_k \} \quad (0 \leq \pi \leq 1)$$

où v_k et w_k définis par (3.225-3.226) et

$$G_{i,j} = \frac{1}{\theta} \sum p_{T,j} \ln(p_{T,j} - 1) + p_{T,j}^* c_{T,j}^*$$

La solution améliorée apparaît comme:

$$\begin{aligned} t_{i,j}^{r+1} &= t_{i,j}^r (1-\pi) + y_{i,j}^r \pi \\ v_k^{r+1} &= v_k^r (1-\pi) + y_{i,j}^r \pi \end{aligned} \quad (3.236)$$

La procédure itérative est conduite de la façon décrite précédemment pour le problème de distribution - affectation. Afin de calibrer proprement β et θ (p étant presumé connu), il convient d'adopter le schème de Boyce et al. [1983] dans lequel l'algorithme de solution est exécuté en alternance avec une procédure de Newton-Raphson visant à respecter des contraintes d'entropie. Ici, la démarche proposée dans la section 3.8.3 semble adéquate si le coût d'équilibre $c_{i,j}^*$ (fourni par une solution de la méthode Frank-Wolfe) et le coût composé $c_{T,j}^*$ sont considérés des coûts modaux associés respectivement aux voyages en automobile et en transport en commun. On voit très bien l'avantage des contraintes d'entropie sur celles de coût total dans la calibration paramétrique lorsque la mesure de coût inter-zonal est un élément variable du problème [Erlander, 1977].

Si nous adoptons l'algorithme d'Evans plutôt que celui de Frank-Wolfe, nous devons redéfinir la fonction objective du sous-problème (3.232). Encore une fois, les différences de forme restent minimales. Ainsi:

$$\min_{\{y_{i,j}\}} \sum_{i,j} y_{i,j} \left[\sum_m p_{T,j} c_{T,j} + \frac{1}{\beta} \ln y_{i,j} + \frac{1}{\theta} \sum_m p_{T,j} (\ln p_{T,j} - 1) \right] \quad (3.237)$$

A chaque itération, la direction de descente $y_{i,j}$ est générée par le modèle de distribution spatiale doté du coût composé $\chi_{i,j}$ de la solution courante. Les proportions modales sont quant à elles calculées avec la fonction de partage issue du problème d'optimisation original.

3.A APPENDICE A: Dérivation du modèle dogit de répartition modale à partir d'un programme mathématique

Le modèle dogit tel que formule aux équations 3.87a et 3.87b possède un programme mathématique équivalent. Celui-ci doit tenir compte autant de la probabilité d'être captif d'un mode que celle d'en choisir un librement. Soit la forme générale initialement proposée:

$$p_{T^n} = (1 - \sum_m \sigma_{mn}) \cdot q_{T^n} + \sigma_{mn} \quad (3A.1)$$

où σ_{mn} : prob. des voyageurs n d'être contraint à utiliser m ; $(1 - \sum_m \sigma_{mn})$: prob. des voyageurs n de n'être captif d'aucun mode; q_{T^n} : prob. de choisir m si on est voyageur non-captif de type n provenant de i .

De même, considérons une variable x_{mn} , caractéristique de la catégorie n et favorisant l'usage exclusif du mode m . La seule maximisation de cette caractéristique ne suffit pas si on ne lui adjoint pas une fonction d'entropie présentant la variabilité courante des comportements. Par exemple, la non-propriété d'une automobile n'exclut pas le voyageur de l'usage de l'auto soit en tant que passager, soit en usage locatif. Voici un programme mathématique correspondant:

$$\begin{aligned} \max_{\{\sigma, Q_1\}} & -\frac{1}{\delta} \sum_{mn} \sigma_{mn} (\ln \sigma_{mn} - 1) - \frac{1}{\delta} \sum_n (1 - \sum_m \sigma_{mn}) (\ln (1 - \sum_m \sigma_{mn}) - 1) + \sum_{mn} \sigma_{mn} x_{mn} \\ & - \frac{1}{\theta} \sum_{imn} q_{T^n} (\ln q_{T^n} - 1) + \sum_{imn} q_{T^n} u_{T^n} \end{aligned} \quad (3A.2)$$

$$\text{sujet à } (3A.1) \text{ et à } \sum_m q_{T^n} = 1 \quad \forall i, n \quad \sum_m \sigma_{mn} \leq 1 \quad \forall n \quad (3A.3-3A.4)$$

Le premier terme désigne l'entropie contenue dans la série $\sigma = \{\sigma_{mn}\}$ tandis que le second exprime celle incluse dans la partie restante (non-captive). La distribution solution de $Q_1 = \{q_{T^n}\}$ prend la forme logistique usuelle:

$$q_{T^n} = \frac{\exp(\theta u_{T^n})}{\sum_m \exp(\theta u_{T^n})} \quad (3A.5)$$

Quant à la proportion de captivité, les conditions d'optimalité révèlent:

$$\sigma_{mn} = (1 - \sum_m \sigma_{mn}) \cdot \exp(\delta x_{mn})$$

soit après substitutions:

$$\begin{aligned}
 (1 - \sum_m \sigma^{mn}) &= \frac{1}{1 + \sum_m \exp(\delta x^{mn})} \\
 \sigma^{mn} &= \frac{\exp(\delta x^{mn})}{1 + \sum_m \exp(\delta x^{mn})} \quad (3A.6)
 \end{aligned}$$

Les niveaux de captivité émanent ici d'une fonction logit consistante avec la distribution du choix libre (3A.5). Selon la forme présumée (3A.1), nous obtenons pour le terme de probabilité totale p_T^n le modèle ci-dessus:

$$p_T^n = \frac{1}{1 + \sum_m \exp(\delta x^{mn})} \cdot \frac{\exp(\theta u_T^n)}{\sum_m \exp(\theta u_T^n)} + \frac{\exp(\delta x^{nn})}{1 + \sum_m \exp(\delta x^{mn})} \quad (3A.7)$$

On notera facilement l'équivalence avec le modèle dogit (3.87b) si l'on fixe $\zeta^{nn} = \exp(\delta x^{nn})$. On a alors une fonction pouvant générer les paramètres de captivité.

3.B APPENDICE B: Formulation du problème d'affectation routière avec agrégation du réseau

L'utilisation d'un modèle d'affectation par trafic d'équilibre requiert un réseau convenablement codé [voir Potts, Oliver, 1972] où l'on connaît, pour chaque lien, le temps minimal de parcours, la capacité technique et la fonction de congestion appropriée. Ces caractéristiques semblent plus facilement évaluables pour les voies importantes comme les autoroutes et les grandes artères. Pourtant, il semble raisonnable qu'une part appréciable des voyageurs opte pour des itinéraires particuliers qui ne sollicitent pas des voies importantes. L'énumération des chemins secondaires reste ardue car ils sont nombreux dans une grille routière urbaine serrée.

On peut sans doute bénéficier des possibilités moindres de congestion sur les routes mineures en considérant les coûts sur celles-ci comme fixes. La fixité présumée des coûts sur chemins mineurs permet d'éviter l'affectation par équilibre et de résumer l'ensemble de ces chemins par un seul itinéraire dit «auxiliaire». On peut également envisager une double affectation, l'une entre les itinéraires principaux et l'itinéraire auxiliaire, l'autre parmi les chemins constitués des grands axes. La repartition entre les itinéraires congestionnées (en général plus efficaces), d'un côté, et l'itinéraire alternatif, de l'autre, peut être dotée d'une certaine variabilité statistique.

En confrontant les itinéraires ordinaires avec l'itinéraire artificiel, on obtiendra sans doute une affectation d'équilibre entre les deux grands types de chemin [Brice, 1989b]. En effet, le chemin auxiliaire ne diffère pas réellement d'un itinéraire régulier, mis à part son insoumission à la congestion. Le problème d'affectation avec agrégation mais sans composante aléatoire peut se formuler ainsi:

$$\min_{(h,g,H^k)} \sum_c \int_0^{v_c} s_c(v) dv + \sum_{i,j} G_{i,j} z_{i,j} \quad (3B.1)$$

$$v_c = \sum_{i,j,k} \delta_{i,j}^k h_{i,j}^k \quad (3B.2)$$

$$\sum_k h_{i,j}^k - H_{i,j} = 0 \quad G_{i,j} + H_{i,j} = t_{i,j} \quad (3B.3-3B.4)$$

$$G_{i,j} \geq 0 \quad H_{i,j} \geq 0 \quad h_{i,j}^k \geq 0 \quad (3B.5-3B.7)$$

où $h = \{H_{ij}\}$; $g = \{G_{ij}\}$; $H^* = \{h_{ij}^k\}$; H_{ij} : trafic $i \rightarrow j$ empruntant le réseau majeur; G_{ij} : trafic $i \rightarrow j$ empruntant le réseau mineur; h_{ij}^k : flux sur chemin $k \in K(ij)$; c_{ij} : coût fixe du chemin auxiliaire.

Il y a deux affectations, l'une entre les deux réseaux, soit entre G_{ij} et H_{ij} , l'autre parmi le trafic H_{ij} , soit entre $\{h_{ij}^k; k=1,2,\dots\}$. Les conditions d'optimalité réitérent une affectation d'équilibre quelque soit la priorité de l'affectation:

$$\sum_c \delta x_c^j s_c(v_c) - g_{ij} \geq 0; \quad h_{ij}^k \geq 0 \quad (3B.8)$$

$$g_{ij} - \lambda_{ij} \geq 0; \quad H_{ij} \geq 0 \quad c_{ij} - \lambda_{ij} \geq 0; \quad G_{ij} \geq 0 \quad (3B.9a)$$

$$\text{ou } g_{ij} - c_{ij} \begin{cases} < 0; & H_{ij} = t_{ij} \geq 0; & G_{ij} = 0 \\ > 0; & H_{ij} = 0; & G_{ij} = t_{ij} \geq 0 \end{cases} \quad (3B.9b)$$

En fait, il n'y a pas lieu de distinguer les deux niveaux de partage puisque les itinéraires agrégés sont traités sur le même pied que les autres.

L'introduction d'une composante aléatoire à chacun des deux mesures de coût de transport, permettra de nuancer la répartition prioritaire entre le réseau des grandes artères et celui des routes mineures. Brice [1989b] propose d'insérer une composante stochastique au sein des coûts pour représenter les erreurs de mesure accompagnant la détermination de coûts fixes et la variabilité perceptuelle des voyageurs. Par conséquent, le problème se réécrit comme suit:

$$\min_{\{h,g,H^*\}} \sum_c \int_0^{v_c} s_c(v) dv + \sum_{ij} H_{ij} e x_{ij}^* + \sum_{ij} G_{ij} (c_{ij} + e c_{ij}^*) \quad (3B.10)$$

sujet à (3B.2-3B.7)

Présumons une distribution de probabilité de type «Gumbel» [voir Gumbel, 1958] pour $e x_{ij}^*$ et $e c_{ij}^*$. La fonction de densité arbore une allure en cloche. Selon sa structure paramétrique, elle peut s'apparenter à la distribution «Weibull» qui a servi pour dériver le logit multinomial dans la section 3.4.3. La fonction de probabilité cumulative s'écrit:

$$\text{Prob}(e_* < e_k) = \bar{g}(e_k) = \exp\{-\exp(-\mu e_k)\} \quad (3B.11)$$

Alors que la fonction de densité est:

$$\varphi(\varepsilon_k) = \partial \tilde{g} / \partial \varepsilon_k = \mu \exp(-\mu \varepsilon_k) \exp\{-\exp(-\mu \varepsilon_k)\} \quad (3B.12)$$

Ordinairement, les dérivations découlant de ces distributions se font au niveau d'un individu. Le problème rencontré ici porte sur un grand nombre d'individus et ne peut guère présumer une simple additivité des comportements individuels. La dérivation s'appuie sur un constat formulé par Cochrane [1975]. Supposons que chaque voyageur possède sa propre distribution d'erreur. Selon certaines hypothèses rationnelles, les voyageurs optent pour ce qu'ils croient les meilleures alternatives. Plus ils choisissent une alternative spécifique, moins il est probable que les autres choix sont meilleurs. D'un point de vue statistique, l'extrémité supérieure de la fonction cumulative s'étend asymptotiquement avec la fréquence de choix [Cochrane, 1975]. La fonction de N_k échantillons indépendants devient la probabilité conjointe de ceux-ci. Alors:

$$\begin{aligned} \varphi(\varepsilon_k) &= [\tilde{g}(\varepsilon_k)]^{N_k} \quad \forall k \\ &= \exp\{-N_k \exp(-\mu \varepsilon_k)\} \end{aligned} \quad (3B.13)$$

Le nombre N_k n'affecte pas l'étendue de la distribution [McFadden, 1978] mais déplace vers la droite la fonction de densité. Seul le paramètre μ est inversement relié à l'écart-type de l'erreur aléatoire [Williams, 1977]. L'espérance statistique de ε_k est:

$$\begin{aligned} E(\varepsilon_k) &= \int_{-\infty}^{\infty} \varepsilon_k \cdot \varphi(\varepsilon_k) \, d\varepsilon_k \\ &= \int_{-\infty}^{\infty} \varepsilon_k \cdot \mu N_k \exp(-\mu \varepsilon_k) \exp\{-N_k \exp(-\mu \varepsilon_k)\} \, d\varepsilon_k \\ &= \frac{1}{\mu} \ln(N_k) + \frac{\gamma}{\mu} \end{aligned} \quad (3B.14)$$

où γ est la constante d'Euler qui égale approximativement 0.5772.

Etant donné que la fonction objective (3B.10) concerne un grand nombre de gens, on peut substituer aux termes d'erreur leurs valeurs espérées:

$$E(\varepsilon_{1j}) = \ln H_{1j} / \mu + \gamma / \mu \quad E(\varepsilon_{2j}) = \ln G_{1j} / \mu + \gamma / \mu \quad (3B.15)$$

$$\begin{aligned} \min_{(h,g,H^k)} \quad & \sum_i \int_0^{V_i} s_i(v) \, dv + \sum_{1j} H_{1j} \left[\ln H_{1j} / \mu + \gamma / \mu \right] + \\ & \sum_{1j} G_{1j} \left[\ln G_{1j} / \mu + \gamma / \mu \right] \end{aligned} \quad (3B.16)$$

sujet à (3B.2-3B.7)

Le lecteur averti notera que l'addition de termes aléatoires aux coûts de transport est équivalente à l'introduction de termes d'entropie à la fonction objective. Le problème (3B.2-3B.7, 3B.16) reste convexe et résolu lorsque:

$$\sum_k \delta h_{kj} s_k(v_k) - g_{1j} \geq 0; \quad h_{1j} \geq 0 \quad (3B.17)$$

$$\frac{1}{\mu} \ln h_{1j} + \frac{(1+\gamma)}{\mu} + g_{1j} - \lambda_{1j} = 0; \quad h_{1j} > 0 \quad (3B.18)$$

$$\frac{1}{\mu} \ln G_{1j} + \frac{(1+\gamma)}{\mu} + z_{1j} - \lambda_{1j} = 0; \quad h_{1j} > 0 \quad (3B.19)$$

Pour satisfaire la contrainte (3B.4), le multiplicateur lagrangien λ_{1j} doit être égal à:

$$\lambda_{1j} = \frac{1}{\mu} \ln t_{1j} - \frac{1}{\mu} \ln [\exp(-\mu g_{1j} + 1 + \gamma) + \exp(-\mu z_{1j} + 1 + \gamma)] \quad (3B.20)$$

Alors:

$$h_{1j} = \frac{t_{1j} \exp(-\mu g_{1j})}{\exp(-\mu g_{1j}) + \exp(\mu z_{1j})} \quad G_{1j} = \frac{t_{1j} \exp(-\mu z_{1j})}{\exp(-\mu g_{1j}) + \exp(\mu z_{1j})} \quad (3B.21)$$

Les conditions de Kuhn-Tucker révèlent qu'un trafic d'équilibre pour $\{h_{1j}\}$ est encore obtenu alors que la répartition $G_{1j}-H_{1j}$ est accomplie par une expression logit binômiale dont le paramètre est relié à l'écart-type des termes stochastiques. Les variables discriminantes restent le coût fixe du chemin auxiliaire et le coût d'équilibre issu du réseau agrégé.

Lorsque que le coût d'équilibre devient élevé dans des conditions de trafic dense, le nombre de voyageurs privilégiant les nombreux chemins secondaires s'accroît en conséquence. Une mesure globale du trajet $i \rightarrow j$ est offerte en recourant au formalisme de la programmation mathématique:

$$c_{1j} = \frac{1}{-\mu} \ln [\exp(-\mu g_{1j}) + \exp(-\mu z_{1j})] \quad (3B.22)$$

L'avantage de ces agrégations réside dans la spécification simplifiée du réseau. Il n'est dorénavant plus nécessaire de coder tous les liens d'un réseau mais seulement les grands axes. On identifie alors les itinéraires k d'un trajet O-D requérant ces arcs et on ajoute, pour chaque trajet, un parcours auxiliaire imaginaire résumant l'ensemble des chemins alternatifs. Il paraît évident qu'une mesure adéquate du coût enregistré sur le chemin auxiliaire doit être faite; elle tient compte des conditions normales de circulation rencontrées sur les meilleurs parcours secondaires.

Chapitre 4

MODELISATION DE L'OFFRE DE TRANSPORT

4.1 CONCEPTUALISATION

4.1.1 Définition de l'offre de transport

L'appellation «offre de transport» semble simplement justifiée par opposition à la demande de transport. Selon Florian et Gaudry [1980], on peut difficilement assimiler cette notion d'offre au concept économique de «fonction d'offre» qui spécifie la quantité d'un certain produit qu'une firme, dans des conditions de compétition parfaite, est disposée à produire à un certain prix. Il existe maintes différences entre le domaine du transport et ceux concernés par les biens et services communs.

L'offre de transport se présente avant tout comme un ensemble de services permettant aux individus de se déplacer efficacement d'un lieu à un autre. Dans une agglomération urbaine, il y a usuellement plus d'un service de transport; certains sont collectifs (autobus, métro, train), d'autres sont privés (véhicule-moteur, bicyclette). Dans tous les cas, l'offre de transport dépend, au moins en partie, d'un palier gouvernemental. Les moyens dits privés empruntent un réseau routier construit et planifié par un gouvernement tandis que les moyens collectifs sont dispensés, opérés et organisés par un organisme de transport (habituellement sans but lucratif).

L'offre peut être scrutée sous divers angles. Manheim [1980] souligne les aspects tangibles de l'offre qui peuvent être modifiés. Voici divers options qui se présentent à un planificateur lorsque qu'il veut améliorer le service.

- Changement technologique: pour le transport public; métro souterrain, train, voie en site propre. Pour le réseau routier; feux synchronisés, chaussées séparées, géométrie des carrefours.

- Modification du réseau: amélioration de la connectivité, addition de liens (route ou ligne).

- Modification des liens: augmentation de la capacité et de la vitesse (au niveau des routes ou des lignes de transport public).
- Véhicules: renouvellement de la flotte, amélioration des performances (accélération, vitesse de pointe).
- Politiques d'opération et d'organisation: fréquences de services, espacement des véhicules, correspondances, horaires, lignes "express", structure de tarification etc...

Ces aspects tangibles de l'offre ne sont pas perçus directement par l'utilisateur. Selon sa perspective, l'offre de transport se mesure par le niveau de service sur les trajets origines-destinations qu'il effectue. Ainsi, la performance globale d'un système s'apprécie par une somme des performances de transport sur chacun des parcours O-D. L'utilisateur ne s'intéresse pas aux dépenses occasionnées par l'ajout ou le maintien d'un service. Il regarde plutôt les éléments du service lui affectant: temps d'accès, d'attente, de transport et de correspondances, tarification, péage, coût du carburant etc... Outre la moyenne de ces mesures, il tient compte des écarts maximaux. En fait, la mesure généralisée du «coût de transport», telle qu'employée dans la modélisation de la demande, synthétise non seulement la «desutilité» d'un déplacement mais la qualité d'un service de transport.

La performance du système, selon une perspective de l'opérateur, diffère dans sa définition. L'organisme responsable du transport apprécie certes le niveau de service fourni aux voyageurs. Cependant, ses préoccupations tiennent avant tout aux aspects tangibles de l'offre car ce sont eux à qui les ressources sont directement affectées. Il n'en reste pas moins que la justesse de l'affectation des ressources s'évaluera en fonction du niveau de service global. En fait, le coût généralisé de transport (pour le voyageur) par trajet O-D reste encore la mesure synthétique la plus appropriée de l'offre. La stratégie de l'opérateur («comportement de l'offre») est de fournir un produit selon ses capacités financières qui maximisera le niveau de service pour le plus grand nombre d'utilisateurs. Certains poursuivent d'autres objectifs comme la maximisation du nombre d'utilisateur sachant que l'affluence est garante de la qualité du service tandis que les organismes privés visent la profitabilité.

Le comportement d'offre le plus approprié demeure quand même la maximisation du niveau de service car il implique sans doute un accroissement de l'achalandage (pour les transports collectifs) ou une baisse de la congestion (pour le transport automobile). Quant à convoiter un maximum de voyageurs, cela semble un but évidemment inadéquat dans le cas des véhicules privés; pour le transport en commun, c'est un objectif incorrect car il peut signifier un accroissement du nombre de déplacements. Enfin, viser la rentabilité paraît un objectif louable mais parfois inatteignable pour un organisme mandaté pour desservir l'ensemble de la population citadine. Le maintien d'une marge de profit peut vouloir dire l'abandon du service dans certaines zones peu denses et l'interruption à certaines heures creuses.

Si le coût de transport se veut une mesure agrégée de l'offre de transport, il n'en reste pas moins qu'il peut difficilement rendre compte de plusieurs composantes du service. La complexité du transport se présente d'une part à travers les caractéristiques non-monnaïres du service. Les temps non reliés au trajet proprement dit comme le temps d'accès, d'attente et de transfert ne peuvent être fixés; ils dépendent de chacun des individus, de leurs sites d'origines et de destinations, de la température (par mauvais temps, ces temps paraissent plus longs). La vitesse des véhicules, autos comme autobus, décroît selon le volume courant de circulation. Les voyageurs, dans l'appréciation du service, tiennent compte de la fiabilité du temps total de transport. Des composantes comme le confort et la sécurité évoluent aussi et ne peuvent être jugées comme fixes. Par exemple, la probabilité de trouver un siège se mesure difficilement car elle décroît avec l'affluence.

Manheim [1980] met en relief les attributs de l'offre qui échappent en partie à l'opérateur. Une part de ceux-là tiennent aux réactions de la demande: la congestion des voies routières, la probabilité de trouver une place assise. Néanmoins, le fait que la demande réagit positivement à une augmentation de la qualité de l'offre (diminution du coût généralisé de transport) couplé au fait que l'offre se détériore quand la demande est excessive, crée une situation «d'équilibre de marché». Cette situation s'apparente ici à celle reconnue dans le mécanisme bien connu de l'offre et de la demande. La comparaison s'arrête là. Il ne s'agit plus de quantité offerte à un prix établi et dans un contexte concurrentiel; le transport

est un service non-substituable dans un marché très peu compétitif, fortement désagrégée géographiquement et dont le véritable prix se compose majoritairement d'éléments non-monétaires. Pour ajouter à la dissemblance, le transport est un bien périssable si non consommé lorsqu'offert [Manheim, 1980].

4.1.2 Un cadre conceptuel pour l'offre de transport urbain

Dans cette section, un cadre conceptuel inspiré de celui dressé par Florian et Gaudry [1980] sera adapté au contexte du transport urbain. Ce cadre établit les relations principales que sous-tendent les mécanismes d'ajustement de l'offre. Ces mécanismes peuvent être repris dans un programme mathématique et, par le fait même, s'inscrire dans un «modèle d'offre».

Florian et Gaudry reitère la dichotomie classique de l'offre et de la demande tout en reconnaissant que le comportement de l'offre peut difficilement être décrit sans référence à la demande. Dans le chapitre précédent, nous avons pu présenter un modèle intégré de la demande en résumant les attributs de l'offre par le coût de transport désagrégé selon le trajet O-D, le mode ou l'itinéraire. A l'égard de l'affectation routière, on admettait néanmoins une variabilité de l'offre à travers une relation coût-traffic bien que les effets de congestion ne relèvent pas directement des instigateurs de l'offre.

Vu du côté de l'offre, le comportement de la demande est davantage pris en compte. Dans l'évaluation du niveau de service, il importe de mesurer, outre les caractéristiques du transport, le nombre de voyageurs bénéficiant de celles-ci. Par exemple, la simple somme:

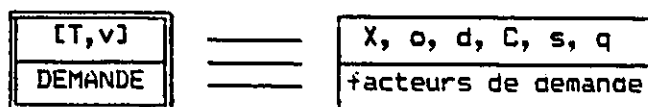
$$\sum_{i,j} c_{ij}$$

renseigne beaucoup moins sur la performance globale du transport que:

$$\sum_{i,j} t_{ij} c_{ij}$$

où c_{ij} est le coût généralisé de transport pour un trajet $i-j$ et selon un mode m tandis que t_{ij} est le nombre de voyageurs empruntant un mode m pour effectuer un déplacement $i-j$.

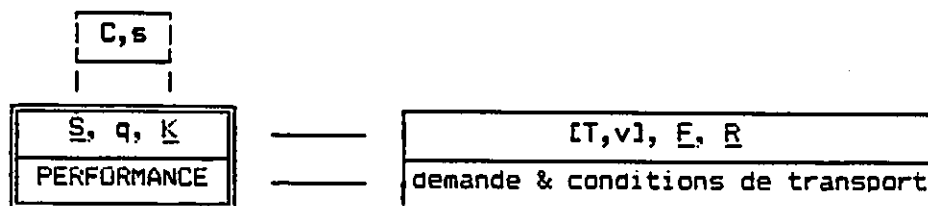
Schématiquement, la demande est modulée par un certain nombre de facteurs bien identifiés. Ici, nous préconiserons une notation apparentée à celle employée dans la modélisation de la demande. La demande origine-destination, par mode et par chemin est symbolisée par une matrice T tandis que le vecteur v désigne les flux sur les éléments des réseaux. Alors:



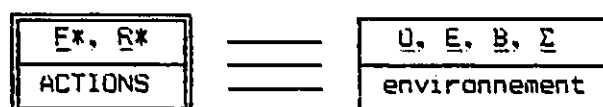
$$\begin{aligned}
 T &\equiv \{hT_j\} \equiv \{tT_j\} \equiv \{t_{ij}\} & v &\equiv \{v_\ell\} \quad \ell \in L \\
 X &\equiv \{X_1, X_2, \dots\}, \quad X_o = \{X_{o1}\} \quad N-1 & C &\equiv \{cT_j\} \equiv \{cT_j\} \equiv \{c_{ij}\} \\
 o &\equiv \{O_i\} \quad N-1 & d &\equiv \{D_j\} \quad N-1 \\
 s &\equiv \{s_\ell(v_\ell)\} \quad \ell \in L & q &\equiv \{q_\ell\} \quad \ell \in L
 \end{aligned}$$

Le matrice X symbolise les activités économiques de divers types (résidentiel, commercial, industriel etc...) pour chaque zone. Les vecteurs o et d représentent la demande générée à l'origine et à la destination. Dans la matrice C , les éléments fixes du coût de transport y sont inscrits alors que les vecteurs s et q évoquent, pour chaque liens de tous les réseaux, la fonction de congestion et la capacité. Le signe entre la demande et ses facteurs n'est pas celui d'une égalité ni même d'une équivalence; il signifie que pour un ensemble donné de caractéristiques de réseau s et q , une distribution des coûts C et une repartition d'activités X générant une mobilité o et d , on a une demande correspondante $[T, v]$.

Poursuivant le schéma de Florian et Gaudry, on passe du côté de l'offre où l'on distingue deux aspects, la «performance» et l'«action». Le volet de la performance reflète l'appréciation globale du service vue par les instigateurs de l'offre. Pour une demande $[T, v]$, une structure de tarification E et des règles d'exploitation R , on obtient une performance résultante définie par le niveau de service $\underline{S}$, des capacités de voyageurs q et des tarifs $\underline{K}$. Ainsi:



Le niveau de service $\underline{S}$ et les tarifs $\underline{K}$ sont résumés par les coûts de transport $\underline{C}$ spatialisés et ses composantes variables sur le réseau (s). La demande $[T, v]$ échappe à l'exploitant dont les pouvoirs se restreignent à la structure de tarification et aux règles d'opération (organisations des réseaux, fréquences). Par conséquent, les «actions» possibles sur l'offre ne procèdent qu'indirectement au moyen de $\underline{F}$ et de $\underline{R}$. Ceux-ci incluent l'ensemble des options comme les caractéristiques technologiques, des réseaux, des liens, des véhicules et les politiques d'opération ainsi que tous les éléments de la tarification (si une telle existe). Ces options émanent d'un ensemble d'objectifs $\underline{O}$, d'un environnement institutionnel $\underline{E}$, des ressources disponibles $\underline{B}$ et d'une évaluation du système actuel ou projeté $\underline{\Sigma}$. Ainsi:



$$\underline{\Sigma} \equiv f(\underline{S}, q, \underline{K}, [T, v])$$

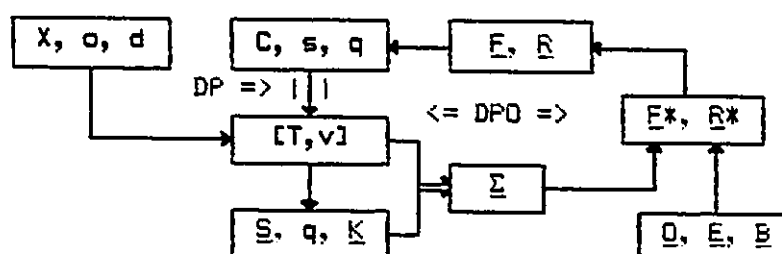
$\underline{O}$ désigne les objectifs du système (minimiser le total des coûts de transport, maximiser les profits, maximiser la clientèle). Les responsables de l'offre sont assujettis à des ressources limitées (budget, personnel) et doivent composer avec une réglementation issue d'un environnement institutionnel $\underline{E}$. Les actions $\underline{F}^*$ et $\underline{R}^*$ dépendent d'un état du système actuel (reconnaissance de lacunes) ou projeté (état désiré); des mesures brutes de la performance composent $\underline{\Sigma}$ ainsi qu'une évaluation de la demande.

Florian et Gaudry insistent sur le fait que les nombreuses interactions entre les aspects occasionnent divers types d'équilibre. Un équilibrage demande - performance survient lorsque la demande du transport donnent lieu à des niveaux de service et des tarifs qui maintiennent cette demande. Ce type d'équilibre s'opère à court terme. L'exemple le plus évident reste certes celui de la répartition des automobilistes sur plusieurs chemins de manière à optimiser le temps de parcours O-D sur un réseau congestionné. Le choix de l'itinéraire en transport public en fonction de la facilité à trouver une place assise est aussi un bon exemple bien que plus difficile à mesurer. Malgré tout, ces équilibres n'impliquent pas une action, du moins à court terme, de la part des responsables de l'offre.

L'équilibre demande - performance - offre engage les «fournisseurs» du transport à une action. Ceux-ci réagissent à un état d'équilibre demande - performance en contrôlant $\underline{F}$ et $\underline{R}$ à travers lesquels sont modifiées les niveaux de service $\underline{S}$, les capacités $\underline{q}$ et les tarifs $\underline{K}$. Comme on sait, ces éléments conditionnent la demande $[T, v]$ pour X , o , d donnés (moyen et long terme) tandis qu'ils provoquent la création de nouveaux équilibres demande - performance (court terme).

Dans un système dépourvu de congestion, les fournisseurs exercent un contrôle plus direct à l'égard des composantes de la performance car la demande n'altère plus cette dernière. Il s'ensuit un équilibre demande - offre. La correspondance entre $[C, s]$, $[\underline{S}, q, \underline{K}]$ et $[\underline{F}, \underline{R}]$ devient désormais directe. C'est le cas de certains systèmes centralisés comme dans le transport de marchandises.

Dans le contexte du transport urbain, la situation la plus complexe, sollicitant un équilibrage demande - performance - offre, est sans doute la plus appropriée. Pour cette raison, un schéma illustrant les interactions émanant de cet équilibre sera présenté ci-dessous:



DP: équilibre demande - performance

DPO: équilibre demande - performance - offre

Le comportement d'offre est inscrit dans la case $[F*, R*]$ qui suscite une mise à jour des aspects tangibles de l'offre soit la case $[F, R]$. Celle-ci correspond aux variables $[C, s, q]$ perçues par les usagers lesquels établissent une demande $[T, v]$ en respect de l'environnement générateur de déplacements $[X, o, d]$. Un état du système Σ constitué de la demande $[T, v]$ et de la performance réalisée en équilibre avec cette demande, soit $[S, q, K]$, servira de mesure d'appréciation pour une correction éventuelle de l'offre. L'environnement restreignant la manoeuvre des fournisseurs de transport est décrit dans $[U, E, B]$.

4.1.3 Structure d'un modèle générique d'offre de transport

Le cadre conceptuel présenté précédemment peut s'assimiler à une programme mathématique d'optimisation. La notion d'équilibre demande - performance a été formalisée dans un modèle intègre de demande doté d'une relation coût-traitic (voir section 3.7.3). Un véritable modèle d'offre décrit le comportement des instigateurs d'un système de transport en réponse à une évaluation du système, à une série d'objectifs et de contraintes. De telles composantes trouvent facilement leurs équivalences dans un problème d'optimisation.

La particularité d'un modèle d'offre est qu'il contient comme argument principal un sous-modèle de demande. Celui-ci adhère à la formulation générale en tant que mesure d'évaluation du système ou mesure inscrite dans la fonction objective. Boyce [1986] favorise un problème hiérarchisé à deux niveaux, l'un strictement relatif à l'offre, l'autre à la demande; nullement dissociables, ces deux niveaux sont fortement connectés. Le sous-problème de demande requiert, comme on sait, des mesures de coût généralisé de transport qui dépendent des éléments modifiables de l'offre. Les options de changement reposent, quant à eux, sur l'optimisation d'une fonction objective qui, hors de tout doute, sollicite une quantification plus ou moins désagrégée de la demande.

Considérons L comme un ensemble de caractéristiques du service de transport; L devient la variable décisionnelle. Une fonction objective appropriée devrait concerner la minimisation du total des coûts de transport. D'autres objectifs peuvent être poursuivis mais cela n'affecte pas la démonstration. Le problème principal s'écrit comme suit:

$$\min_{\{L\}} f(L, [C, s], q, [T, v]) \quad \text{sujet à } g(L, P, B) \leq 0 \quad (4.1-4.2)$$

$$\text{où } [C, s], q \equiv L$$

$g(\cdot)$ est un vecteur fonctionnel décrivant les contraintes. P désigne un ensemble de dépenses relié à l'adoption d'un ensemble de caractéristiques L et B est le budget autorisé (ou les ressources disponibles). La matrice C de coûts de transport, le vecteur s de fonctions de congestion et le vecteur q de capacités sont définis en fonction des attributs de transport L . $[T, v]$ résolvent le sous-problème de minimisation suivant:

$$\min_{[T,v]} + ([T,v],[C,s],q,L) \quad \text{sujet à } h([T,v],Y) \leq 0 \quad (4.3-4.4)$$

La notation Y symbolise les facteurs externes pouvant contraindre la demande (ex: totaux marginaux, niveau d'entropie). L'application est générale puisqu'elle considère des effets de congestion: si tel n'est pas le cas, les vecteurs v , s , et q deviennent sans pertinence. Le sous-problème (4.3-4.4) peut être interprété par de nombreux problèmes d'optimisation de la demande comme ceux décrits dans le chapitre 3.

La type définitif de programmation est connu lorsque les caractéristiques contenues dans L sont identifiées. Si les modifications de l'offre portent sur tout élément monétaire ou temporel du coût généralisé de transport (tarif, vitesse des véhicules, prix du carburant, prix du stationnement) mais que la structure du réseau reste inchangée, on assiste à une variation linéaire de l'impédance du déplacement pour les trajets concernés. Toutefois, les opérateurs d'un système sont plutôt confrontés à des décisions regardant les éléments rattachés au réseau. La synchronisation des feux de circulation, l'augmentation de la capacité d'un segment routier, la fréquence des autobus et leur correspondance, l'addition et soustraction de liens sur le réseau soulèvent des problèmes d'optimisation beaucoup plus complexes.

Evidemment, bien que le cadre conceptuel spécifie que les options de transport peuvent se faire à divers niveaux, un programme mathématique manipulable se contentera d'optimiser un ensemble de choix. Abdulaal et LeBlanc [1979] considérèrent l'allocation de capacité sur un réseau de manière à réduire les coûts de déplacement. Scheele [1980] proposa un problème de détermination de la fréquence optimale des lignes d'autobus en accord avec une configuration des déplacements. Toutefois, le problème d'offre le plus connu reste indubitablement celui de «conception discrète d'un réseau» (communément appelé "network design problem"). Il réfère à l'addition ou soustraction de liens dans un réseau mais se concilie aussi à celui de modification.

Nous privilégierons le modèle de conception car il évoque l'aspect le plus déterminant du développement d'un système de transport, l'évolution spatiale du réseau. Plus que tout autre modification, l'avènement d'une

nouvelle connexion de transport change le patron géographique d'accessibilité. La demande O-D de transport peut être substantiellement perturbée et à plus long terme, la distribution spatiale des activités, à la fois tributaire et responsable des interactions, procèdent à quelques lents reajustements.

4.2 PROBLEME DE CONCEPTION DE RESEAU

4.2.1 Formulation avec coûts fixes de transport

Le fameux «probleme de conception de reseau» consiste à déterminer la meilleure amélioration qu'on puisse faire à un réseau. Le problème peut être étudié sous une forme «discrète» (addition et soustraction de liens ou de voies) ou sous une forme «continue» (augmentation et réduction de capacité des liens) [voir Addulaal, LeBlanc, 1979]. En dépit de sa légitimité mathématique, l'approche continue apparaît plutôt invraisemblable dans un contexte d'aménagement de transport puisque des solutions comprenant des parties fractionnaires de voies routières sont possibles. Ici, la forme discrète sera d'ailleurs privilégiée.

Le singulier employé pour désigner le problème discret de conception de réseau ne signifie pas pour autant qu'il n'existe qu'une formulation du problème. Un des éléments constants reste la variable décisionnelle (disons y_z) de type binaire (0 ou 1) qui caractérise le statut d'un lien projeté. La valeur positive indique que le lien est ajouté au réseau alors que la valeur nulle témoigne de son inexistence. Ordinairement, on cherche parmi un ensemble de liens projetés L^p , un sous-ensemble de liens qui minimisera une fonction de coût total. Le réseau amélioré, symbolisé par l'ensemble $L(y)$, est formé de l'union du réseau initial L avec les éléments de L^p dont $y_z = 1$.

La fonction objective peut se présenter sous différentes formes. Steenbrink [1974] exposa une forme très générale sollicitant tant la minimisation des coûts de déplacements que celle des coûts de construction. Selon la notation employée ici, écrivons:

$$\min_{\{y\}} \sum_z [v_z c_z + y_z p_z] \quad (4.5)$$

$$\text{où } v_z = \sum_{i,j} \delta_{ij} t_{i,j} \quad \delta_{ij} \begin{cases} = 1 & \text{si } i \in K \cap (ij) \\ = 0 & \text{sinon} \end{cases} \quad (4.6)$$

$$y^* = \{y_1, y_2, \dots, y_z, \dots, y_m\} \quad y_z \begin{cases} = 1 & \text{si } z \text{ ajouté} \\ = 0 & \text{sinon} \end{cases}$$

$$\text{sujet à } My_z \geq v_z \geq 0 \quad (4.7)$$

v_l : volume de circulation sur le lien l ; c_l : coût (fixe) de transport sur le lien l ; p_l : coût de construction du lien l ; $\tau_{i,j}$: nombre (fixe) de déplacements entre les lieux i et j ; $K_v(i,j)$: ensemble de liens constituant le plus court chemin entre i et j ; M : un nombre arbitraire très grand ($M > \max\{v_l\}$).

La sommation dans (4.5) ne concerne pas que les m liens (unidirectionnels ou bidirectionnels) inclus dans le sous-ensembles des liens projetés L^p . Elle porte sur les liens déjà existants, donnant ainsi une mesure globale du réseau, mais prend soin de mettre les coûts d'amélioration p_l à zéro pour les liens appartenant au réseau initial.

En confrontant deux types de coût dans la fonction objective, on se bute à un problème d'incomparabilité. Pour cette raison, les formulations les plus courantes procèdent à une sélection optimale des liens qui minimise uniquement les coûts de transport mais qui doit satisfaire aussi une contrainte de budget limite. Le problème de conception de réseau décrit par Scott [1969] peut être énoncé comme suit:

$$\min_{\{y\}} \sum_l v_l c_l \quad (4.8)$$

$$\text{sujet à (4.6), (4.7) et } \sum_l y_l p_l \leq B \quad (4.9)$$

Le lecteur averti remarquera que la lagrangienne de (4.8-4.9) est équivalente à celle du programme précédent à l'exception d'un multiplicateur lagrangien accolé à la contrainte de budget. Ce problème de «programmation entière» ("integer programming") consiste à sélectionner un sous-ensemble parmi les L liens projetés qui améliorera le plus la performance du réseau selon les disponibilités budgétaires. Plusieurs méthodes algorithmiques furent proposées pour résoudre ce problème véritablement difficile bien que simple en apparence. Le sous-chapitre suivant abordera ce thème.

4.2.2 Formulation avec un réseau congestionné

Dans un contexte de réseau urbain, la présomption de fixité du coût de transport devient inappropriée puisque l'affectation de trafic modifiera le coût de parcours. L'effet de congestion peut être représenté par une fonction croissante et monotone comme celle préconisée par le "U.S. Federal

Highway Administration" (voir section 3.6.2):

$$s_z(v_z) = s_z^0 \left[1 + \alpha (v_z/q_z)^\eta \right] \quad (4.10)$$

Dans la formulation avec coûts fixes, l'allocation des flux se faisait selon le principe du plus court chemin. Avec une relation coût-traffic, le principe des temps égaux de Wardrop [1952] devient propre à une affectation à chemin multiple. Dans la section 3.6.2, un modèle d'affectation de trafic d'équilibre a été présenté. La signification d'un tel principe d'affectation, la structure du programme mathématique et les conditions d'optimalité ont été abondamment discutées. LeBlanc [1975] formula un problème discret de conception de réseau qui considère un trafic d'équilibre optimum pour l'utilisateur. Désormais, le programme mathématique est de nature mixte puisque le vecteur entier $\{y_z\}$ et le vecteur continu $\{v_z\}$ sont à déterminer. Écrivons le programme selon une notation conforme au modèle de demande:

$$\min_{\{y\}} \sum_z v_z s_z(v_z) \quad \text{sujet à} \quad \sum_{z \in L^p} y_z p_z \leq B \quad (4.11-4.12)$$

où v_z et $s_z(v_z)$ sont respectivement le flux et le coût d'équilibre sur le lien z . Ils sont obtenus par le programme interne suivant:

$$\min_{\{H^k\}} \sum_z \int_0^{v_z} s_z(v) dv \quad (4.13)$$

$$\text{sujet à} \quad v_z = \sum_{i,j,k} \delta_{ij}^k h_{ij}^k, \quad \forall z \in L(y) \quad v_z \leq M y_z \quad \forall z \in L^p \quad (4.14-4.15)$$

$$\sum_k h_{ij}^k = t_{ij} \quad \forall ij \quad h_{ij}^k \geq 0 \quad \forall ijk \quad (4.16-4.17)$$

L'ensemble $L(y)$ définit le réseau courant i.e. l'union du réseau initial L avec les liens $z \in L^p$ dont $y_z=1$. La variable d'incidence route - lien δ_{ij}^k est définie selon la configuration courante du réseau. LeBlanc rappelle que y_z reste la seule variable décisionnelle; en effet, les aménagistes choisissent les liens à construire, non pas les flux. Toutefois, afin d'évaluer le patron de flux résultant d'un réseau modifié $L(y)$, on doit résoudre le sous-problème d'affectation (4.13-4.17). Globalement, le problème formulé par LeBlanc ressemble à celui décrit par (4.6-4.9) avec coûts fixes. Dans un cas, un algorithme du plus court chemin détermine l'utilisation des arcs. Dans l'autre, un modèle d'affectation plus sophistiqué alloue les voyageurs parmi différents chemins.

4.2.3 Incorporation d'une demande inter-zonale variable

Dans une échelle de complexité des problèmes de conception de réseau, la version comprenant un trafic d'équilibre est certes située plus haut que celle avec des coûts fixes de transport. Elle n'occupe pourtant pas l'échelon le plus élevé. Boyce et Janson [1980] proposeront une formulation qui tenait compte, outre de l'affectation routière, de la demande origine-destination variable. L'endogénéisation de la demande O-D dans un problème de conception ne semble pas qu'un simple exercice intellectuel. La qualité d'une amélioration d'un réseau s'apprécie à long terme. Si l'amélioration s'avère très pertinente, elle ne modifiera pas que les parcours empruntés par les voyageurs mais, à plus long terme, affectera la distribution de l'accessibilité origine-destination. Une modélisation de l'offre de transport qui se veut consistante doit ainsi considérer les effets rétroactifs sur la demande [Wong, 1984].

La démarche intuitive qui consiste à remettre à jour une distribution spatiale des déplacements avec un modèle de gravité après chaque correction du réseau reste peu prometteuse. D'une part, il n'y a aucune preuve de convergence vers un réseau optimal. D'autre part, le coût inter-zonal est donné par le trafic d'équilibre dont l'obtention n'est pas garantie dans une procédure répétitive des modèles de distributions spatiale et routière [Florian, Nguyen, Ferland, 1975]. Une formulation unifiée s'impose. La formulation du problème combiné de distribution et d'affectation est bien connue et a été traitée dans le sous-chapitre 3.7. Le problème comporte toutefois un paramètre inconnu qui influence la variabilité de la distribution spatiale [Erlander, 1977]. Dans un problème de distribution spatiale où l'on maximise l'entropie, le paramètre est estimé en solutionnant une contrainte de coût total (il est alors son multiplicateur de Lagrange). Dans un problème combiné, la non-fixité du coût de même que la nature de la fonction objective convexe (eq. 4.11) prohibe l'emploi d'une contrainte de coût. La satisfaction d'une contrainte relative au niveau d'entropie reste alors la méthode la plus appropriée [Erlander, Nguyen, Stewart, 1979]. Boyce et Janson [1980] reconnurent eux aussi la convenance de la contrainte d'entropie dans le problème de conception de réseau. Dans ce cas, la fonction objective consiste aussi à minimiser les dépenses globales de transport. Le problème avec demande inter-zonale variable et affectation

d'équilibre (optimum pour l'utilisateur) qu'ils énoncèrent s'écrit comme suit:

$$\min_{\{y\}} \sum_e v_e s_e(v_e) \quad \text{sujet à } \sum_{e \in L^P} y_e p_e \leq B \quad (4.18-4.19)$$

où v_e et $s_e(v_e)$ proviennent du sous-programme suivant:

$$\min_{(H^k, T)} \sum_e \int_0^{v_e} s_e(v) dv \quad (4.20)$$

$$\text{sujet à } v_e = \sum_{i,j,k} \delta_{ij}^k h_{ij}^k \quad \forall e \in L \quad v_e \leq M y_e \quad \forall e \in L^P \quad (4.21-4.22)$$

$$\sum_k h_{ij}^k - t_{ij} = 0 \quad \forall i,j \quad h_{ij}^k \geq 0 \quad \forall i,j,k \quad (4.23-4.24)$$

$$\sum_j t_{ij} = O_i \quad \forall i \quad \sum_i t_{ij} = D_j \quad \forall j \quad (4.25-4.26)$$

$$-\sum_{i,j} t_{ij} (Int_{ij} - 1) \geq S^T \quad (4.27)$$

Le programme mathématique ci-dessus procède à une sélection des liens $e \in L^P$ qui minimise les coûts totaux de déplacement et qui satisfont la contrainte budgétaire. La répartition du trafic sur le réseau s'effectue selon un principe d'optimalité pour l'utilisateur et la demande O-D est établie en accord avec un niveau de variabilité S^T et des contraintes marginales d'origine et de destination. Comme il a été mentionné dans le sous-chapitre 3.6, le principe d'affectation ne minimise pas le total des coûts de transport car il émane d'une rationalité individuelle et non de celle du système. Dans le cas où le réseau investigué est contrôlé par un organisme central, on peut substituer la fonction (4.20) par une véritable somme des coûts comme [Boyce, Janson, 1980]:

$$\min_{\{H^k\}} \sum_e v_e s_e(v_e) \quad (4.28)$$

Dès lors, l'affectation est telle qu'elle minimise véritablement les dépenses. Une telle présomption dans un contexte de déplacements individuels implique que les voyageurs affichent un comportement coopératif, ce qui semble peu vraisemblable.

L'un des problèmes rencontrés dans un modèle de conception couplé avec un trafic d'équilibre optimal pour l'utilisateur demeure la dissemblance des fonctions objectives. Le problème de conception avec coûts fixes est convexe (tout optimum est global) et le problème d'affectation optimale pour l'utilisateur avec un réseau fixe est aussi convexe. Pourtant, la sélection d'un sous-ensemble de liens (évaluation du vecteur y) avec un tel trafic d'équilibre n'est plus convexe puisque la fonction objective du sous-

problème se distingue de celle du problème de conception. En effet, le coût total de transport pourrait croître avec l'addition d'un lien dans le réseau. Ces résultats plutôt étonnants sont connus sous le nom de «paradoxe de Braess» [Braess, 1968; Murchland, 1970; LeBlanc, 1975]. Les effets du paradoxe sont à considérer dans la composition d'un algorithme. De toute évidence, en optant pour un trafic d'équilibre optimal pour le système, on évite ces résultats contre-intuitifs car les fonctions objectives visent toutes deux la minimisation du total des coûts.

4.2.4 Application multi-modale

Dans un contexte multi-modal où la répartition n'apparaît pas définitive, l'analyste doit considérer que l'amélioration d'un réseau augmentera sensiblement la part de marché octroyée au mode bénéficiaire. En s'inspirant de la formulation avec demande O-D variable et du modèle unifiant affectation routière, répartition modale et distribution spatiale (3.144-3.151), on peut formuler un problème de conception de plusieurs réseaux. Une notation selon l'affiliation modale des liens semble nécessaire si on présume que chacun des liens appartient exclusivement à un réseau modal.

La formulation du problème de conception ne présente pas de difficulté particulière car l'affiliation modale des liens n'apparaît ni plus ni moins comme une désagregation supplémentaire des flux sur le réseau. Selon Wong [1984], le problème générique de conception autorise n'importe quelle séparation de la demande; il est formulé pour différentes «commodités». Un trajet O-D spécifique y est considéré comme une commodité distincte. Un parcours sollicitant un mode peut alors prétendre à ce même titre.

Dans la présente thèse, une application bi-modale d'un modèle d'offre de transport sera montrée. Une distinction classique des réseaux de transport urbain est faite; l'un est celui des routes urbaines à l'usage des automobiles, l'autre, des parcours de transporteurs publics. Pour faciliter la modélisation, on présume l'indépendance des réseaux. Ainsi, les conditions de circulation sur les routes urbaines ne dépendent ni n'affectent la performance du transport collectif. De plus, nous considérerons que les effets de congestion restent l'apanage du réseau réservé aux automobilistes. Gaudry [1975] remarqua qu'ordinairement le service de trans-

port en commun ne connaît guère de variation des conditions de déplacement puisque les responsables des organismes ajustent régulièrement la capacité de la flotte de véhicules afin de conserver une capacité excédentaire.

L'absence d'une relation coût-traffic pour l'un des réseaux semble simplifier le problème de conception car l'affectation des usagers pourra se faire sur l'itinéraire le plus court. Toutefois, on entrevoit certaines complications dues à la conciliation de deux réseaux dotés de caractéristiques globales différentes. Du point de vue conceptuel, il ne s'agit pas d'un obstacle énorme puisqu'il est facile de supposer pour le réseau "sans limite de capacité", une fonction de congestion de type uniforme. Par exemple:

$$s_c(v_c) = s_c \quad \forall l \in L(m)$$

où $L(m)$ est l'ensemble des liens du réseau modal m à capacité illimitée et s_c est un coût fixe par lien. Une formulation bi-modale, où $L(1)$ et $L(2)$ désignent respectivement le réseau routier et les parcours de transport public, peut s'écrire de la façon suivante:

$$\min_{\{y\}} \sum_{l \in L(1)} s_c(v_l) v_l + \sum_{l \in L(2)} s_c v_l \quad \text{sujet à} \quad \sum_{l \in L^p} y_l p_l \leq B \quad (4.29-4.30)$$

v_l , $l \in L(2)$, est calculé d'après:

$$v_l = \sum_{i,j} \delta_{ij} t_{ij} p_{ij} \quad \delta_{ij} = 1, \text{ si } l \in K_{ij}, = 0, \text{ sinon} \quad (4.31)$$

$s_c(v_l)$, v_l , $l \in L(1)$, t_{ij} , p_{ij} proviennent de programme suivant:

$$\min_{\{H^k, P^m, T\}} \sum_l \int_0^{v_l} s_c(v) dv + \sum_{i,j} t_{ij} p_{ij} c_{ij} \quad (4.32)$$

$$\text{sujet à } v_l = \sum_{i,j,k} \delta_{ij} h_{ijk} \quad \forall l \in L(1) \quad v_l \leq M y_l \quad \forall l \in L^p \quad (4.33-4.34)$$

$$\sum_k h_{ijk} - t_{ij} p_{ij} = 0 \quad \forall ij \quad h_{ijk} \geq 0 \quad \forall ijk \quad (4.35-4.36)$$

$$-\sum_{i,j} t_{ij} (\ln t_{ij} - 1) \geq S^T \quad \sum_j t_{ij} = O_i \quad \forall i \quad \sum_i t_{ij} = D_j \quad \forall j \quad (4.37-4.39)$$

$$-\sum_{i,j} t_{ij} \sum_m p_{ij} (\ln p_{ij} - 1) \geq S^P \quad \sum_m p_{ij} = 1 \quad \forall ij \quad (4.40-4.41)$$

c_{ij} : plus court chemin entre i et j selon le mode 2; h_{ijk} : flux d'automobilistes sur l'itinéraire k du trajet (ij) ; p_{ij} : proportion des voyageurs du trajet (ij) accordé au mode m .

Dans le programme precedent, la selection des liens se fait sans discrimination de l'affiliation modale. Les deux types de liens ont des unites comparables de coût de construction. De même, les coûts de transport epargnés par l'amélioration des reseaux peuvent être controntés. Dans cette formulation, il n'y a pas de préférence predeterminée quant au reseau favorisé. Un budget B peut autoriser des options mixtes de construction tout comme il peut privilegier un seul des reseaux en autant que la diminution totale des coûts de transport soit maximale.

4.3 ALGORITHMES POUR LE PROBLEME DE CONCEPTION DE RESEAU

Dans le cadre de cette recherche, le modèle discret de conception de réseau avec trafic d'équilibre, demande O-D variable et partage modal devient certes le plus intéressant et le plus consistant avec la modélisation unifiée de la demande de transport présentée au chapitre précédent. La conjugaison de l'affectation routière et de la distribution spatiale se fait de la même manière dans les modèles d'offre et de demande. Une version multi-modale reste plausible car elle n'implique qu'une partition de la contrainte d'entropie (voir section 3.5.3).

Pourtant, on n'abordera pas tout de suite la méthodologie algorithmique associée à la recherche d'une solution du problème de conception d'un réseau congestionné et avec demande endogène. Il semble préférable de traiter avant tout du problème avec coûts fixes pour la simple raison que les problèmes de conception sont des programmes non-linéaires considérés comme très difficiles (dans la terminologie de la recherche opérationnelle, ils sont appelés "NLP-Hard") [Dionne, Florian, 1979]. En fait, ils deviennent des problèmes combinatoires de grande dimension: un sous-ensemble de liens projetés L^p de taille m peut occasionner jusqu'à 2^m solutions possibles (deux états possibles, 0 et 1, m éléments). On comprend qu'une énumération exhaustive des configurations projetées de réseau devient vite impensable lorsque le nombre de liens atteint la dizaine. Une démarche algorithmique écartant efficacement les alternatives sub-optimales et menant vers l'unique solution optimale est nécessaire. Une telle méthode convient également aux applications où le réseau est propice à la congestion et où le patron migratoire est endogène. On privilégiera les algorithmes «exacts» au dépens des «heuristiques» car les applications prévues restent plutôt de tailles modestes (tout au plus une vingtaine de liens projetés).

4.3.1 Solution du problème avec coûts fixes

Il existe des algorithmes très efficaces pour résoudre des cas particuliers du problème de conception de réseau comme celui du «plus court chemin», du «réseau arborescent minimal» et du «voyageur de commerce». Magnanti et Wong [1984] notèrent que la recherche du plus court chemin

entre deux points dans un réseau est un cas spécial où l'on minimise les coûts de déplacement mais où les flux sont uniformes et les coûts de construction sont omis. Quant au réseau arborescent minimal, il ne concerne pas les dépenses de transport mais uniquement celles occasionnées par la construction (il n'y a pas de budget limite). Le problème du voyageur de commerce consiste par ailleurs à définir un parcours minimal passant par tous les noeuds; les flux sont uniformes sur tous les liens et le total des coûts de transport est minimisé.

De nombreuses variantes de l'algorithme du plus court chemin existent et une excellente revue comparative fut faite par Golden [1976]. Gallo et Pallottino [1984] attribuerent cet intérêt au fait que ce problème est rarement absent d'un programme de transport, y apparaissant souvent comme un sous-problème. D'ailleurs, le problème de conception de réseau (4.6-4.9) requiert la solution du plus court chemin pour chacune des paires nodales. Parmi les procédures les plus classiques mentionnons celles de Dantzig [1960] (structure arborescente) et de Whiting et Hillier [1960] (reclassement nodal).

Le réseau arborescent minimal (en anglais "minimal spanning tree") est soluble par un algorithme très efficace. En fait, il demeure une solution spécifique du problème de conception et il peut servir à titre de comparaison avec d'autres solutions [voir Dionne, Florian, 1979]. Pourtant, la méthode algorithmique associée reste de peu d'utilité dans la recherche de la configuration optimale d'un réseau avec flux différenciés et coûts de déplacement à minimiser.

Par sa complexité, les méthodes résolvant le problème discret de conception de réseau attirèrent l'attention de maints chercheurs. La plupart optèrent pour une approche algorithmique du type "branch & bound" (ou «branche-et-borne»), notamment Scott [1969], Boyce, Fahri et Weischedel [1973], Hoang [1973], Boyce et Soberanes [1979] ainsi que Dionne et Florian [1979]. La structure de ce type d'algorithmes est adéquatement résumée dans Magnanti et Wong [1984] et convient particulièrement à la résolution des «programmes entiers» (variables discrètes) et des «programmes mixtes-entiers» (variables discrètes et continues).

Un arbre d'énumération «branche-et-borne» comprend des «noeuds» correspondant chacun à une solution partielle i.e. où un certain nombre d'alternatives ont été fixées tandis que d'autres restent à déterminer. Subséquentement à chaque noeud, on procède à une division (ici, 0 et 1) opérée sur un élément encore non fixé. À chacun des noeuds, est octroyée une «borne inférieure» ("lower bound") équivalant au minimum possible de la fonction objective qui puisse être obtenu dans le meilleur des cas. Dans le problème concerné, il s'agit d'évaluer la fonction comme si tous les liens indéterminés seraient ajoutés. De cette façon, on apprécie la "valeur prometteuse" d'une solution partielle. Si cette borne inférieure devient supérieure à la meilleure performance rencontrée dans une autre branche, il paraît inutile de progresser dans toutes les branches succédant à ce noeud. L'élimination des alternatives sub-optimales se fait ainsi et, évidemment, selon la violation des contraintes.

Examinons quelques mécanismes de bornage appropriés au problème de conception. Définissons pour chaque noeud D de l'algorithme, un sous-ensemble $L^F(D)$ de liens avec valeurs fixées ($y_L=0$ ou $y_L=1$ pour $L \in L^F(D)$). Un sous-ensemble complémentaire des liens indéterminés est symbolisé par $L^G(D)$. De plus, denotons $F(y)$ comme la valeur de la fonction objective calculée avec le réseau initial L et les liens $L \in L^F(D)$; $F(y^D)$ est la valeur associée à la meilleure solution issue d'un réseau $L \cup L^F(D) \cup L^G(D)$ où pour tout $L \in L^G(D)$, $y_L=1$. Boyce et al. [1973] remarquèrent qu'à toute position de l'arbre d'énumération, l'inégalité suivante est satisfaite:

$$F(y) \leq F(y^D) \quad (4.42)$$

$F(y^D)$ est alors considérée comme une borne inférieure au noeud D. Hoang [1973] suggéra une borne plus sévère permettant ainsi d'éliminer plus rapidement des configurations non prometteuses. Cette fonction de borne inférieure se formule comme suit:

$$F(y) \leq F(y^D) + \sum_{L \in L^G} (1-y_L) I_L(y^D) \quad (4.43)$$

où $I_L(y^D) = c_{i,j}(y^D - L) - c_{i,j}(y^D)$ $(ij) \equiv L$

$c_{i,j}(y^D)$ est le coût du plus court chemin dans un réseau (y^D) entre les lieux i et j identifiés aux extrémités du lien L . Le symbolisme $(y^D - L)$ denote un réseau (y^D) à lequel on a soustrait le lien L . La fonction

$I_\ell(y^D)$ quantifie alors l'incrément du coût minimal entre les noeuds delimitant le segment ℓ occasionné lorsque celui-ci est omis du réseau défini par le vecteur y^D (dans ce cas, $y_\ell=0$). La borne 4.43 apparaît plus élevée que 4.42 car l'omission de chacun des liens indéterminés doit au moins augmenter la meilleure valeur possible $F(y^D)$ par $I_\ell(y^D)$ [Magnanti, Wong, 1984]. Dionne et Florian [1979] examinèrent une borne intérieure plus forte mais requérant davantage de calculs. Dans 4.43, on ne considère que les increments liés aux paires origines-destinations situées aux extrémités des liens investigués. Considérons l'ensemble des noeuds ou origines-destinations (ij) du réseau et repartissons-les dans des sous-ensembles S_ℓ identifiant les paires dont le plus court chemin sollicite le passage du lien ℓ . Pour tout $(ij) \in S_\ell$, $Q_{ij}^\ell(y^D)$ est défini comme l'incrément rattaché au plus court chemin entre i et j lorsque ℓ est retranché du réseau y^D . Alors:

$$F(y) \leq F(y) + \sum_{\ell \in L^D} (1-y_\ell) \sum_{(ij) \in S_\ell} Q_{ij}^\ell(y^D) \quad (4.44)$$

où $Q_{ij}^\ell(y^D) = c_{i,j}(y^D - \ell) - c_{i,j}(y^D) \quad \forall (ij)$

devient une borne inférieure au point D de l'arbre d'énumération. Dionne et Florian démontrèrent toutefois que cette borne n'est pas inviolable et que dans certains cas spéciaux, elle peut conduire vers des configurations non-optimales. Elle peut pourtant être utilisée dans un algorithme heuristique car elle se montre très efficace dans la maintes applications réalistes.

Outre la détermination d'une borne inférieure convenable aidant à évincer les branches dont les solutions subséquentes sont assurément non optimales, l'algorithme peut être aussi amélioré par une règle de branchage plus perspicace. Boyce et al. [1973] préconisèrent une énumération sans ordre particulier, chaque noeud D se branchant sur un lien $\ell \in L^D(D)$ sans considération explicite de ses caractéristiques. Dionne et Florian [1979] proposerent un branchage sur y_ℓ ($\ell \in L^D(D)$) dont le lien maximise le rapport $I_\ell(y^D)/p_\ell$. Le numérateur relève du coût de transport direct épargné par l'addition du lien tandis que le dénominateur évoque les dépenses de construction du lien projeté. Dans une application avec limite budgétaire, les liens les moins onéreux à construire mais les plus efficaces en terme de transport ont plus de chance d'appartenir au réseau optimal. En bran-

chant en priorité ces liens, les configurations les plus pertinentes se dessinent plus rapidement. Ces auteurs constatèrent, qu'en prônant cette stratégie de branchage, une amélioration importante en terme de temps ordinateur.

4.3.2 Algorithme avec trafic d'équilibre

Du point de vue de la structure algorithmique, l'insertion des effets de congestion de même que des flux O-D variables n'implique pas de changement majeur. La fonction objective du strict problème de conception reste identique bien que, dorénavant, la longueur du plus court chemin n'est plus déterminée par l'algorithme homonyme mais plutôt par un problème d'affectation routière.

Dans un premier temps, on traitera spécifiquement du problème avec trafic d'équilibre seulement. LeBlanc [1975] favorisa une méthode "branch & bound" pour la recherche de la solution. Dans le calcul des bornes inférieures, il prévint l'emploi de la fonction (4.13). Sa non-convexité lorsque des liens sont ajoutés ou enlevés (paradoxe de Braess) empêche d'établir une véritable borne inférieure. LeBlanc énonça et prouva un lemme permettant de trouver une issue: pour tout noeud D de l'arbre de branche-et-borne, la solution optimale du programme (4.13-4.17) — la congestion minimale selon les flux d'équilibre de l'utilisateur dans un réseau $L \cup L^F(D)$ — est supérieure ou égale à celle du programme (4.28, 4.14-4.17) lequel donne la congestion minimale selon les flux d'équilibre du système dans un même réseau. Par conséquent, on établit la borne inférieure au noeud D comme étant le total des coûts de transport selon un trafic optimal pour le système sur un réseau $L \cup L^F(D) \cup L^G(D)$ où $y_i = 1 \quad \forall i \in L^G(D)$. En somme, déterminer la borne inférieure $F(y^D)$ équivaut à résoudre le problème d'affectation (4.28, 4.14-4.17) avec le réseau identifié au vecteur y^D .

La différence entre les problèmes pour l'utilisateur et pour le système réside essentiellement dans leurs fonctions objectives. La fonction (4.28) apparaît, selon LeBlanc, très similaire à (4.13), seul un coefficient étant distinct. En effet, si on reprend la fonction de congestion (4.10) et on développe selon (4.13), on a:

$$\sum_e \int_0^{v_e} s_e(v) dv = \sum_e \int_0^{v_e} s_e^* \left[1 + \alpha (v/q_e)^\eta \right] dv$$

par ailleurs,

$$\int s + s\alpha q^{-\eta} v^\eta dv = sv + s\alpha q^{-\eta} \frac{1}{\eta+1} v^{(\eta+1)} + C$$

donc,

$$\sum_e \int_0^{v_e} s_e(v) dv = \sum_e s_e^* \left[1 + \{\alpha/(\eta+1)\} (v_e/q_e)^\eta \right] v_e \quad (4.45)$$

alors que

$$\sum_e s_e(v_e) v_e = \sum_e s_e^* \left[1 + \alpha (v_e/q_e)^\eta \right] v_e \quad (4.46)$$

Les fonctions (4.45) et (4.46) étant isomorphes, le coefficient α se substituant à $\{\alpha/(\eta+1)\}$, le problème d'affectation du système devient alors soluble par l'algorithme de Frank-Wolfe, technique généralement associée à la résolution du trafic d'équilibre pour l'utilisateur (voir section 3.8.4).

Le calcul des coûts de transport s'avérant particulièrement exigeant, chaque fois nécessitant l'application de la méthode de Frank-Wolfe, il convient que la borne supérieure ne sera pas rendue plus sévère en rajoutant des incréments comme dans (4.43) et (4.44). Non plus, il n'y aura de règle de branchage comme celle maximisant le rapport entre incrément et coût de construction. Malgré tout, un ordre d'énumération établi en fonction du coût minimal de parcours (s_e^*), de la capacité (q_e) et du coût d'amélioration (p_e) perfectionnerait sans nul doute l'algorithme.

L'algorithme branche-et-borne est conduit de la façon suivante. A chaque noeud D de l'arbre d'énumération, on résout le problème d'affectation (4.28, 4.14-4.17) pour un réseau dans lequel les liens indéterminés sont considérés comme ajoutés. Ainsi:

$$F(y^D) = \min_{\{H^D\}} \sum_e s_e(v_e) v_e$$

sujet à (4.14-4.17) et pour $l \in L^F(D)$, $y_l = 0$ ou 1 (fixe)

$$l \in L^G(D), y_l = 1$$

LeBlanc rappelle, qu'en pratique, on calcule une borne inférieure pour deux noeuds. En effet, le noeud associé à l'option $y_l=1$ affiche une borne inférieure identique au noeud précédent puisque la borne y est calculée

comme si y_e indéterminé prenait une valeur de 1. Seul le noeud complémentaire où $y_e=0$ nécessite l'évaluation d'un nouveau $F(y^p)$.

En tout temps, si $F(y^p)$ est plus élevée que la fonction (4.11) associée à une configuration antérieurement complétée (le vecteur y entièrement défini), il est inutile de poursuivre la recherche sur les noeuds successeurs et on continue sur le noeud complémentaire. De plus, on met systématiquement à l'écart les branches dont les coûts violent la contrainte budgétaire (4.12). Cela empêche toute progression vers des configurations inacceptables. Une configuration est dite optimale lorsque le vecteur y est complet, la contrainte (4.12) est satisfaite et que le total (4.11) calculé selon un trafic d'équilibre solutionnant le programme (4.13-4.17) reste en deça de tous les autres totaux liés aux configurations respectant les mêmes conditions.

4.3.3 Distribution spatiale et répartition modale

L'incorporation d'une demande origine-destination et d'un partage modal endogènes se fait, comme on sait déjà, en introduisant des contraintes d'entropie dotées d'arguments t_{ij} et p_j , dans le sous-programme d'affectation des flux sur le réseau. Boyce et Janson [1980] privilégierent, encore une fois, une algorithmie arborescente pour solutionner des problèmes discrets de conception de réseau en présence d'un trafic d'équilibre et d'une distribution O-D endogène. A la méthode de LeBlanc [1975], ils imposèrent la satisfaction des contraintes de couplage (4.23), d'origine (4.25), de destination (4.26) et d'entropie (4.27). Ainsi, la procédure branche-et-borne décrite précédemment tient encore mais dans l'évaluation de la fonction objective (vecteur y complète), il faut désormais résoudre le programme combiné (4.20-4.27). Quant aux bornes inférieures, la solution optimale d'un programme greffant la fonction (4.28) aux contraintes (4.21-4.27) est sollicitée. La résolution de ces deux sous-problèmes se fait à l'aide de l'algorithme de Frank-Wolfe adapté au problème combiné de distribution et d'affectation (voir section 3.8.4).

Apparemment, une telle procédure semble onéreuse puisque la solution du sous-programme demande une phase de calibration (selon la technique de Newton-Raphson décrite dans 3.8.3) effectuée en alternance avec une méthode de

solution du problème combine [voir Boyce et al., 1983]. Une fois résolu, le sous-programme ne fournit qu'une borne inférieure alors qu'il y en a tant à évaluer dans l'algorithme branche-et-borne.

Boyce et Janson [1980] discutèrent d'un mécanisme permettant d'abréger substantiellement la démarche énumérative. Il s'agit de composer un «seuil inconditionnel» ("unconditional threshold") qui, au début de l'arbre, permet de reconnaître certains liens qui doivent nécessairement constituer la configuration optimale. Pour chacun des liens $l \in L^*$, un seuil $U(l)$ est évalué comme étant la valeur de la fonction objective pour un réseau incluant tous les liens projetés à l'exception de l . Le seuil inconditionnel est utilisé à chaque iteration de la recherche afin de tester si le lien en question doit être inclus dans la solution optimale. Si, à toute étape de la recherche, $U(l) > F(y^*)$, où $F(y^*)$ est la valeur la plus basse de la fonction objective pour une configuration complétée, alors tout réseau qui n'inclut pas l peut être éliminé de la recherche. Plus le nombre de liens dont le seuil excède la fonction objective est élevé, plus un nombre considérable de configuration est écarté. Le seuil inconditionnel peut être vu comme un mécanisme auxiliaire de la procédure branche-et-borne. Il ne réduit pas pour autant la difficulté de calculer les bornes inférieures ainsi que toute évaluation de la fonction objective.

L'application dans un contexte bi-modal comme celui décrit à la section 4.2.4, n'exige pas plus de réajustement méthodologique. Le calcul des bornes inférieures requiert une affectation optimale du système mais pour un mode seulement car, en absence de congestion, la minimisation des coûts de transport est identique pour le système et pour l'utilisateur. La méthode de Frank-Wolfe adaptée pour le problème combiné d'affectation, de répartition et de distribution de Florian et Nguyen [1978] convient spécifiquement en raison que ces auteurs considérèrent deux modes dont l'un avec coûts de transport fixes et l'autre, avec des relations coût-traffic.

La méthode proposée de Florian et Nguyen se basait toutefois sur une connaissance préalable des paramètres associées aux termes d'entropie. Boyce et al. [1983], dans une formulation hiérarchisée de la demande, suggérèrent une procédure alternant la solution du programme et la calibration des multiplicateurs de Lagrange liés aux contraintes d'entropie

(ces multiplicateurs sont équivalents aux paramètres ci-haut mentionnés). Dans une modélisation de la demande de transport, une telle procédure est convenable puisqu'elle suffit à évaluer cette demande selon une occupation du sol donnée et des réseaux prédéterminés. Néanmoins, en tant que simple étape d'un algorithme d'un problème de conception, elle s'avère plutôt complexe. Examinons quelques possibilités de simplification.

La calibration des multiplicateurs lagrangiens semblent particulièrement alourdir la résolution du sous-programme d'affectation, de répartition et de distribution. Deux des multiplicateurs lagrangiens sont en fait les paramètres associés respectivement à un modèle logit de partage modal et à un modèle gravitaire. La fixation des valeurs à priori faciliterait grandement la recherche de la solution de (4.32-4.41) mais on doit recourir à des contraintes d'entropie car la modification du réseau influence largement ces paramètres [Erlander, Stewart, 1978]. Toutefois, comme le soulignèrent Boyce et Janson, il n'est pas plus sûr que l'entropie d'une distribution demeure constante au gré des changements d'un réseau sous-jacent. En effet, Erlander [1977] jugea que l'entropie d'une distribution O-D s'assimilait à une mesure d'accessibilité globale d'un système; le développement d'un réseau augmente cette accessibilité si bien que l'entropie peut difficilement prétendre à une exogénéité parfaite. Quant à l'entropie des proportions modales, elle dépend implicitement, tout comme le paramètre modal, des coûts comparatifs des modes pour chaque trajet.

Evidemment, ces difficultés se dissipent si les paramètres spatial et modal sont préalablement connus. De cette manière, les contraintes d'entropie pourraient être substituées par des termes équivalents dans la fonction objective du modèle de demande. L'interposition alourdissante de phases de calibration dans l'algorithme de solution de la demande n'est alors plus nécessaire.

4.4 CONCLUSION SUR LE CADRE DE MODELISATION DU TRANSPORT

Le chapitre 4 a rassemblé les éléments permettant de parvenir à une approche unifiée du transport urbain. Il apparaît sans aucun doute comme un dénouement de cette partie consacrée à la modélisation (statique) du transport. En effet, si les modèles de demande peuvent se passer de ceux d'offre, ces derniers peuvent difficilement être formulés sans recourir à une composante variable de la demande. Les problèmes de modification optimale de l'offre se fondent sur quelques paramètres exogènes mais surtout sur la performance anticipée, elle-même conditionnée par l'état de la demande associé à l'offre modifiée.

Comme il a été démontré dans la section 4.1.3, la résolution d'un problème d'optimisation de l'offre passe habituellement par la résolution d'un sous-problème de demande. Le problème discret de conception de réseau est un exemple supportant cette conceptualisation. Au niveau supérieur, on imite le comportement des aménagistes qui consiste en des gestes discrets (améliorations d'un réseau), posés dans une perspective de long terme, inspirés par un ou des objectifs, restreints par des limites budgétaires et basés sur une demande prédite selon des paramètres actuels. Au niveau inférieur, on reproduit le comportement des voyageurs qui comprend plusieurs phases de décisions hiérarchisées mais non-successives. Parmi les facteurs qui caractérisent ces phases, on notera la dispersion des choix empêchant un comportement totalement optimal, la désagrégation des voyageurs (par trajet O-D, par mode et par chemin) traduisant une rationalité fondée sur des conditions locales de transport et non globales ainsi que la courte perspective temporelle des décisions.

Du fait que la demande de transport réagit promptement aux changements de son environnement (conditions de transport, occupation du sol), il n'y a pas lieu de préconiser une modélisation dynamique. Le comportement d'offre s'accommode aussi d'un cadre statique car les décisions d'aménagement de transport s'appuient généralement sur (1) un budget de dépense étalé sur une période fixe et (2) sur une prédiction de la demande réagissant aux seules modifications projetées. Il s'agit alors d'un cadre «statique-comparatif». Le plus souvent, la distribution géographique projetée des activités est quand même considérée non-constante car on y apporte des cor-

rections relatives au plan d'aménagement en cours et aux développements prévus (ou espérés).

Pourtant, les théories de localisation stipulent unanimement que la distribution spatiale des activités est un produit des interactions spatiales suscitées par la complémentarité de ces activités. Cette relation rétroactive reste négligée dans un contexte d'aménagement de transport car elle s'exerce à moyen et long terme et, par conséquent, échappe à la période d'analyse, usuellement courte, favorisée par les planificateurs du transport. Dans ce contexte véritablement dynamique, il est parfaitement concevable que les modifications de réseau jugées optimales à l'intérieur de cette période d'étude, provoquent une reorganisation spatiale des activités telle qu'elles ne se révéleront pas les meilleurs choix. Les exemples d'écart entre les effets escomotés et obtenus foisonnent dans le domaine du transport urbain.

Dans ce cas, le cadre statique-comparatif convient particulièrement ici puisque nous désirons reproduire les comportements des voyageurs et des planificateurs de transport. Avant de passer à l'action, ces derniers tentent habituellement d'évaluer les impacts probables en terme de demande de transport mais rarement il peuvent mesurer les conséquences à long terme sur le développement urbain. Leur démarche reste donc foncièrement statique-comparative et, pour cette raison, nous adoptons un cadre qui y est conforme.

Deuxième partie

**MODELISER LA
STRUCTURE URBAINE**

Chapitre 5

ANALYSE DES SYSTEMES SPATIO-TEMPORELS

5.1 INTRODUCTION

La recherche d'un cadre de modélisation spatio-dynamique de la structure urbaine demande, dans un premier temps, la définition d'une représentation de processus. Cette représentation, il va sans dire, sollicite le langage mathématique mais doit spécifiquement s'accommoder de relations se déployant tant au niveau des variables, du temps et de l'espace.

Le modèle de régression est une représentation très connue d'un processus causal ou associatif s'exerçant d'une variable à l'autre. Bien que les observations puissent être des unités spatiales, il ne s'agit pas d'un modèle spatial puisque les relations se cantonnent dans le même lieu. Les modèles de séries temporelles décrivent des processus où les variables sont conditionnées par leurs valeurs passées. Ces liaisons «découlées» se transmettent ordinairement au sein d'une même variable mais peuvent s'appliquer à des processus bivariés. A cet égard, l'analyse régressive et celle des séries temporelles peuvent être entreprises conjointement. Analogiquement aux séries temporelles, des modèles linéaires d'interactions spatiales imitent des processus purement géographiques où les relations se produisent par des effets de voisinage. Ces modèles spatiaux n'exhibent habituellement de références au temps [voir Cliff, Ord, 1973].

La représentation requise conjugue la structure des modèles émanant des trois dimensions d'un processus, variables, temps, espace [Sheppard, 1979]. En fait, ces structures sont compatibles si une additivité des composantes est présumée; cette restriction ne s'impose pas aux composantes mêmes. Plus que théorique, cette représentation semble opérationnelle comme l'ont témoigné Hordjik et Nijkamp [1977; 1978], Bennett et Chorley [1978] et surtout Bennett [1979]. Ce domaine de modélisation souvent appelé "Spatial Time Series" à cause de ses fondements reposant sur l'analyse des séries temporelles de Box et Jenkins [1970], peut aussi s'appeler «analyse des systèmes spatio-temporels» car les processus décrits sont assimilés au fonctionnement d'un système.

Le chapitre présent démontrera les possibilités et la versatilité d'une telle analyse. D'abord, nous traiterons des systèmes purement temporels. La compréhension de tels systèmes reste fondamentale car leur structure particulière conditionne la représentation de tout système référant au temps. Ensuite, sera abordée la représentation générale des systèmes spatio-temporels. A ce stade, la composante spatiale demeure très abstraite mais fort complexe. Avec des méthodes de spécification et le recours à des modèles purement spatiaux, la composante géographique prend forme et se marie plus facilement à la désagrégation temporelle.

Enfin, un sous-chapitre se consacrera à l'estimation des paramètres. Des méthodes conventionnelles comme les moindres carrés ordinaires, multivariés et généralisés de même que la vraisemblance maximale se prêtent assez bien à cette phase estimative. Toutefois quelques précautions doivent être prises concernant la combinaison des structures spatiales et temporelles d'un système.

5.2 LES SYSTEMES PUREMENT TEMPORELS

5.2.1 Introduction

Les systèmes purement temporels se décrivent ici comme étant des systèmes univariés, bivariés ou multivariés possédant des observations temporelles équidistantes et qui sont sujets à des relations à travers le temps. Ils ne réfèrent pas à une dimension spatiale, les interactions du système ne pouvant être désagrégées géographiquement. La théorie supportant l'«analyse des séries temporelles» ("Time Series Analysis") apparaît relativement articulée. Parmi les plus importantes contributions à l'élaboration de cette théorie, il faut noter les ouvrages classiques de Jenkins et Watts [1968] sur l'analyse spectrale et surtout de Box et Jenkins [1970] sur l'analyse des séries temporelles. L'analyse purement temporelle reste encore aujourd'hui un domaine d'investigation privilégié en économétrie car de telles analyses sont essentielles dans un cadre prévisionnel. On trouve dans Johnston [1972] une excellente revue des méthodes économétriques parmi lesquelles les méthodes liées au temps tiennent une place prépondérante.

Box et Jenkins [1970] exposèrent les principaux types de modèles du temps comme étant ceux qui décrivent les processus autoregressifs (AR), de moyennes mobiles (MA pour "moving average"), autoregressifs - moyennes mobiles (ARMA) et les fonctions de transfert (TF).

5.2.2 Processus autorégressifs

Le processus autorégressif linéaire reste l'un des plus simples car, étant univariés, il n'évoque qu'une dépendance entre les observations de périodes différentes. Faut-il préciser que l'objet d'analyse est une série chronologique ou «chronique» qui est, tout simplement, une suite d'observations dans le temps, par exemple:

$$\{y_t\} = y_1, y_2, \dots, y_{t-1}, y_t$$

Un modèle AR temporel a spécifiquement des relations de décalage ("lag") i.e. des dépendances d'une observation quelconque envers une ou des observations antérieures. Le modèle impose une restriction particulière, la stationnarité, sur la série. Un processus stationnaire remplit les trois

conditions suivantes:

$$E\{y_t\} = 0 \quad \forall t; \text{ la moyenne est nulle et constante}$$

$$E\{(y_t)^2\} = \sigma_y^2 \quad \forall t; \text{ la variance est constante}$$

$$E\{y_t y_{t-K}\} = A_V(K) \quad \forall t, K; \text{ l'autocovariance est décrite par la fonction d'autocovariance } A_V(K)$$

Ces restrictions impliquent que la série n'a pas de tendance (la nullité de la moyenne peut être facultative), qu'elle est homoscedastique (les variations sont généralement d'amplitudes régulières) et que l'autocorrélation ne dépend du rang de l'observation (i.e. t) mais uniquement du décalage K . Nous reparlerons plus attentivement de l'autocorrélation. Si une série ne rencontre pas les deux premières conditions, il y a diverses méthodes de correction [Philips, Blomme, 1979]. Un processus autoregressif est alors décrit par le modèle générique suivante:

$$y_t = \omega_1 y_{t-1} + \omega_2 y_{t-2} + \dots + \omega_p y_{t-p} + u_t$$

$$y_t = \sum_{k=1}^p \omega_k y_{t-k} + u_t \quad (5.1)$$

La série $\{\omega_k\}$ contient des paramètres autoregressifs. p est l'ordre du processus et le terme u_t désigne une série de «bruits blancs» i.e. une série complètement aléatoire. Le modèle est tel que le processus $\{y_t\}$, d'une part, est décrit par son passé et, d'autre part, comprend une variation aléatoire. La composante de perturbation, par définition, n'est pas autocorrélée:

$$E\{u_t u_{t-K}\} \begin{cases} = \sigma_u^2 & K = 0 \\ = 0 & K \neq 0 \end{cases}$$

La fonction d'autocovariance $A_V(K)$ est dotée d'arguments discrets symbolisant les décalages K . Le concept d'autocovariance s'inspire du concept de covariance de deux séries distinctes. Il s'agit plutôt de mesurer la variance commune entre deux observations séparées d'un écart K . Cependant, il s'avère souvent plus commode de raisonner sur les «autocorrélations» plutôt que sur les autocovariances. Le passage de l'un vers l'autre se fait aussi aisément que de la covariance vers la corrélation. Il suffit de standardiser la mesure en divisant par la variance de la série (la variance étant constante dans une série stationnaire). Ainsi la fonction d'autocorrélation empirique $\rho_V(K)$ s'écrit comme suit:

$$\rho_V(K) = \frac{E\{y_t y_{t-K}\}}{E\{(y_t)^2\}} = \frac{T \cdot \sum_{t=1+K}^T y_t y_{t-K}}{(T-K) \cdot \sum_{t=1}^T (y_t)^2} \quad (5.2)$$

D'après cette définition, il semble bien évident que $\rho_V(0) = 1$ du fait que $A_V(0) = \text{Var}\{y_t\}$. De plus, aucune valeur de la fonction ne peut dépasser l'unité car aucun croisement décalé de la série peut avoir une covariance supérieure à celle de la série par elle-même. Une autre propriété fondamentale est que $\rho_V(K)$ est symétrique tel que $\rho_V(K) = \rho_V(-K)$. La forme graphique de l'autocorrélation dans un intervalle $[-p, p]$ ressemble donc à une pyramide dotée de lobes. Il faut aussi rappeler que des valeurs négatives de $\rho_V(K)$ sont fréquentes et traduisent des relations inverses entre des observations.

L'intérêt premier de la fonction d'autocorrélation réside dans l'identification du modèle autoregressif [Box, Jenkins, 1970 chap.5]. En pratique, rares sont les processus temporels qui ne peuvent pas être décrits par un nombre restreint de décalages. Le procédé d'identification consiste à détecter les décalages les plus significatifs dans la description d'une série. Cette étape doit précéder celle de l'estimation car celle-ci s'occupe d'évaluer les paramètres selon une structure paramétrique préalablement choisie. Cette structure est souvent appelée «structure de mémorisation» [Bennett, 1979] car elle réfère à la mémoire qu'a un système envers son passé.

Pourtant, il ne suffit pas de retenir les termes de (5.1) dont le décalage correspondant de la fonction d'autocorrélation possède une valeur absolue élevée. Il faut se rabattre, en général, aux estimates d'autocorrélation partielle qui, à la différence des premiers, enlèvent les effets indirects [Box, Jenkins, 1970 chap.5]. Prenons, par exemple, un processus strictement de premier ordre:

$$y_t = \omega_1 y_{t-1} + u_t \quad (5.3)$$

L'autocovariance $A_V(1)$ se calcule ainsi:

$$\frac{1}{T-1} \sum_{t=2}^T y_t y_{t-1} = \frac{1}{T-1} \sum_{t=2}^T (\omega_1 y_{t-1} + u_t) y_{t-1} = \frac{\omega_1}{T-1} \sum_{t=2}^T y_{t-1} y_{t-1} \quad (5.4)$$

$$A_V(1) \approx \omega_1 \text{Var}\{y_t\} \text{ donc } \rho_V(1) = A_V(1) / \text{Var}\{y_t\} \approx \omega_1$$

Théoriquement, le paramètre autoregressif est égal à l'estime d'autocorrélation du premier degré. Le symbole $\approx$ «approximatif» est utilisé car l'estime de covariance a une observation de moins de celui de la variance et ne peut donc pas être strictement équivalent. Par contre, si N est grand, la distorsion devient infime si bien que $p_Y(1)$ est un bon estimateur de ω_1 . Cette équivalence implique que le paramètre autoregressif est borné à l'intervalle unitaire ouvert, soit $] -1, 1[$. On démontre ici que les $p_Y(K)$; $K \geq 2$ ne peuvent se considérer des estimateurs de ω_K ; $K \geq 2$, ni même servir à l'identification des termes significatifs. Reprenons (5.3) et développons-le:

$$y_t = \omega_1 y_{t-1} + u_t \quad \text{de même que} \quad y_{t-1} = \omega_1 y_{t-2} + u_{t-1}$$

$$\text{et } y_{t-2} = \omega_1 y_{t-3} + u_{t-2} \quad \text{et ainsi de suite} \quad (5.5)$$

$$\text{alors } y_t = \omega_1 (\omega_1 y_{t-2} + u_{t-1}) + u_t = \omega_1 (\omega_1 (\omega_1 y_{t-3} + u_{t-2}) + u_{t-1}) + u_t$$

Le développement (5.5) est possible du fait que $\{y_t\}$ étant stationnaire, l'autocorrélation dépend seulement du décalage et non de t . Dans la forme initiale (5.3), les paramètres de degrés supérieurs, soient ω_2 et ω_3 , sont présumés nuls. On remarque néanmoins que des modèles du deuxième et troisième degrés deviennent vraisemblables d'après ci-dessus:

$$y_t = (\omega_1)^2 y_{t-2} + \varepsilon_t \quad \text{où } \varepsilon_t = \omega_1 u_{t-1} + u_t \quad (5.6a)$$

$$y_t = (\omega_1)^3 y_{t-3} + \varepsilon_t \quad \text{où } \varepsilon_t = (\omega_1)^2 u_{t-2} + \omega_1 u_{t-1} + u_t \quad (5.6b)$$

La fonction d'autocorrélation détectera ces effets indirects découlant de la relation du premier ordre mais l'attribuera aux estimates correspondant aux décalages $K=2$ et $K=3$. En effet, dans l'exemple précédent, il suffit de procéder comme dans (5.5-5.6) pour s'apercevoir que:

$$p_Y(2) \approx (\omega_1)^2 \quad \text{et} \quad p_Y(3) \approx (\omega_1)^3$$

$$\text{ainsi } p_Y(K) \approx (\omega_1)^K$$

La fonction d'autocorrélation d'un processus animé par une relation du premier ordre aura la forme d'une série géométrique décroissante puisque $\omega_1 \leq 1$. Pour détecter les vraies autocorrélations du 2^e degré, il faut, par conséquent, soustraire la part due au 1^{er} degré. Un coefficient partiel pour $K = 2$, symbolisé par $\tilde{p}_Y(2)$, s'écrira ainsi:

$$\hat{p}_v(2) = p_v(2) - [p_v(1)]^2$$

Ceci vaut pour tous les ordres mais cela devient complexe lorsque le processus est animé par une structure de memorisation élaborée. On trouve dans Box et Jenkins [1970 chap.5] une méthode de calcul pour les coefficients partiels. Heureusement, en pratique, les relations les plus fréquemment détectées sont associées à des ordres peu élevés sinon très évidents (délai de douze mois pour les séries économiques mensuelles).

A partir d'ici, on fera usage d'une notation conventionnelle qui simplifie passablement les formulations. Il s'agit de l'«opérateur de décalage» ou «opérateur de retard» "z" (en anglais "lag operator" ou "backshift operator"). Il se définit comme suit:

$$y_{t-1} = zy_t \quad y_{t-2} = z^2y_t \quad y_{t-k} = z^ky_t$$

Cet opérateur se manipule au sein d'une fonction comme une variable habituelle [Goldberg, 1958]. Plutôt qu'une constante mathématique, z doit être vu comme une représentation algébrique d'un processus de translation. On reprendra la démonstration précédente à l'aide de z:

$$y_t = \omega_1 zy_t + u_t$$

$$y_t - \omega_1 zy_t = u_t$$

$$y_t = (1 - \omega_1 z)^{-1} u_t \quad (5.7)$$

Le lecteur notera la similitude entre cette fonction et les multiplicateurs économiques (voir Samuelson, 1976). On sait, par ailleurs, que le terme inversé exprime une série géométrique convergente du fait que $|\omega_1| < 1$:

$$y_t = \sum_{k=0}^{\infty} (\omega_1)^k z^k u_t = (1 + \omega_1 z + \omega_1^2 z^2 + \omega_1^3 z^3 + \dots) \cdot u_t \quad (5.8a)$$

$$= H(z)u_t = u_t + \omega_1 u_{t-1} + \omega_1^2 u_{t-2} + \dots \quad (5.8b)$$

La fonction $H(z)$ est un polynôme de degré infini de z et est appelée «fonction de transfert» car elle convertit (convolue) la série aléatoire en une série $\{y_t\}$. La transformation d'un processus autoregressif en un processus de moyenne mobile (MA) comme (5.8b) via l'opérateur z se nomme «transformée de z». On peut donc complètement décrire une chronique par son terme d'erreur actuel et un polynôme autorégressant les erreurs antérieures. La transformée de z convient à tout processus AR:

$$y_t = \sum_{k=1}^p \omega_k z^k y_t + u_t = \Omega(z) y_t + u_t \quad (5.9a)$$

$$y_t = (1 - \Omega(z))^{-1} u_t = (1 - \xi_1 z - \xi_2 z^2 - \dots) \cdot u_t \quad (5.9b)$$

La fonction $(1 - \Omega(z))^{-1}$ est une forme générale d'une fonction de transfert. Dans un processus AR de 1^{er} ordre, les paramètres sont telles que $\xi_1 = -(\omega_1)$, $\xi_2 = -(\omega_1)^2$, etc... Dans un processus de 2^e ordre, AR(2), la fonction de transfert adopte la forme d'une exponentielle-négative dotée d'oscillations sinusoidales qui s'atténuent progressivement. Toutefois, la stationnarité du processus AR(2) requiert des restrictions différentes d'un AR(1) quant aux valeurs paramétriques: l'invertibilité polynômiale en dépend.

5.2.3 Processus de moyennes mobiles

Un modèle de moyennes mobiles d'ordre fini peut être spécifié. Il s'agit d'une formulation servant à décrire un processus à partir des bruits. Cette forme de modélisation est très usitée dans l'analyse de certains procédés industriels où les perturbations avec leurs effets directs et retardés permettent d'obtenir une série temporelle très régulière. Cette régularité tient particulièrement à la durabilité des effets. En effet, une impulsion isolée peut, selon le processus de retard envisagé, se répercuter très longtemps si bien que la série qui en dépend montrera une évolution non-accidentelle. Les effets à retardement provenant d'une impulsion unique sont exprimés dans une «fonction impulsion-réponse» avec comme arguments les périodes suivant l'impulsion. Cette fonction sera discutée plus tard.

Un modèle MA(q) général prend l'allure suivante:

$$y_t = u_t - \xi_1 u_{t-1} - \xi_2 u_{t-2} - \dots - \xi_p u_{t-p} = H(z) u_t \quad (5.10)$$

Cette fois, la fonction polynômiale $H(z)$ est une fonction de transfert d'ordre fini (=q). Box et Jenkins [1970 chap.3] montreront qu'un processus MA(q) se convertit en processus AR(∞) tout comme il a été possible de faire avec un AR(p) vers un MA(∞) (voir 5.7, 5.8b). En effet:

$$y_t = H(z) u_t \quad \text{alors} \quad H^{-1}(z) y_t = u_t \quad (5.11)$$

On sait que:

$$H(z) = 1 - \xi_1 z - \xi_2 z^2 - \dots - \xi_n z^n \quad (5.12a)$$

donc $H^{-1}(z)$ a une forme polynômiale de degré infini comme suit:

$$(1 - \omega_1 z - \omega_2 z^2 - \omega_3 z^3 - \dots) \equiv 1 - \Omega(z) \quad (5.12b)$$

$$y_t = \sum_{k=1}^{\infty} \omega_k z^k y_t + u_t = \Omega(z) y_t + u_t \quad (5.13)$$

L'invertibilité des processus autorégressifs et de moyennes mobiles permet d'utiliser alternativement les deux formes. Cependant, on convient qu'il reste préférable de préconiser la forme qui n'implique pas des décalages à l'infini. En pratique, heureusement, on peut représenter maintes séries avec un ordre ne dépassant pas deux sous l'une ou l'autre des deux formes.

Un modèle ARMA(p,q), souvent référé au nom «modèle Box-Jenkins», a la prétention de représenter n'importe quel processus temporel stationnaire car il combine une composante autorégressive et une de moyennes mobiles. De cette façon, les effets de mémorisation de la série sont strictement contenus dans la part AR tandis que les effets retardés des perturbations reviennent à la part MA. On utilise donc un minimum de paramètre pour le modèle car on attribue les relations temporelles aux véritables sources de variation. Un modèle ARMA(p,q) se présente ainsi:

$$y_t = \sum_{k=1}^p \omega_k z^k y_t + u_t - \sum_{k=1}^q \xi_k z^k u_t = \Omega(z) y_t + H(z) u_t \quad (5.14a)$$

$$(1 - \Omega(z)) y_t = H(z) u_t \quad (5.14b)$$

La fonction de transfert appropriée est une division polynômiale de z:

$$y_t = \frac{H(z)}{(1 - \Omega(z))} u_t \quad (5.14c)$$

Une division de polynômes de z n'est que soluble que sous la forme de somme infinie de termes z. Cette fonction de transfert expose donc, via son numérateur, les effets à retardement des impulsions aléatoires et, via son dénominateur, les effets indirects de la mémorisation du terme y_t . La mémorisation (issue du processus AR) peut donc se traduire en terme de la variable de perturbation; par exemple, si $y_t = f(y_{t-1}, u_t, u_{t-k})$, alors $y_t = f(y_{t-2}, u_{t-1}, u_{t-k-1})$ et ainsi de suite. De plus, chacun des termes décalés de y_t est lui-même fonction de ses erreurs correspondantes. Il en résulte un effet durable des impulsions aléatoires qui s'estompera (si la série est

stationnaire) au rythme de l'intensité de la mémorisation.

5.2.4 Processus de décalage distribué et fonctions de transfert

Dans le domaine des sciences humaines, les phénomènes purement aléatoires restent tout de même peu fréquents [voir Bennett, Chorley, 1978]. Ainsi, les modèles de moyennes mobiles paraissent peu employés pour l'explication d'une série. Pourtant, en économétrie, on est conscient depuis longtemps du rôle non-négligeable de l'autocorrélation du terme d'erreur d'un modèle de régression. Pour soustraire cette autocorrélation, on considère le résidu comme un processus autoregressif doté d'une véritable erreur aléatoire [Cochrane, Orcutt, 1949]. Cette représentation ressemble à la forme MA. Par exemple:

$$y_t = a + bx_t + \varepsilon_t \quad \text{où} \quad \varepsilon_t = p_1 \varepsilon_{t-1} + u_t \quad (5.15a)$$

$$\text{ou} \quad y_t - a - bx_t = p_1 \varepsilon_{t-1} + u_t \quad (5.15b)$$

x_t est un régresseur, ε_t est le résidu de la régression et u_t est une composante aléatoire. Pourtant, il ne s'agit pas d'un processus MA car la série $\{\varepsilon_t\}$ n'est pas aléatoire puisqu'autocorrélée.

Un véritable modèle de moyennes mobiles est une fonction autoregressive de la composante aléatoire. Or, cette composante est inconnue. Une procédure particulière de déconvolution existe pour reproduire u_t d'après y_t [voir Bennett, 1979 chap.3]. Pour décrire les systèmes temporels sociaux, on préfère des processus qui sollicitent plutôt des variables explicatives exogènes. Le modèle de régression classique appliqué à une chronique recourt à une ou des variables indépendantes prises à leurs valeurs actuelles. Ainsi:

$$y_t = a_0 + a_1 x_t + \varepsilon_t \quad (5.16)$$

Il s'agit d'une régression linéaire simple. La présomption qui y est sous-jacente consiste en une dépendance linéaire de la série $\{y_t\}$ à l'égard de la série $\{x_t\}$. On suppose aussi que la transmission se fait simultanément (ou à l'intérieur d'une même période). Cette hypothèse ne se vérifie pas lorsque les unités de temps sont rapprochées ou si le processus de transmission est lent. Pourtant, dans de tels cas, le modèle sera probablement

valide statistiquement du fait que la série $\{x_t\}$ étant autocorrélee, la relation détectée sera non-nulle. Ainsi, le décalage réel L de la relation $\{x_{t-L}\} \rightarrow \{y_t\}$ est reconduit partiellement à $\{x_t\} \rightarrow \{y_t\}$ du fait que $E\{x_t x_{t-L}\} \neq 0$.

Dans les phénomènes sociaux et économiques, il semble fort probable que les relations de dépendance soient décalées. Le délai symbolise les retards qui accompagnent les changements d'attitude des individus ou bien l'inertie d'un mécanisme collectif (firme, institution, gouvernement). On appelle «processus de délai» ce genre de phénomène qui se formule de la façon suivante [Bennett, 1979 chap.2]:

$$y_t = \alpha_0 + \alpha_L x_{t-L} + \varepsilon_t \quad (5.17)$$

En fait, il y a très peu de différence conceptuelle entre ce modèle de régression avec délai et celui présenté en (5.16). Les séries apparaissent seulement croisées avec un décalage fixe, soit L . Théoriquement, ce modèle reste un cas particulier de modèle de «décalage distribué» (de l'anglais "distributed lag"):

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_q x_{t-q} + u_t \\ &= \alpha_0 + \sum_{k=1}^q \alpha_k x_{t-k} + u_t \end{aligned} \quad (5.18)$$

Le processus de décalage distribué autorise la modélisation d'effets étalés sur plus d'une période. Le délai de réaction de y_t à l'égard de x_t n'est plus fixe et constant. On représente ainsi la variabilité de délai de réaction qui se manifeste, dans un système socio-économique, par la diversité des acteurs et des moyens de transmission.

L'identification des délais significatifs s'effectue analogiquement à celle des processus autoregressifs. Au lieu d'une fonction d'autocorrélation $\rho_v(K)$, on a une «fonction de corrélation croisée» $\rho_{vx}(K)$. Celle-ci, de par sa morphologie, s'inspire des concepts d'autocorrélation et de corrélation. La fonction, symétrique dans un intervalle centré sur 0, consiste en un rapport de covariance entre les séries décalées sur les écarts-types [Jenkins, Watts, 1968 chap.5]:

$$\rho_{vx}(K) = \frac{\text{Cov}(y_t x_{t-K})}{[\text{Var}(y_t) \cdot \text{Var}(x_t)]^{1/2}} \quad (5.19)$$

$$\text{où } \text{Cov}(y_t, x_{t-k}) = \frac{1}{T-k} \sum_{t=k+1}^T y_t x_{t-k} \quad \text{Var}(y_t) = \frac{1}{T} \sum_{t=1}^T (y_t)^2$$

$$\text{Var}(x_t) = \frac{1}{N} \sum_{t=1}^N (x_t)^2 \quad \text{si } E\{x_t\} = \bar{x} \neq 0, \text{ remplacer } x_t \text{ par } x_t - \bar{x}$$

Si aucun processus autoregressif anime la serie $\{y_t\}$, la fonction de correlation croisee peut être utilisée directement pour identifier les décalages les plus significatifs. Il n'y a alors plus de phénomène de mémorisation chez la variable dependante et donc plus d'effets indirects.

Sous les presomptions usuelles concernant la distribution du terme d'erreur (moyenne nulle, variance constante, non-autocorrelé) de l'indépendance de $\{x_t\}$ et $\{u_t\}$, il n'y a pas en principe de problèmes additionnels d'estimation des paramètres α_k [Johnston, 1972 chap.10]. La methode des moindres carrés fournira des estimateurs sans biais et de variance minimale. Pourtant, des difficultés surviennent si la fonction $\rho_{yx}(K)$ nous livre un nombre important de décalages significatifs. Cela risque, du point de vue de l'estimation, de réduire considérablement les degrés de liberté et, du point de vue de la modélisation, de spécifier un modèle lourd. Koyck [1954] proposa une methode palliant à ce problème en remplaçant la serie de coefficients $\{\alpha_k\}$ par un seul mais seconde par une serie geometrique declinante de z munie d'un paramètre. Selon cette procédure, on étale régressivement la ponderation à travers plusieurs décalages.

Le schème de Koyck convient spéciquement aux processus dont les relations décalées sont fortement et régulièrement étalées à partir d'un délai particulier. Cette régularité apparaît un signe révélateur d'un processus autoregressif chez la variable dependante. Par exemple, supposons un modèle combinant un terme autoregressif du premier ordre et une fonction de délai unique:

$$y_t = \omega_1 y_{t-1} + \alpha_L x_{t-L} + u_t \quad (5.20)$$

$$y_t - \omega_1 z y_t = \alpha_L x_{t-L} + u_t$$

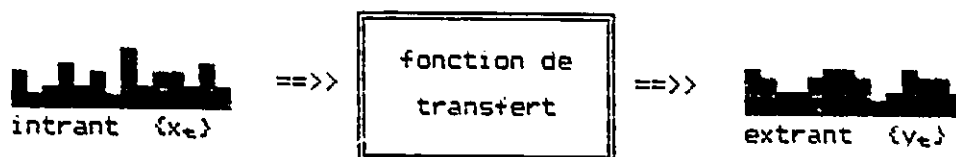
$$y_t = (1 - \omega_1 z)^{-1} (\alpha_L x_{t-L} + u_t) = \sum_{k=0}^{\infty} (\omega_1)^k z^k (\alpha_L x_{t-L} + u_t) \quad (5.21a)$$

$$\equiv \alpha_L x_{t-L} + \alpha_{L+1} x_{t-L-1} + \alpha_{L+2} x_{t-L-2} + \dots + \Omega(z) u_t \quad (5.21b)$$

où $\alpha_{L+k} = \alpha_L (\omega_1)^k$; $k=0,1,2,\dots$

La fonction $(1-\omega_1 z)^{-1} \alpha_L z$ transforme la série exogène $\{x_t\}$ en série $\{y_t\}$ moins la composante aléatoire. Il s'agit presque d'une fonction de transfert. Dans le formalisme de la modélisation de transfert, on ne considère pas le terme stochastique (du moins pas séparément) mais un «intrant» ("input"). L'intrant peut être une variable exogène seule ou assistée d'une modulation aléatoire. La variable dependante est désormais nommée «extrant» ("output"). Le système dynamique comprend une série temporelle intrant et une autre extrant ainsi qu'une fonction de transfert convoluant la première vers la dernière [Box, Jenkins, 1970 chap.10; Bennett, Chorley, 1978]:

Figure 5.1 Représentation d'une fonction de transfert temporelle



Un modèle temporel de transfert peut être formulé de la façon générale suivante. Initialement, on l'écrit sous une forme autoregressive et de décalage distribué:

$$y_t = \sum_{k=1}^p \omega_k z^k y_t + \sum_{k=0}^q \alpha_k z^k (z^L x_t) = \Omega(z) y_t + A(z) (z^L x_t) \quad (5.22)$$

$$y_t = \left[(1 - \Omega(z))^{-1} A(z) z^L \right] \cdot x_t \quad (5.23a)$$

$$= (v_L z^L + v_{L+1} z^{L+1} + v_{L+2} z^{L+2} + \dots) \cdot x_t \equiv \underline{v}(z) x_t \quad (5.23b)$$

Les relations directes entre x_t sont sujettes à un délai de longueur L . Ce retard correspond au temps minimal de transmission d'une impulsion. La fonction entre crochets de (5.23a) apparaît la véritable fonction de transfert du système. Le développement de la fonction en termes polynômiaux de z est une série infinie. Si cette série converge, alors le système est dit stable. Les coefficients $v_0, v_1, v_2, \dots$ forment une fonction particulière qu'on appelle «fonction d'impulsion-réponse (IR)». Chacun d'eux évalue l'effet qu'une unité d'intrant produit sur l'extrant pour un décalage donné. Ils traduisent le degré de persistance d'une impulsion dans le système.

Dans l'exemple ci-haut, les coefficients $v_k = 0$ pour $k < L$: de ce fait, une valeur x_{40} engendre des réactions de la part de $\{v_k\}$ qu'à partir de $t = 40+L$. L'étude des différentes fonctions d'impulsion-réponse dérivées de modèles simples reste intéressante. Selon les combinaisons d'ordre des processus, $\Omega(z)$ et $A(z)$, on retrouve divers comportements allant de la série géométrique décroissante à la série sinusoïdale déclinante. Souvent, la fonction IR denote à ses débuts ($k \geq L$) un sursaut ("overshoot"), signalant une amplification retardée de la réponse. L'évaluation de la fonction IR exige la solution du rapport polynomial inscrit dans la fonction de transfert. Dans Box et Jenkins [1970 chap.10], on décrit adéquatement les comportements algébriques des fonctions IR en rapport avec les types de fonction de transfert.

Si la fonction IR est convergente (condition d'un système stable) et si l'intrant est tenu constant ($x_t = x; \forall t$), alors l'extrant atteindra éventuellement une valeur d'équilibre. Par exemple:

$$y_{\infty} = \left[(1 - \Omega(z))^{-1} A(z) z^L \right] \cdot x = \underline{y}(z) x \quad (5.24)$$

$$= x \cdot \sum_{k=0}^{\infty} v_k \equiv g \cdot x \quad (5.25)$$

Dès lors, l'opérateur de retard z n'est plus requis puisque l'impulsion devient constante. La valeur d'équilibre apparaît le produit de x avec une constante g équivalant à la somme des poids v_k . Cette sommation de la fonction d'impulsion-réponse s'appelle le «gain d'état soutenu» et interprète la contribution totale d'une unité de la série impulsive dans le système. Le gain reste facile à calculer car il n'implique pas les opérateurs de retard z . Dans l'exemple (5.22-5.23):

$$g = \frac{\alpha_L + \alpha_{L+1} + \dots + \alpha_q}{1 - \omega_1 - \omega_2 - \dots - \omega_p} \equiv \sum_{k=0}^{\infty} v_k \quad (5.26)$$

5.2.5 Systèmes temporels multivariés

La structure d'un système temporel multivarié diffère peu, selon Bennett [1979 chap.2] de celle d'un système bivarie à un intrant et à un extrant. Le modèle de régression multiple reste un exemple bien connu de système multivarie. Dans ce cas, la variable dépendante désagrégée temporellement figure de «réponse» tandis que les «impulsions» sont symbolisées

par les régresseurs et le terme résiduelle. Quant à la fonction de transfert, elle est interprétée par le vecteur des coefficients de régression:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_S x_{St} + \varepsilon_t \quad (5.27)$$

Cependant, ce modèle demeure un cas particulier de systèmes temporels car il ne sollicite aucune structure de mémorisation ou de décalage. De plus, la variable dépendante reste unique. Dans une représentation généralisée de système temporel multivarie, il n'y a pas de composante résiduelle (concept associé à la régression) et non plus restriction quant au nombre d'extrants. Voici un système complet:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Rt} \end{bmatrix} = \begin{bmatrix} \sigma_{11}(z) & \sigma_{12}(z) & \dots & \sigma_{1S}(z) \\ \sigma_{21}(z) & \sigma_{22}(z) & \dots & \sigma_{2S}(z) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{R1}(z) & \sigma_{R2}(z) & \dots & \sigma_{RS}(z) \end{bmatrix} \times \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{St} \end{bmatrix}$$

$$y_t = \sum_{r=1}^R \sum_{s=1}^S \sigma_{rs}(z) x_{st} \quad (5.28)$$

Le vecteur y_t est spécifique au temps t et contient R "outputs" différents tandis que le vecteur x_t comprend S "inputs". La «matrice de transfert» Σ possède des $R \cdot S$ entrées polynômiales $\sigma_{rs}(z)$. Chacun des polynômes de z n'est pas tenu d'avoir la même structure paramétrique, ni le même ordre et peut être nul. Il interprète l'effet d'une variable exogène s sur une variable endogène r . Par exemple, un système à deux intrants et deux extrants peut se représenter ainsi:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} (v_{11}^0 + v_{11}^1 z) & (v_{12}^0 z^2) \\ 0 & (v_{22}^1 z + v_{22}^2 z^2) \end{bmatrix} \times \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \quad (5.29)$$

Chacune des entrées de $\Sigma = \{\sigma_{rs}(z)\} 2 \cdot 2$ est individualisée de telle sorte que les relations bivariées en terme d'intensité et de structure de décalage y apparaissent propres à chaque paire de variables. Néanmoins, cette représentation semble simple du point de vue temporelle. Théoriquement, les polynômes $\sigma_{rs}(z)$ peuvent avoir un ordre infini: c'est le cas notamment quand existe une mémorisation du vecteur y_t :

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_{11}(z) & 0 \\ 0 & \Omega_{22}(z) \end{bmatrix} \times \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} + \begin{bmatrix} A_{11}(z) & A_{12}(z) \\ A_{21}(z) & A_{22}(z) \end{bmatrix} \times \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \quad (5.30)$$

dont la fonction de transfert pour chaque extrants est:

$$y_{1t} = \frac{A_{11}(z)x_{1t} + A_{12}(z)x_{2t}}{1 - \Omega_{11}(z)} \quad y_{2t} = \frac{A_{21}(z)x_{1t} + A_{22}(z)x_{2t}}{1 - \Omega_{22}(z)} \quad (5.31)$$

La matrice de transfert Σ correspondante a les elements:

$$\sigma_{rs}(z) = \frac{A_{rs}(z)}{1 - Q_{rr}(z)} = \frac{\alpha_{rs}^0 + \alpha_{rs}^1 z + \dots + \alpha_{rs}^p z^p}{1 - \omega_{rr}^1 z - \omega_{rr}^2 z^2 - \dots - \omega_{rr}^p z^p} \quad (5.32)$$

Comme on sait déjà, cette division de polynômes de termes de décalage donne une série infinie de z . Il s'agit donc de la matrice de transfert d'un système temporel multivarie doté d'un processus autoregressif et de décalage distribué. Cette représentation paraît sans doute très théorique mais peu pratique car elle implique des polynômes de z de degré infini. On pourrait tout simplement tronquer les polynômes de manière à retenir que les termes les plus significatifs. Usuellement, il reste préférable de conserver la forme initiale couplant autoregression et décalage distribue comme (5.30).

Dans le dernier exemple, les variables y_{1t} et y_{2t} entretenaient des relations autoregressives avec chacune d'elles. Pourtant, l'une d'elles pourrait dépendre des deux: dans ce cas, dans la matrice de droite de (5.30), un des deux zéros serait remplacé par un polynôme de z . Le système est alors évalué récursivement en calculant d'abord l'un pour ensuite calculer l'autre à partir du premier [Bennett, 1979 chap.2]. Quand relations autoregressives et regressives (avec décalages) s'exercent au sein des extrants, le calcul doit s'effectuer simultanément en recourant au calcul matriciel:

$$y_t = Q y_t + A x_t \quad \text{où } Q = \{Q_{rr}(z)\}_{R \times R}, \quad A = \{A_{rs}(z)\}_{R \times S} \quad (5.33)$$

$$y_t = [I - Q]^{-1} A x_t \quad (5.34)$$

La détermination de la matrice de transfert requiert l'inversion de la matrice $[I - Q]$ de dimension $R \times R$, opération qui fournit des éléments sous forme de rapports polynômiaux avec un dénominateur commun. Ce dénominateur est égal au déterminant de la matrice $[I - Q]$ et apparaît, par conséquent, comme un polynôme fini de z .

5.3 MODELISATION DE PROCESSUS SPATIO-TEMPORELS

5.3.1 Représentation générale de systèmes spatio-temporels

Lorsque que le système étudié se compose de N zones spatiales en interaction, la représentation avec relations temporelles diffère peu à première vue de celle d'un système temporel multivarie. En effet, il suffit de substituer un vecteur de zones à un de variables dans la formulation (5.28). On a alors une fonction de transfert du type [Bennett, 1979 chap.2]:

$$y_t = \sum_{N=1}^N \sigma_{1j}(z) x_t \quad \text{où } y_t = \{y_{1t}\}, \quad x_t = \{x_{1t}\} \quad (5.35)$$

$$\Sigma = \{\sigma_{1j}(z)\}$$

La matrice de transfert Σ devient carrée (mais pas nécessairement symétrique) pour représenter l'ensemble des interactions spatiales possibles. Selon Sheppard [1979b], la matrice Σ est une matrice d'interaction complète car elle tient compte de l'intensité des flux de chaque paire zonale, de la relation fonctionnelle $x - y$ et du temps de réaction. Cette matrice peut provenir d'un processus autoregressif du genre:

$$y_t = \Omega \cdot y_t + A \cdot x_t \quad \text{où } \Omega = \{\Omega_{1j}(z)\}, \quad A = \{A_{1j}(z)\} \quad (5.36)$$

La matrice A désigne les interactions directes découlant de la relation fonctionnelle $x - y$; ses éléments sont polynômes de l'opérateur de retard. Si Ω est diagonale ($\Omega_{1j}(z) = 0$; $i \neq j$), on observe une simple autocorrelation temporelle parmi la variable $\{y_{1t}\}$. Si Ω est non-diagonale, alors autocorrélations temporelles et spatiales se mêlent au sein de la variable endogène. Celle-ci dépend non seulement de ses valeurs passées mais également des valeurs enregistrées chez ses voisines. La matrice de transfert prend la forme suivante:

$$y_t = [I - \Omega]^{-1} A x_t \quad (5.37)$$

Les éléments de transfert sont, comme on sait, des divisions polynômiales de z . On remarque facilement l'isomorphie de la matrice de transfert avec le multiplicateur spatial [Sheppard, 1976 chap.6]. Le multiplicateur spatial «pur» somme les effets directs et indirects sans référence temporelle. Dans la section 6.3.5, on démontrera qu'un tel multiplicateur fait partie de la solution d'équilibre du modèle de Lowry. La matrice de transfert

reste du point de vue conceptuel semblable au multiplicateur spatial mais considère, en surcroît, un décalage des interactions de même qu'une persistance de l'influence. On peut s'en rendre compte en développant l'inversion matricielle sous la forme de série géométrique:

$$[I - \underline{\Omega}]^{-1} = \sum_{r=0}^{\infty} \underline{\Omega}^r \quad \text{où } \underline{\Omega}^r \equiv \underline{\Omega} \text{ multipliée } r \text{ fois par elle-même}$$

Si les polynômes $\Omega_{ij}(z)$ sont tels que le processus dans (5.36) est stationnaire, alors la série est convergente [Murata, 1977]. Le modèle (5.37) usant de la fonction d'impulsion-réponse correspondante serait une somme infinie de matrices comme celles-ci:

$$y_t = \left[V_0 + V_1 z + V_2 z^2 + V_3 z^3 + \dots \right] x_t \quad (5.38)$$

Le gain du système évalué dans des conditions d'état soutenu de la part de x_t équivaut à:

$$G = \sum_{r=0}^{\infty} V_r \equiv [I - \underline{\Omega}]^{-1} \underline{A} \quad (5.39)$$

où $\underline{\Omega} = (\underline{\Omega}_{ij})$; $\underline{\Omega}_{ij} = \sum_{k=0}^p \omega_{ijk}^k$ et $\underline{A} = (\underline{A}_{ij})$; $\underline{A}_{ij} = \sum_{k=0}^p \alpha_{ijk}^k$

Les matrices soulignées correspondent aux matrices homonymes dont les entrées polynômiales ont été démunies de leurs termes z . L'opérateur n'est donc plus nécessaire si on constate des conditions d'état soutenu ($x_{1t} = x_1$; $\forall t$). Dans de telles conditions, le vecteur d'équilibre y est obtenu en multipliant la matrice de gain G par le vecteur constant x ; le gain du système spatio-temporel en équilibre devient alors équivalent à un multiplicateur spatial (sans référence temporelle).

5.3.2 Spécification de la structure spatiale

La représentation de système spatio-temporelle qui a été exposée précédemment reste en pratique inutilisable. En fait, il n'y a aucune procédure d'estimation qui puisse justifier une configuration spatiale des interactions puisqu'il y a N^2 flux à estimer d'après $2N$ observations de variables localisées [Sheppard, 1979b]. Dans les modèles du temps, les structures de délai sont usuellement présumées constantes à travers le temps. Dans les modèles spatiaux, la structure spatiale ne peut être jugée constante à travers l'espace bien que relativement stable dans le temps. La structure spatiale informe sur les relations de voisinage qu'entre-

tiennent les zones entre elles. Il ne s'agit pas de mesurer l'intensité des échanges inter-zonaux. Néanmoins, on sait pertinemment que les interactions sont véhiculées via une structure spatiale et que celle-ci est déterminante dans l'intensité des relations [Griffith, 1978].

L'idée de représenter une structure spatiale par une matrice de pondération revient à Cliff et Ord [1970] qui en firent usage pour tester l'autocorrélation spatiale d'une série. Cette matrice de structure exprime l'intensité du voisinage de chaque paire zonale. La spécification du type de voisinage et la façon de mesurer incombe à l'analyste car il n'existe aucune procédure d'estimation possible étant donné le nombre d'éléments à estimer en rapport avec les observations disponibles. Griffith [1978] identifia quelques unes des structures les plus courantes dans les systèmes socio-économiques. Une structure «hiérarchique» définit les liens de domination fonctionnelle ou politique entre les zones. La matrice W ($=\{w_{ij}\}$ $N \cdot N$) est nécessairement asymétrique puisque les unités spatiales de niveau supérieure ne dépendent pas de celles de niveau inférieure. Les éléments matriciels sont spécifiés de la façon suivante:

$$w_{ij} \begin{cases} = 1 & \text{si zone } i \text{ est subordonnée à la zone } j \text{ dans la hiérarchie} \\ = 0 & \text{sinon} \end{cases}$$

Lorsque les relations spatiales sont canalisées à travers un réseau, on choisira une structure de «connectivité». On obtient alors une matrice dichotomique et symétrique du type:

$$w_{ij} \begin{cases} = 1 & \text{si zones } i \text{ et } j \text{ reliées directement par un lien du réseau} \\ = 0 & \text{sinon} \end{cases}$$

La structure spatiale la plus répandue dans les modèles linéaires spatiaux reste sans aucun la structure de «contiguïté». Il s'agit plus ou moins d'une transposition du décalage temporel dans la dimension spatiale. Un bon nombre de processus purement spatiaux se transigent d'une unité géographique aux unités adjacentes notamment dans les phénomènes de diffusion [par exemple, Hägerstrand, 1968; Phipps, 1987]. La matrice W se décrit:

$$w_{ij} \begin{cases} = 1 & \text{si zone } i \text{ est adjacente à zone } j \\ = 0 & \end{cases}$$

Cliff et Ord [1973] privilégièrent une version de la matrice de contiguïté qui tenait compte de la longueur de frontière commune; cette pondération vise à décrire davantage l'intensité des contacts zonaux dans un système géographique irrégulier. Ils donnerent également un sens élargi à cette matrice en lui permettant de contenir toute mesure susceptible de décrire des relations géographiques. En outre, W peut être définie par une fonction de transport:

$$w_{ij} = f(c_{ij}) \quad \text{ou} \quad f(c_{ij}) \equiv \exp(-\beta c_{ij}) \quad \text{ou} \quad f(c_{ij}) \equiv k \cdot (c_{ij})^{-\alpha}$$

Il y a certes des avantages à recourir aux structures dichotomiques de connectivité ou de contiguïté. D'une part, l'usage d'ordres supérieurs n'est pas interdit. D'autre part, les propriétés d'une matrice binaire (avec diagonale nulle) sont bien connues [Tinkler, 1974]. En effet, si W décrit les contacts de premier ordre, son carré décrit le nombre de contacts de deuxième ordre tandis que son cube, les contacts de troisième degré. Notons que la matrice W élevée à la puissance zéro équivaut à la matrice identité I , définissant ainsi les "contacts internes". Un fait tout de même l'analogie avec l'opérateur de retard z dont la puissance correspond au nombre de décalage; du fait, une telle matrice de structure spatiale est souvent considérée comme un opérateur d'«écarts spatiaux» (en anglais "spatial lag"). Le net inconvénient de cette version bi-dimensionnelle est la redondance des liens des niveaux supérieurs. En effet, chaque entrée d'une matrice d'ordre r énumère tous les "chemins" possibles sollicitant r contacts entre deux lieux. Ces "chemins" se révèlent, du point de vue conceptuel, très importants car chacun d'eux est un itinéraire emprunté par les interactions indirectes [Sheppard, 1979b]. Ce mécanisme rétroactif est celui utilisé par un multiplicateur spatial.

Tinkler [1974] nota que les matrices binaires élevées à une puissance r avait une magnitude λ fois plus grande que celle élevée à $r-1$ ou λ est la valeur propre principale de la matrice de départ. Pour compenser cette magnitude croissante, on devrait multiplier chacune des matrices d'ordre r par un scalaire p^r où $p \leq 1/\lambda$. Ce procédé équivaut à rendre stochastique la matrice W de telle sorte que la somme:

$$I + pW + p^2W^2 + p^3W^3 + \dots \equiv [I - pW]^{-1} \quad (5.40)$$

est mathématiquement finie. Cette série géométrique est analogue à celle dérivée par un multiplicateur spatial ou à la transformée de z dans un processus unilatéral du temps (voir 5.8a).

Il y a plusieurs méthodes qui sont proposées dans le choix d'une structure spatiale appropriée. Dans une application univariée où des autocorrélations spatiales sont recherchées, on est tenté d'opter pour la structure qui maximisera la corrélation de la série par elle-même. Plusieurs structures alternatives peuvent être testées par l'«indice de Moran» [Moran, 1950] généralisé par Cliff et Ord [1970]. L'indice se présente comme le coefficient de corrélation mais intègre la matrice spatiale:

$$I_w = \frac{N \cdot \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{W \cdot \sum_i (y_i - \bar{y})^2} \quad (5.41)$$

$$\text{où } \bar{y} = \sum_i y_i / N \quad W = \sum_i \sum_j w_{ij}$$

L'usage de l'indice commande la détermination a priori de la matrice de structure. Martin et Oeppen [1975] proposeront une version bivariable de l'indice qui met en relief plutôt la corrélation croisée entre deux séries spatiales:

$$I_w = \frac{\sum_i \sum_j w_{ij} (y_i - \bar{y})(x_j - \bar{x})}{W \cdot [\text{Var}(y_i) \cdot \text{Var}(x_i)]^{1/2}} \quad (5.42)$$

$$\text{où } \bar{x} = \sum_i x_i / N \quad W = \sum_i \sum_j w_{ij}$$

D'autres méthodes apparaissent plus exploratoires car elles visent à détecter une structure rationnelle à partir des données. Lorsque l'arrangement des zones semble très régulier (cellules carrées ou hexagonales), il est facile d'identifier des écarts spatiaux (contiguïté) de degrés différents: il s'agit ni plus ni moins de conférer à l'espace le concept de décalage. Par analogie, on calcule un «corrélogramme spatial» [voir Cliff et al., 1975 chap.8] qui est en fait une fonction d'autocorrélation des écarts. Par exemple:

$$\rho_y(r) = \frac{\sum_i y_i y_i(r) - N \bar{y}^2}{[\text{Var}(y_i) \cdot \text{Var}\{y_i(r)\}]^{1/2}} \quad (5.43)$$

où $y_i(r)$: moyenne de y_j des zones j situées à r écarts de la zone i

Ici, on distingue la variance de la série $\{y_i\}$ à celle de la série $\{y_i(r)\}$ obtenue avec la moyenne du voisinage de degré r car la deuxième est nécessairement inférieure ou égale à la première. L'examen de cette fonction

d'autocorrelation spatiale permet de deceler l'association statistique entre les valeurs et sa variation dans un système géographique. Le défaut principal de la méthode reste indubitablement l'identification des paires de zones distantes de r écarts; l'opération assez évidente dans un système cellulaires devient complexe dans un système zonal irrégulier (mais davantage réaliste). On peut par contre utiliser les matrices de contiguïté élevées aux puissances correspondant à r et retenir pour chaque paire, le plus court chemin.

Pour identifier la structure spatiale sous-jacente à un processus d'interaction, l'approche préconisée par Haggett [1972; Cliff, Haggett et al., 1975 chap.5-7] recourt à la comparaison des séries temporelles de régions différentes en se basant sur la similitude des séries. Les fonctions d'autocorrelation temporelle peuvent être comparées pour chaque paire. Toutefois, dans des processus stationnaires, la fonction ne dépend que du décalage et, de ce fait, ne permet pas de dévoiler le déphasage de deux chroniques. L'analyse spectrale produit un spectre de fréquence des cycles à partir de la fonction d'autocorrelation temporelle. Sa version bivariable, l'analyse co-spectrale peut ainsi être employée pour calculer la «cohérence» cyclique et le «déphasage» des fluctuations entre deux séries [Granger, 1969]. Dans un système multi-régional, ces mesures permettent de repérer les ensembles régionaux ayant des affinités et le réseau de transmission des fluctuations (économiques surtout) [voir, par exemple, Marchand, 1981a; 1981b]. Ces méthodes de comparaison des séries peuvent alors servir à formuler des structures spatiales hypothétiques. Néanmoins, elles ne conviennent pas à tous les contextes car elles exigent des bases de données temporelles excessivement désagrégées (mois, trimestres) et longues (au moins 20 ans).

5.3.3 Formulation de modèles purement spatiaux

Les modèles purement spatiaux sont des modèles ne faisant intervenir que des interactions spatiales et des relations fonctionnelles. N'ayant aucune référence temporelle, ils misent sur la simultanéité des relations. L'un des premiers modèles linéaires spatiaux fut celui de Whittle [1954] qui représentait un processus d'autocorrelation multilatéral dans un système régulier de quadrats. Chaque cellule ayant habituellement quatre

voisins immédiats (à l'exception des zones en marge du système), la stationnarité du processus est assurée que si chaque valeur dépend d'un quart tout au plus de chacune des valeurs voisines. En fait, ce modèle apparaît comme une transposition spatiale d'un processus temporel markovien de premier ordre. La différence essentielle réside dans l'établissement de moyennes de voisinage du fait que chaque observation côtoie jusqu'à quatre autres observations.

La généralisation apportée par Cliff et Ord [1970; 1973; 1980] du modèle autorégressif multilatéral de Whittle reste aujourd'hui la formulation conventionnelle du modèle «autorégressif spatial» (appelé sous le sigle SAR). Le modèle représente un processus linéaire s'accordant avec toute structure spatiale adéquatement spécifiée par une matrice W . La diagonale de celle-ci est conventionnellement fixée à zéro pour s'assurer que les effets d'autocorrélation proviennent des zones voisines. La forme fonctionnelle du modèle SAR s'inspire du modèle de régression. En effet:

$$Y_i = \rho_0 + \rho_1 \sum_j W_{ij} Y_j + \varepsilon_i \quad (S.44a)$$

ρ_0 est une constante. ρ_1 est un paramètre autorégressif et ε_i , un terme d'erreur. Pour donner une interprétation logique à ρ_1 , une pondération telle que $\sum_j W_{ij} = 1$ ou $\sum_i W_{ij} = 1$ confine le paramètre à l'intervalle unitaire $[-1, 1]$ [Ord, 1975]. Cette pondération permet aussi de standardiser le poids du voisinage de sorte que, pour une structure de contiguïté, le nombre de zones adjacentes devient sans pertinence. Par ailleurs, rien n'empêche que certaines sommes de rangée ou de colonne soient inférieures à un. La standardisation de W en rangée diffère toutefois de celle en colonne; chez l'une, l'apport total du voisinage est considéré constant tandis que, chez l'autre, il est variable mais chaque zone conditionne les autres avec une même intensité. Cette propriété qu'on attribue à la matrice structurelle s'avère en fait quasi-essentielle. En notation matricielle, le modèle (S.44a) devient:

$$Y = \rho W Y + \varepsilon \quad (S.44b)$$

La constante disparaîtra si la pondération de W concerne les colonnes. On sait pertinemment que l'équation peut être reformulée autrement, soit:

$$Y = [I - \rho W]^{-1} \varepsilon \quad (S.45)$$

$$\text{où } [I - \rho W]^{-1} \equiv \sum_{r=0}^{\infty} \rho^r W^r = I + \rho W + \rho^2 W^2 + \dots$$

Si W est standardisée à l'unité, $|\rho| \leq 1$ et l'inversion de matrice donne une matrice finie. On remarquera que le terme d'erreur est convolué à travers un multiplicateur spatial de telle sorte que le vecteur y est affecté à prime abord par son propre terme d'erreur (via I), ensuite par ceux de ses voisins immédiats (via ρW) et enfin par ceux des voisins d'ordres plus élevés mais à chaque fois dans une moindre mesure. On comprend alors mieux le rôle du paramètre autoregressif. Dans la formulation initiale (5.44), il semble simplement mesurer l'impact des valeurs voisines. Dans la forme usant du multiplicateur spatial, il contrôle l'atténuation des interactions indirectes.

Une formulation préconisant les moyennes mobiles spatiales est possible [Cliff, Ord, 1980; Haining, 1981]. Cependant, elle ne sera point discutée ici car elle implique la modélisation d'une véritable composante aléatoire qui, en réalité, est inobservable. On se concentrera plutôt sur les modèles de régression spatiale, version bivariable du modèle cité précédemment.

La variable endogène (ou dépendante) est maintenant conditionnée par les valeurs d'une variable endogène prises chez les localisations voisines. Les éléments diagonaux de la matrice de structure spatiale peuvent être positifs, indiquant ainsi que de simples relations fonctionnelles aspatiales peuvent s'exercer au sein du système. La standardisation de la matrice n'apparaît plus obligatoire du fait que deux variables distinctes y sont confrontées; néanmoins, la standardisation reste désirable par souci de compatibilité. Voici un modèle de régression spatiale:

$$y_i = \delta_0 + \delta_1 \cdot \sum_j q_{ij} x_j + \varepsilon_i \quad (5.46)$$

La formulation additive et linéaire ne présente aucune difficulté en fait d'estimation. Le paramètre δ_1 y est vu comme un coefficient de régression rattaché à un régresseur composé d'une moyenne spatiale de valeur x_j . Usuellement, comme c'est aussi le cas dans l'autorégression, le coefficient reste faible car il y a peu de variance chez une variable calculée comme une moyenne spatiale. Cette carence de variabilité est particulièrement prononcée lorsque Q est standardisée par rangée plutôt que par colonne. Dans le premier cas, le poids cumulé par zone "réceptrice" de la variable

exogène y demeure constant tandis que, dans le deuxième cas, la constance s'applique sur l'effet total de la zone "émettrice", permettant ainsi de différencier les effets reçus.

La formulation la plus intéressante parmi les modèles purement spatiaux reste sans aucun doute celle combinant autoregression et régression. La variable endogène y est affectée par les valeurs prises chez ses voisins et par une autre variable calculée tant chez elle que chez les zones qui la côtoient. Représentons un système bivarie dépourvu d'une composante aléatoire (composante facultative dans une représentation théorique non soumise aux problèmes d'estimation):

$$y = \rho Wy + \delta Qx \quad \text{où } W = \{w_{ij}\} N \cdot N; \quad w_{ii} = 0 \quad \forall i; \quad \sum_i w_{ij} = 1 \\ Q = \{q_{ij}\} N \cdot N; \quad \sum_i q_{ij} = 1 \quad (5.47)$$

Par ailleurs, cette équation non-linéaire en terme du vecteur y devient:

$$y = [I - \rho W]^{-1} Qx \quad (5.48)$$

Par définition, W est stochastique et, par conséquent, $|\rho| < 1$. De ce fait, l'inversion correspond à une somme géométrique infinie mais convergente composée de matrices élevées à des puissances successives (voir S.40). La matrice $[I - \rho W]^{-1} Q$ apparaît comme un multiplicateur spatial tel que décrit par Sheppard [1976, 1979b] convertissant la série géographique x en une série y . Les véritables impulsions du système sont contenues dans x mais les effets d'autocorrelation chez y amplifie considérablement (par rétroaction) l'effet total reçu.

Curry [1972] considéra que la «fonction de transfert spatial» est une approche alternative à celle dite «gravitaire». La fonction de transfert qu'on assimile au multiplicateur spatial reste un «opérateur» convoluant un «patron d'origines» en un «patron de destinations». Curry établit qu'une «régionalisation en cascade» suivie d'une différentiation du premier ordre était équivalente à une fonction d'impulsion-réponse exponentielle-négative. La régionalisation en cascade est un procédé hiérarchique de moyenne spatiale; à chaque niveau régional, une moyenne est calculée et est considérée comme un décalage spatial. Une structure spatiale de contiguïté ressemble à cette régionalisation; le premier degré correspond à la couronne immédiate de la zone, le second, à la couronne suivante et ainsi de

suite. Sheppard [1976 chap.6] rappelle que la fonction d'impulsion-réponse décrite pour les processus temporels par Jenkins et Watts [1968] s'accorde avec tout processus spatial doté d'un opérateur établissant la connectivité ou la contiguïté des zones. La fonction IR spatiale équivaut alors aux éléments des matrices sommées dans (5.40) correspondant à chaque niveau d'écart spatiaux. Dans un processus de premier ordre comme (5.44) ou (5.47), il y a N^2 fonctions IR (N intra-zonales et N^2-N inter-zonales) qui arborent une forme exponentielle-négative.

5.3.4 Conciliation des séries temporelles et des modèles spatiaux

Selon Bennett [1979, chap.7], les modèles purement spatiaux semblent essentiellement théoriques car les interactions spatiales de toute sorte sont normalement transmises en requérant un certain temps. De plus, la «réponse» ne peut se manifester en simultanéité avec l'«impulsion». Chez les modèles spatiaux, l'inertie d'une structure, la mémorisation d'un terme, la persistance d'un effet et le décalage des réactions paraissent des concepts complètement impertinents. Bennett et Chorley [1978] identifient trois types d'application où des processus strictement spatiaux sont concevables:

1- Ils existent sous certaines conditions, notamment à l'atteinte d'un état d'équilibre statique (impulsions tenues constantes):

2- Ils semblent exister lorsque les données espace-temps sont très écartées temporellement de sorte que la dynamique de décalage devient imperceptible. Dans d'autres occasions, les réactions s'exercent dans un temps si court que l'on leur confère la simultanéité.

3- Il existe des problèmes purement spatiaux. Les problèmes d'extrapolation spatiale (données pour des zones hors-région), l'interpolation spatiale (données manquantes pour zones de la région) [voir Tobler, 1973] et l'analyse des résidus de régression [Cliff, Ord, 1973] ne nécessitent aucune référence du temps.

Dans la section 5.3.1, des représentations générales de systèmes spatio-temporels ont été exposées. Elles consistaient en des matrices spatiales contenant des termes polynômiaux de retard. Les procédés d'identification temporelle servent à simplifier la structure de décalage. Ainsi, du point de vue paramétrique, la spécification des relations tempo-

nelles se révèle moins gourmande. Analogiquement, nous montrions dans la section précédente, que des structures spatiales peuvent être identifiées au moyen de mesures d'autocorrelation et de corrélation spatiales. Si ces structures sont spécifiées par une matrice de pondération, un seul paramètre inconnu devient nécessaire pour un processus d'interaction de N-N zones. La simplification apparaît davantage substantielle.

L'inconvénient principal du modèle théorique spatio-temporel résidait dans le trop grand nombre de paramètres dont chacun exprime une relation de décalage couplée avec une interaction spatiale. A l'égard de l'estimation, le problème apparaissait insurmontable [Hordijk, Nijkamp, 1977]. Il y a une approche conciliant les méthodes d'identifications spatiale et temporelle; Bennett [1981] la nomme «approche par matrice de pondération» car elle emploie la matrice de structure spatiale comme opérateur liant les observations géographiques. Les modèles de prévision suggérés par Cliff et Ord [1975a; 1975b; 1980; Cliff et al., 1975] et de diffusion de Hordijk et Nijkamp [1977; 1978] appartiennent à cette catégorie. Dans ce type de modélisation, les processus autoregressif et regressif spatiaux (comme S.47) sont dotés de fonctions de délai. Usuellement, les structures de décalage sont choisies de manière à maintenir une parcimonie paramétrique.

La conciliation des dimensions spatiales et temporelles amènent un nouveau problème de spécification. En effet, pour une structure spatiale préalablement sélectionnée, il y a 3 différents niveaux d'agrégation de la structure temporelle:

- 1- La structure de décalage est propre à chaque paire de zones (y compris la zone envers elle-même). Il y a alors N^2 structures.
- 2- Il n'y a que deux structures, l'une concernant les relations intra-zonales et l'autre, les relations inter-zonales.
- 3- La structure de décalage est unique et se rapporte à toutes les interactions.

Le niveau le plus désagrégé implique que la dynamique diffère selon le couple de lieux. Du point de vue estimatif, cette approche reste inacceptable pas plus que du point de vue conceptuel où il est difficile d'admettre que les processus de transmission font la distinction entre les paires de lieux. Le niveau le plus agrégé simplifie considérablement le

probleme d'estimation. Toutefois, il semble important de distinguer les relations intra-zonales des relations inter-zonales comme on le fait au second niveau. Les relations internes étant évidemment aspatiales, elles ne solliciteront pas les mêmes délais que des interactions. De plus, dans un processus autoregressif, les relations internes sont necessairement des memorisations des valeurs passées alors que les relations externes désignent les effets d'autocorrelation transmis a retardement. Dans un processus de nature regressive, les relations intra-zonales relèvent a la fois de liaisons purement fonctionnelles et d'interactions spatiales (a courte distance). Les relations inter-zonales restent strictement des interactions spatiales bivariées.

Un modele spatio-temporel specifie par des matrices de ponderation et par deux types de polynôme de retard prend l'allure suivante:

$$y_{1t} = \Omega_I(z)y_{1t} + \Omega_W(z) \cdot \sum_j w_{1j} y_{jt} + A_I(z)x_{1t} + A_O(z) \cdot \sum_j q_{1j} x_{jt} \quad (5.49a)$$

ou en notation matricielle:

$$y_t = \Omega_I(z)y_t + \Omega_W(z)W y_t + A_I(z)x_t + A_O(z)Q x_t \quad (5.49b)$$

$$y_t = \left[\Omega_I(z)I + \Omega_W(z)W \right] y_t + \left[A_I(z)I + A_O(z)Q \right] x_t \quad (5.49c)$$

La formulation ci-dessus (5.49) possède une forme identique à la representation généralisée (5.36). Les matrices entre crochet equivalent respectivement aux matrices Ω et A composees d'entrees polynômiales de z . La correspondance se fait comme suit:

$$\Omega_{1j}(z) \begin{cases} = \Omega_I(z) & i=j \\ = \Omega_W(z)w_{1j} & i \neq j \end{cases} \quad A_{1j}(z) \begin{cases} = A_I(z) + A_O(z)q_{1j} & i=j \\ = A_O(z)q_{1j} & i \neq j \end{cases}$$

Désormais, il n'y a que quatre polynômes de decalage differents à estimer. Martin et Geppen [1975] ainsi que Hordjik et Nijkamp [1977] suggererent des variantes temporelles à l'indice de Moran. La mesure d'autocorrelation spatiale peut être convertie en une fonction du délai K :

$$\rho_{y_{1t}}(K) = \frac{\text{Cov}(y_{1t}, \bar{y}_{1t-K})}{[\text{Var}(y_{1t}) \cdot \text{Var}(\bar{y}_{1t})]^{1/2}} \quad (5.50)$$

$$\text{où } \text{Cov}(y_{1t}, \bar{y}_{1t-K}) = \frac{1}{NT} \sum_{t=1}^N \sum_{\tau=K+1}^T y_{1t} \bar{y}_{1\tau-K} - \bar{y}^2$$

$$\bar{y}_{1t} = \sum_j w_{1j} y_{jt} \quad \bar{y} = \sum_i \sum_t y_{it} / NT$$

Cette fonction reste isomorphe à la fonction d'autocorrelation temporelle (5.2), exception faite de l'«opérateur spatial» qui fait contronter les observations voisines à l'instar des décalages. Cette fonction peut servir à l'identification du polynôme $\Omega_w(z)$ concerne par les interactions auto-regressives au sein de la série $\{y_{1t}\}$. Quant au polynôme $\Omega_x(z)$ de relations strictement intra-zonales, la fonction aspatiales (5.2) suffira à l'identification. Parallèlement, la fonction de corrélation croisée temporelle (5.19) aidera à la spécification de $A_x(z)$ traitant des relations bivariées $\{x_{1t}\}$ versus $\{y_{1t}\}$ mais intra-zonales. La forme la plus générale de fonctions spatio-temporelles reste celle où l'on compare les deux séries à travers le temps et une structure spatiale nommée Q :

$$\rho_{yx}(K) = \frac{\text{Cov}(y_{1t}, x_{1t-K})}{[\text{Var}(y_{1t}) \cdot \text{Var}(x_{1t})]^{1/2}} \quad (5.51)$$

$$\text{où } \text{Cov}(y_{1t}, x_{1t-K}) = \frac{1}{NT} \sum_{t=1}^N \sum_{\tau=t-K+1}^T y_{1t} x_{1\tau-K} - \bar{y}\bar{x}$$

$$\bar{x}_{1t} = \sum_j q_{1j} x_{jt} \quad \bar{x} = \sum_{t=1}^T \bar{x}_{1t} / NT$$

La matrice Q n'a pas de restriction de nullité de sa diagonale; elle comprend donc des "interactions intra-zonales". Le lecteur remarquera que dans le système (5.49), le polynôme $A_0(z)$ concerne aussi les relations temporelles s'exerçant à l'intérieur du même zone. On notera également que les dénominateurs des fonctions (5.50) et (5.51) distinguent la variance de la série originale ($\{y_{1t}\}$) de celle de la série transformée ($\{\bar{y}_{1t}\}$ ou $\{\bar{x}_{1t}\}$) contrairement à l'indice de Moran (5.41) et aux fonctions temporelles (5.2, 5.19). Quand il s'agit strictement de décalage, le couplage se fait un à un, ce qui fait que les variances des séries initiale et décalée sont asymptotiquement identiques [Jenkins, Watts, 1968 chap.5]. Dans l'espace, les matrices standardisées W et Q produisent des séries à partir de moyennes du voisinage de i , séries dont les variances sont considérablement atténuées. L'usage des séries originales dans les dénominateurs réduit donc la taille des estimées [Cliff et al., 1975 chap.8].

5.4 ESTIMATION D'UN SYSTEME SPATIO-TEMPOREL

Ce sous-chapitre traitera des problèmes d'estimation qui sont rencontrés dans l'évaluation des paramètres qui composent une équation de système. De par la linéarité de ces équations, on privilégiera une méthodologie conventionnelle basée sur les moindres carrés. Ceux-ci conviennent aux applications temporelles et sont équivalents à une estimation de vraisemblance maximale si les erreurs sont présumées normales. La considération d'un système spatial ne nuit pas à la pertinence de la méthode et peut même être favorable car elle fournit un plus grand nombre d'observations. Les phénomènes spatio-temporels décrits par plus d'une variable y trouveront leur compte car la méthode s'accommode d'un cadre multivarié. La combinaison de l'espace avec le temps occasionne certes des problèmes d'estimation, d'une part, en y confondant une structure d'autocorrelation spatiale des erreurs à celle du temps et, d'autre part, en y ajoutant un problème de non-stationnarité spatiale des variables sujetes à un processus d'autorégression dans l'espace.

5.4.1 Paramètres de systèmes temporels

Cette section abordera l'estimation des paramètres de modèles purement temporels (univariés et bivariés) dont la ou les structures de décalage ont été préalablement identifiées (voir ss-chap. 5.2). La méthode d'estimation appropriée est celle qui s'accorde avec les présomptions distributionnelles de la série endogène. Pour éviter de souscrire à un critère estimatif choisi arbitrairement, il convient de référer avant tout à la méthodologie de la «vraisemblance». Celle-ci paraît particulièrement apte à des situations où un petit ensemble de paramètres doit être estimé à partir d'observations [Jenkins, Watts, 1968, chap.5]. Le principe de cette méthode se décrit comme suit. Lorsqu'une série est réalisée (et observée), cette série est connue mais les paramètres décrivant la distribution restent inconnus. Une fonction des paramètres est obtenue en substituant ceux-ci par les valeurs observées dans la «fonction de densité de probabilité» de la distribution (souvent référée par le sigle PDF de l'anglais "probability density function"). Cette fonction devient alors une «fonction de vraisemblance» des paramètres. Elle décrit la vraisemblance d'apparition parmi les valeurs paramétriques possibles.

Considérons une dépendance linéaire de la série $\{y_t\}$ à l'égard d'une série exogène $\{x_t\}$ où $t=1,2,\dots,T$. Ces séries ont des moyennes égales à zéro. Par conséquent, $\{y_t\}$ peut être décrite par l'équation suivante:

$$y_t = \beta x_t + \varepsilon_t \quad (5.52)$$

La série résiduelle $\{\varepsilon_t\}$ ou $\{y_t - \beta x_t\}$ possède par présomption un comportement normal $N(0, \sigma^2)$ où σ^2 est la variance résiduelle qui est inconnue. La fonction de vraisemblance des paramètres β et σ^2 est donnée par la PDF normale à T dimensions de la série $\{\varepsilon_t\}$:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_t (y_t - \beta x_t)^2\right] \quad (5.53)$$

Les valeurs de β et de σ^2 les plus vraisemblables sont celles qui maximisent $L(\beta, \sigma^2)$ ou, de façon équivalente, son logarithme:

$$\max_{\{\beta, \sigma^2\}} \ln L = (-T/2) \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_t (y_t - \beta x_t)^2 \quad (5.53)$$

Soit l'annulation des dérivées partielles en respect de σ^2 et de β :

$$\frac{\partial(\ln L)}{\partial \sigma^2} = \frac{-T}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_t (y_t - \beta x_t)^2 = 0 \quad (5.54)$$

L'estimateur ML de la variance devient donc:

$$\hat{\sigma}^2 = \frac{\sum_t (y_t - \beta x_t)^2}{T} \quad (5.55)$$

D'autre part:

$$\frac{\partial(\ln L)}{\partial \beta} = \frac{-2 \sum_t y_t x_t}{\sigma^2} + \frac{2 \sum_t (x_t)^2}{\sigma^2} = 0 \quad (5.56)$$

$$\hat{\beta} = \frac{\sum_t y_t x_t}{\sum_t (x_t)^2} \quad (5.57)$$

Le lecteur averti notera que l'estimateur ML du paramètre β est identique à celui qui provient du principe des moindres carrés ordinaires (référé conventionnellement au sigle OLS). En effet, la fonction maximisée $\ln L(\beta, \sigma^2)$ contient explicitement un terme s'identifiant à la somme des erreurs au carré (avec un signe négatif) qui reste le seul terme variable en respect de β . En général, les estimateurs ML liés à des modèles linéaires dotés d'une composante d'erreur distribuée normalement sont équivalents aux estimateurs OLS [Jenkins, Watts, 1968]. Cela vaut pour les processus de décalage distribué et les processus autorégressifs. Néanmoins, il y a certaines mises en garde spécifiques aux modèles temporels.

Le schème de régression appliqué aux séries chronologiques exige la stationnarité des erreurs. Celles-ci ont une moyenne nulle et constante à travers la période investiguée. Leur variance satisfait la condition d'homoscédasticité. La non-autocorrélation des erreurs reste sans doute l'une des conditions les plus difficiles à respecter. Le caractère ordonné d'une chronique et l'inertie des variables nuisent certe à l'indépendance observations résiduelles adjacentes. On connaît bien les méfaits de l'autocorrélation des erreurs sur les estimateurs OLS: ceux-ci demeurent sans biais mais ils ne sont plus de variance minimale [Philips, Blomme, 1981]. Des tests appropriés sont connus pour reconnaître la dépendance autoregressive des erreurs [Durbin, Watson, 1950]. Advenant le cas, on emploiera en remplacement de la procédure OLS, une procédure d'estimation en deux étapes (mieux connus sous le sigle 2SLS "Two-Stage Least Squares"), l'une pour l'autocorrélation, l'autre pour le modèle dont on a soustrait l'effet nocif [voir Cochrane, Orcutt, 1949]. On pourrait recourir aussi à la méthode d'Aitken [1934] dite «des moindres carrés généralisés» (GLS) dans lequel on spécifie la structure de covariance du terme d'erreur (dans la section 5.4.3, il sera question de l'estimation GLS dans l'éventualité d'autocorrélation spatiale).

Dans le type d'application qui concerne cette thèse, l'autocorrélation temporelle n'apparaît pourtant pas comme le problème le plus épineux. En effet, il y a peu de chance de déceler une telle dépendance au sein des séries temporelles disponibles dans le domaine du développement urbain. Le peu d'observations à travers le temps donnera des séries très courtes. Les risques se situent plutôt au niveau de la brièveté de la période d'observation. Il y a différents problèmes qui s'y rattachent. Les questions relevant de la taille de l'échantillon sont évitées grâce à l'utilisation d'un système spatial lequel multiplie les observations à chaque unité de temps. Par ailleurs, la reconnaissance des décalages devient le problème le plus évident et ce, pour deux raisons. D'une part, pour chaque série, il y a un nombre restreint de croisements décalés; l'estimation à zones multiples permettra, par contre, de s'appuyer sur un plus grand nombre de croisements. D'autre part, on n'est plus autorisé à des ordres élevés de décalage car, à chaque fois, une observation temporelle est perdue. Il n'est d'ailleurs guère recommandé de poursuivre au-delà du tiers de longueur de la série.

A propos des processus autoregressifs, il importe que la série observée soit stationnaire. La nullité de la moyenne n'est pas exigée car elle est compensée par l'introduction d'une ordonnée à l'origine. La contravention à la stationnarité la plus fréquente semble la tendance dans le temps: les valeurs moyennes de la série tendent à augmenter ou décroître avec le temps. L'existence d'une tendance ne nuit pas à la fiabilité des paramètres autoregressifs mais confond plutôt la mémorisation avec l'évolution de la magnitude. La tendance peut être corrigée par un procédé de différentiation temporelle [voir Box, Jenkins, 1970, chap.4]. Une série différenciée au premier degré est constituée des increments (et decrements) entre chaque observations successives:

$$\nabla y_t = y_t - y_{t-1} \equiv (1-z)y_t \quad (5.58)$$

L'«opérateur de différence» ∇ possède des propriétés similaires à celui de retard z . Ordinairement une tendance linéaire est corrigée par la transformation ∇ tandis qu'une tendance quadratique se satisfait d'une différence du second degré $\nabla^2 = (1-\nabla)^2 = (1-2z+z^2)$ et ainsi de suite. Néanmoins, dans le cas présent, on évitera un tel procédé car il soustrait, à chaque degré supplémentaire, une observation à la série. Une technique bien connue consiste à évaluer la tendance à l'aide de moyennes mobiles ou d'un polynôme du temps et à la soustraire de la série originale. Les moyennes mobiles sont reconnues pour le lissage des séries [Spiegel, 1981]. On doit sa grande flexibilité à la possibilité d'adopter des degrés de lissage différents selon l'ordre de récursion. Par exemple, une moyenne mobile d'ordre 1:

$$\bar{y}_t = (y_{t-1} + y_t + y_{t+1})/3 \quad (5.59a)$$

produit une série filtrée $\{\bar{y}_t\}^{<1>}$ moins lisse qu'une série $\{\bar{y}_t\}^{<2>}$:

$$\bar{y}_t = (y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2})/5 \quad (5.59b)$$

Ce procédé non-paramétrique a l'inconvénient de nécessiter des ordres plus élevés pour décrire les tendances les plus simples. De plus, il faut établir une convention pour les observations en marge de la série sinon elles doivent être exclues. L'utilisation d'un polynôme du temps reste la technique la plus courante et n'implique pas l'amputation de la série. Le degré polynomial dépend avantageusement de la curvilinéarité de la tendance. Une équation additive décrit facilement une progression linéaire:

$$y_t = a + bt + y_t \quad t=1,2,\dots,T \quad (5.60)$$

où $\{y_t\}$ devient la série dénuée de tendance. Il s'agit d'une démarche consistante car les paramètres a et b peuvent être estimés par moindres carrés. L'estimation de cette pente peut s'effectuer en simultanéité avec le processus autoregressif, notamment en adoptant la formulation suivante:

$$y_t = a + bt + \omega y_{t-1} + \varepsilon_t \quad (5.61)$$

Ce modèle distingue une composante de mémorisation du 1^{er} degré d'une tendance linéaire.

Quant à l'éventuelle «hétéroscédasticité», elle s'explique souvent par une progression non-linéaire de la série; l'homoscédasticité sera rétablie par une transformation adéquate permettant de lineariser la série [Philips, Blomme, 1981 chap.2]. Ici, de telles corrections apparaissent incompatibles avec les systèmes spatio-temporels à moins qu'elles s'appliquent invariablement à toutes les séries zonales. En surcroît, les interactions spatiales sont généralement établies à partir d'une échelle numérique absolue sans laquelle toute comparaison inter-zonale est gâchée.

5.4.2 Paramètres de modèles spatio-temporels

Les méthodes d'estimation pour les modèles linéaires d'interaction spatiale (sans référence du temps) diffèrent selon la nature du processus étudié. Les processus spatiaux régressifs ne se distinguent pas du modèle de régression classique, l'effet de l'opérateur géographique pouvant se traduire par une nouvelle variable synthétique. Par exemple, le modèle:

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 \sum_j q_{ij} x_j + \varepsilon_i \quad (5.62)$$

est isomorphe à la formulation multivariée et aspatiale suivante:

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 z_i + \varepsilon_i \quad \text{où } z_i = \sum_j q_{ij} x_j \quad (5.63)$$

Par conséquent, la série $\{z_i\}$ constitue une variable indépendante établie d'après les valeurs de la variable $\{x_i\}$ au voisinage de la zone i . La procédure OLS fournit alors une estimation non-biaisée et de variance minimale. L'introduction d'une structure de décalage ne détruit aucunement les propriétés de l'estimation par moindres carrés.

Quant aux processus spatiaux autoregressifs, il est difficile d'envisager l'indépendance d'un régresseur engendré par une transformation spatiale de la variable régressée. Whittle [1954] et Ord [1975] montrèrent que l'estimateur OLS du paramètre autoregressif était «inconsistant» i.e. qu'il ne converge pas vers la distribution du vrai paramètre (il est asymptotiquement biaisé). Ord [1975] spécifia la fonction de vraisemblance du paramètre autoregressif ω_1 d'un modèle comme celui-ci :

$$y_1 = \omega_1 \sum_j w_{1j} y_j + \varepsilon_1 \quad \text{ou} \quad y = \omega_1 W y + \varepsilon \quad (5.64)$$

où W est une matrice de pondération de taille $N \times N$; elle est standardisée de telle sorte qu'une constante pour l'ordonnée à l'origine devient inutile. En supposant le terme d'erreur ε distribué selon une loi $N(0, \sigma^2 I)$ i.e. de loi normale avec une moyenne nulle et une variance constante et non-autocorrélée, la fonction de vraisemblance apparaît comme suit :

$$L(\omega_1, \sigma^2) = \frac{|A|}{(2\pi\sigma^2)^{N/2}} \exp \{-(y'A'Ay)/2\sigma^2\} \quad (5.65)$$

où $A = I - \omega_1 W$, $\varepsilon = Ay$ et $|A|$: déterminant de A

La fonction dérive directement de la PDF multivariée laquelle implique une matrice variance-covariance. Celle-ci est implicitement incorporée par la matrice $(A'A)^{-1}\sigma^2$ (voir Jenkins, Watts, 1968 chap.4). Le logarithme de $L(\omega_1, \sigma^2)$ s'écrit :

$$\ln L(\omega_1, \sigma^2) = -(N/2) \ln(2\pi\sigma^2) + \ln|A| - (y'A'Ay)/2\sigma^2 \quad (5.66)$$

L'estimateur ML de la variance reste identique à la convention :

$$\hat{\sigma}^2 = (y'A'Ay)/N \quad (5.67)$$

tandis que celui du paramètre est celui qui maximise :

$$-(N/2) \ln(\hat{\sigma}^2 |A|^{-2/N}) \quad (5.68)$$

La méthode de vraisemblance maximale ne génère pas dans ce cas-ci un estimateur identique à celui des moindres carrés. La maximisation de (5.68) sollicite certes une minimisation des erreurs au carrés mais aussi une augmentation du log du déterminant de la matrice A . D'après sa définition, celle-ci est une fonction du paramètre autoregressif. Dans une procédure numérique itérative, la recherche répétitive du déterminant est évitée en utilisant les N valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_N$ de W qui sont invariables

[Ord, 1975]:

$$|A| = |I - \omega_1 W| \equiv \prod_{i=1}^N (1 - \omega_1 \lambda_i) \quad \text{alors} \quad \ln|A| = \sum_i \ln(1 - \omega_1 \lambda_i) \quad (5.69)$$

Une des conditions essentielles est que A soit positive définie ou bien que W soit stochastique (sommes des rangées ou des colonnes égales ou inférieures à un).

Cette procédure ML demeure particulière au modèle d'autorégression spatiale sans décalage. L'introduction d'un délai rend la méthode des moindres carrés propre à l'estimation puisqu'on peut assimiler l'opération spatiale Wy_{t-k} à un régresseur z_{t-k} distinct (mais non véritablement indépendant) de la distribution de y_t .

Pour l'estimation d'un système spatio-temporel où les relations autorégressives ne sont pas simultanées, Bennett [1979, chap.4] reconnaît l'adéquation des «moindres carrés multivariés» conventionnellement notés par les lettres MOLS [voir Anderson, 1958]. Dans l'esprit de Bennett, les séries chronologiques décalées sont considérées comme des régresseurs de la série actuelle: dans un système non-spatial, l'estimation OLS multiple convient. Dans un système géographique soumis à des interactions, chacune des chroniques zonales est prise comme une variable. Un système multivarié d'équation en résulte si bien que la méthode MOLS y trouve sa justification. Les interactions entre régions sont distinctement paramétrisées au même titre que des relations fonctionnelles entre variables. Néanmoins, l'approche adoptée ici preconise l'emploi de matrices de structure spatiale lesquelles résument l'ensemble des relations spatiales dans une variables. La paramétrisation reste donc conventionnelle i.e. qu'un coefficient se rattache à chacune de ces variables synthétiques. Un schéma multivarié n'apparaît donc plus nécessaire si pas plus d'une variable endogène n'est impliquée.

Reprenons la formulation générale (5.49) qui comprend des relations temporelles et spatiales de nature régressives et autorégressives. Dans un contexte estimatif, prenons soin d'y ajouter une composante d'erreur:

$$y_t = \Omega_x(z)y_t + \Omega_w(z)Wy_t + A_x(z)x_t + A_\varepsilon(z)\varepsilon_t + \varepsilon_t \quad (5.70)$$

$$\text{ou } \Omega_1(z) = \omega_1^1 z + \omega_1^2 z^2 + \dots + \omega_1^p \cdot z^p$$

$$\Omega_w(z) = \omega_w^1 z + \omega_w^2 z^2 + \dots + \omega_w^p \cdot z^p$$

$$A_1(z) = \alpha_1^0 + \alpha_1^1 z + \alpha_1^2 z^2 + \dots + \alpha_1^q \cdot z^q$$

$$A_w(z) = \alpha_w^0 + \alpha_w^1 z + \alpha_w^2 z^2 + \dots + \alpha_w^q \cdot z^q$$

Le système nécessite l'estimation de H paramètres ou $H = p' + p'' + q' + q'' + 2$. On convient que les ordres de décalage restent en pratique peu élevés et qu'il se peut que certains ordres intermédiaires soient omis. De même, si des déphasages existent, les premiers termes du polynôme seront éclipsés. Dans la recherche des estimateurs par moindres carrés, il est préférable de reformuler (5.70) par une équation composée d'observations spatiales et temporelles. Ainsi:

$$\underline{Y}_{NT \times 1} = \underline{X}_{NT \times H} \underline{\theta}_{H \times 1} + \underline{\varepsilon}_{NT \times 1} \quad (5.71)$$

$$\text{où } \underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix}$$

$$\underline{X} = \begin{bmatrix} y_0 & \dots & y_{1-p'} & \omega y_0 & \dots & \omega y_{1-p''} & x_1 & \dots & y_{1-q'} & Qx_1 & \dots & Qx_{1-p''} \\ y_1 & \dots & y_{2-p'} & \omega y_1 & \dots & \omega y_{2-p''} & x_2 & \dots & y_{2-q'} & Qx_2 & \dots & Qx_{2-p''} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ y_{T-1} & \dots & y_{T-p'} & \omega y_{T-1} & \dots & \omega y_{T-p''} & x_t & \dots & y_{T-q'} & Qx_t & \dots & Qx_{T-p''} \end{bmatrix}$$

$$\underline{\theta}' = \left[\omega_1^1 \dots \omega_1^p \quad \omega_w^1 \dots \omega_w^p \quad \alpha_1^0 \dots \alpha_1^q \quad \alpha_w^0 \dots \alpha_w^q \right]$$

Les vecteurs $\underline{Y}$, $\underline{\varepsilon}$ et la matrice $\underline{X}$ sont composés d'éléments vectoriels (indices du temps) dans lesquels s'incrivent les observations spatiales:

$$y_t = [y_{1t}, y_{2t}, \dots, y_{Nt}]' \quad x_t = [x_{1t}, x_{2t}, \dots, x_{Nt}]'$$

$$\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt}]'$$

On obtient une équation de régression multiple à H paramètres et NT observations où N est le nombre de zones et T la longueur de la période qui tient compte du maximum de décalage. Ainsi, si $K = \max(p', p'', q', q'')$, il faut $T+K$ observations réelles pour construire un système à T unités de temps. Ici, l'unité de temps "1" correspond à la première observation avec laquelle un décalage d'ordre K devient possible.

Pour estimer le vecteur paramétrique $\underline{\theta}$, il suffit de minimiser la somme des erreurs au carré ou, de façon équivalente, la multiplication du vecteur $\underline{e}$ avec sa transposée, soit:

$$\min_{\underline{\theta}} S = \underline{e}'\underline{e} \equiv [\underline{Y}-\underline{X}\underline{\theta}]'[\underline{Y}-\underline{X}\underline{\theta}] \equiv \underline{Y}'\underline{Y} - 2\underline{\theta}'\underline{X}'\underline{Y} + \underline{\theta}'\underline{X}'\underline{X}\underline{\theta} \quad (5.72)$$

Pour minimiser cette expression, on calcule la dérivée:

$$\frac{\partial S}{\partial \underline{\theta}} = -2\underline{X}'\underline{Y} + 2\underline{X}'\underline{X}\underline{\theta} = 0 \quad (5.73)$$

En l'annulant, on trouvera l'équation normale de régression avec une notation matricielle:

$$\hat{\underline{\theta}} = [\underline{X}'\underline{X}]^{-1}\underline{X}'\underline{Y} \quad (5.75)$$

Encore une fois, les propriétés des estimateurs dépendent surtout des hypothèses concernant la distribution des erreurs. Si le vecteur $\underline{e}$ est de loi $N(0, \sigma^2 \underline{I})$, $\hat{\underline{\theta}}$ est un estimateur non-biaisé et de variance minimale de $\underline{\theta}$. Ici, 0 est un vecteur de zéro de dimension $NT \cdot 1$ et la matrice identité $\underline{I}$ est de taille $NT \cdot NT$. En fait, $\sigma^2 \underline{I}$ symbolise une matrice variance-covariance $\underline{ee}'$ à variance constante (entrées diagonales égales à σ^2) et dépourvue d'autocorrélation temporelle et spatiale (entrées triangulaires nulles). Reprenons la démonstration classique des propriétés des estimateurs [voir Johnston, 1972]:

$$\hat{\underline{\theta}} = [\underline{X}'\underline{X}]^{-1}\underline{X}'\underline{Y} \quad \text{et selon} \quad (5.71)$$

$$\hat{\underline{\theta}} = [\underline{X}'\underline{X}]^{-1}\underline{X}'[\underline{X}\underline{\theta} + \underline{e}] \equiv \underline{\theta} + [\underline{X}'\underline{X}]^{-1}\underline{X}'\underline{e}$$

L'espérance mathématique de l'estimateur est établie selon:

$$E(\hat{\underline{\theta}}) = E(\underline{\theta}) + [\underline{X}'\underline{X}]^{-1}\underline{X}'E(\underline{e}) \quad (5.76)$$

$$\text{Puisque } E(\underline{e}) = 0, \quad E(\hat{\underline{\theta}}) = E(\underline{\theta}) \quad (5.77)$$

Par conséquent, si la moyenne des erreurs est constamment nulle à travers le temps et l'espace, l'estimateur OLS est «non-biaisé». La variance des paramètres estimés est déterminée par le carré des écarts aux valeurs théoriques. Ainsi:

$$\text{Var}(\hat{\underline{\theta}}) = E([\hat{\underline{\theta}} - \underline{\theta}][\hat{\underline{\theta}} - \underline{\theta}]') \quad (5.78)$$

D'après (5.75) et sachant la constance de la matrice $\underline{X}$:

$$\text{Var}(\hat{\underline{\theta}}) = E([\underline{X}'\underline{X}]^{-1}\underline{X}'\underline{ee}'\underline{X}[\underline{X}'\underline{X}]^{-1})$$

$$= [\underline{X}'\underline{X}]^{-1}\underline{X}'E(\underline{ee}')\underline{X}[\underline{X}'\underline{X}]^{-1} \quad (5.79)$$

Les hypothèses concernant le comportement de distribution de $\underline{e}$ nous disent:

$$E(\underline{ee}') = \sigma^2 \underline{I}, \quad \text{alors}$$

$$\text{Var}(\hat{\underline{\beta}}) = \sigma^2 [\underline{X}'\underline{X}]^{-1} \quad (5.80)$$

Dans ce cas, la variance des paramètres est minimale car Johnston [1972] démontre que tous les autres estimateurs non-biaisés ont une variance plus grande que celle de $\hat{\underline{\beta}}$. La matrice $\text{Var}(\hat{\underline{\beta}})$ est de taille $H \cdot H$ dont la diagonale mesure la variance des estimateurs et les éléments triangulaires s'identifient aux covariances des paramètres. D'après (5.80), on en déduit que la fiabilité des estimateurs est conditionnée par l'ajustement du modèle de régression avec les observations (via la variance résiduelle σ^2). Indirectement, la variance des estimateurs se gonflent sous l'effet de la multicollinéarité des régresseurs; des covariances élevées inscrites dans les portions hors-diagonales de $\underline{X}'\underline{X}$ tendent à réduire son déterminant, accroissant ainsi la magnitude de la matrice inversée [Philips, Blomme, 1981 chap.2].

Le schéma d'estimation OLS s'appliquent également dans les systèmes où coexistent plus d'une variable endogène (voir modèle 5.28). Les vecteurs régresses et paramétriques de (5.71) s'allongent en proportion respective du nombre de variables dépendantes et du nombre de paramètres tandis que la matrice de régresseurs contient, outre des relations spatio-temporelles, les nouvelles relations fonctionnelles des variables endogènes vis-à-vis les autres variables endogènes et les variables exogènes. On assiste à extension multivariée de la méthode des moindres carrés (MOLS) qui, du point de vue théorique, ne diffère en rien de sa version univariée (au sens des variables dépendantes) [voir Bennett, 1979, chap.4].

5.4.3 Les problèmes liés à la structure spatiale

Dans la section 5.4.1, on nota la faible possibilité d'observer une autocorrélation temporelle significative de l'erreur de par la brièveté des séries disponibles dans un contexte d'aménagement. Ce constat n'exclut cependant pas l'occurrence d'autocorrélation spatiale. Plusieurs facteurs engendrent une association statistique des déviations parmi les lieux qui

se côtoient. La configuration géographique des zones est un important facteur responsable; les limites administratives ne coïncident souvent pas avec des aires homogènes si bien que les attributs débordent les limites. Aussi, parmi les zones voisines, existe une certaine inertie des relations fonctionnelles de sorte que les erreurs d'ajustement de régression se ressemblent d'un lieu à l'autre.

La présence d'autocorrélation spatiale contrevient à la présomption de distribution du terme d'erreur. La matrice variance-covariance $\underline{ee'}$ possède une structure en partition du fait que $\underline{e} = \{e_t; t=1,2,\dots,T\}$ et $e_t = \{e_{it}; i=1,2,\dots,N\}$. En absence d'autocorrélation temporelle, les sous-matrices $e_t e_{t'}'$; $t \neq t'$ sont nécessairement nulles. Une corrélation spatiale implique toutefois des matrices $e_t e_t'$ non-nulles du type:

$$e_t e_t' = \sigma^2 R \quad (5.81)$$

$$\text{où } R_{N \times N} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1N} \\ r_{21} & 1 & r_{23} & \dots & r_{2N} \\ r_{31} & r_{32} & 1 & \dots & r_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & r_{N3} & \dots & 1 \end{bmatrix} \quad \text{et } R = R' \text{ (symétrique)}$$

La matrice R indique la corrélation spatiale présumée entre les N observations. Les éléments r_{ij} peuvent être nuls, négatifs ou positifs mais toujours contenus dans l'intervalle unitaire. Pour évaluer approximativement une telle matrice, on peut utiliser la matrice (de même taille) de structure spatiale la plus significative pour les erreurs en la multipliant toutefois par l'indice de Moran correspondant [voir Hordjik, Nijkamp, 1977]. Pour l'ensemble de la période d'observation, on a une matrice variance-covariance des erreurs:

$$E(\underline{ee'}) = \sigma^2 \underline{R} \quad (5.82)$$

$$\text{où } \underline{R}_{NT \times NT} = \begin{bmatrix} R & 0 & 0 & \dots & 0 \\ 0 & R & 0 & \dots & 0 \\ 0 & 0 & R & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & R \end{bmatrix}$$

Les sous-matrices spatiales R ne sont pas indicées du temps car elles sont présumées constantes à travers la période. Les différencier dans le temps ne changent en rien à la démonstration.

La méthode des moindres carrés généralisés d'Aitken [1934] (conventionnellement appelée GLS) s'accommode aux applications où les erreurs possèdent une structure de covariation connue. Elle convient aussi au cas d'hétéroscédasticité. Dans l'exemple présent, l'homoscédasticité est admise tant dans le temps que dans l'espace, d'une part, parce que les périodes d'observation (dans un contexte d'aménagement) ne peuvent démontrer une modulation significative de la variance et, d'autre part, une variance inconstante géographiquement peut être facilement rétablie en substituant les sous-matrices $\sigma^2 R$ par des formes plus générales G admettant des comportements variationnels distincts par zone.

La procédure GLS met à profit les propriétés particulières d'une matrice symétrique et positive définie comme $\underline{R}$. Une telle matrice peut être réexprimée par une forme PP' où P est non-singulière et de même taille que $\underline{R}$ [Johnston, 1972 chap.4]:

$$\underline{R} = PP' \quad \text{ou bien} \quad P^{-1}\underline{R}P'^{-1} = \underline{I} \quad \text{ou alors} \quad P'^{-1}P^{-1} = \underline{R}^{-1}$$

En pré-multipliant le modèle (5.71) par la matrice P^{-1} , on obtient:

$$P^{-1}\underline{Y} = P^{-1}\underline{X}\underline{\theta} + P^{-1}\underline{e} \quad (5.83)$$

dont la matrice variance-covariance des erreurs est:

$$\begin{aligned} E([P^{-1}\underline{e}][P^{-1}\underline{e}]') &= P^{-1}E(\underline{e}\underline{e}')P'^{-1} \\ &= \sigma^2 P^{-1}\underline{R}P'^{-1} = \sigma^2 \underline{I} \end{aligned} \quad (5.84)$$

de sorte que la formulation transformée (5.83) satisfait les présomptions des moindres carrés ordinaires. De plus:

$$\begin{aligned} \hat{\underline{\theta}} &= [(P^{-1}\underline{X})'(P^{-1}\underline{X})]^{-1}(P^{-1}\underline{X})'(P^{-1}\underline{Y}) \\ &= [\underline{X}'\underline{R}^{-1}\underline{X}]^{-1}\underline{X}'\underline{R}^{-1}\underline{Y} \end{aligned} \quad (5.85)$$

est un estimateur non-biaisé de variance minimale de $\underline{\theta}$ doté d'une matrice variance-covariance des paramètres:

$$\begin{aligned} \text{Var}(\hat{\underline{\theta}}) &= \sigma^2 [(P^{-1}\underline{X})'(P^{-1}\underline{X})]^{-1} \\ &= \sigma^2 [\underline{X}'\underline{R}^{-1}\underline{X}]^{-1} \end{aligned} \quad (5.86)$$

L'estimateur (5.85) est dit «estimateur GLS» car il s'accorde avec n'importe quelle structure de corrélation des erreurs définie par une matrice R . Dans les moindres carrés ordinaires, R est presumée égale à une matrice identité I , négligeant ainsi une éventuelle hétéroscédasticité et/ou une autocorrélation des erreurs.

La plupart des distributions spatiales ne sont pas stationnaires dans l'espace [Cliff et al., 1975]. Si un modèle contient une composante autorégressive spatiale (comme 5.49), il a fort à parier que le paramètre sera affecté par l'existence d'une tendance. Dans un système géographique, une tendance se manifeste par une régionalisation des variables, faisant apparaître des groupements de valeurs. Malheureusement, il semble difficile de distinguer ces effets «régionaux» de ceux émanant d'interactions spatiales. En effet, dans les deux cas, des associations (positives) surgissent.

A l'examen d'un graphique illustrant une série chronologique, la forme générale de la tendance se repère facilement car elle dessine souvent une droite ou des courbes quadratique et exponentielles. Sur une carte univariée, les agglomérations de valeurs sont propres à l'arrangement géographique de l'aire étudiée. L'appréhension de ces variations générales reste le domaine privilégié de l'«analyse de surface de tendance» ("Trend Surface Analysis") qui consiste à ajuster un polynôme bi-dimensionnel (des coordonnées géographiques) à la distribution spatiale [voir Norcliffe, 1969; Berlyant, 1970]. Comme on sait, l'élévation du degré polynômial (linéaire, quadratique, cubique, quartique...) améliore la flexibilité de la surface théorique et augmente alors le raffinement avec lequel on capture les variations. La surface appropriée s'identifie à une «composante régionale» tandis que ses résidus sont perçus comme une «composante locale» [Haggett, 1969]. La technique est parfois utilisée comme procédé d'élimination de l'autocorrélation [Cliff et al., 1975, chap.8] en l'appliquant successivement et en élevant les degrés jusqu'à l'obtention d'une distribution résiduelle aléatoire. Par conséquent, pour abstraire la vraie composante régionale sans supprimer les interactions univariées (du moins leurs manifestations), il faut choisir judicieusement le polynôme.

Pour établir ce choix, il convient de définir qu'est-ce qu'une composante régionale. Une tendance spatiale peut-être créée par le simple fait que, dans un secteur de la région, les zones ont généralement une taille supérieure à la moyenne: les phénomènes y apparaissent donc avec plus d'ampleur. Un système zonal avec tailles homogènes permet éviter ce biais géographique. Sinon, il y a lieu de standardiser la variable avec la superficie ou la population zonale. Toutefois, certains processus (bivariés) d'interaction spatiale ne s'évaluent qu'en terme absolu comme dans le cas d'échange de personnes, de biens ou de services. Il devient alors préférable d'utiliser des résidus d'une surface théorique qui ne modifie pas l'échelle de la variable initiale.

Evidemment, dans une application spatio-temporelle, il faut prendre garde de ces transformations de variables qu'elles soient exogènes ou endogènes. Les modifications numériques dans la dimension spatiale affectent le caractère stationnaire de la tranche temporelle. Les corrections des séries temporelles détruisent, pour leur part, les relations d'ordre de grandeur des variables mesurées dans les zones: les processus spatiaux sont alors impunément corrompus. Le maintien de la non-stationnarité spatiale de la variable endogène reste par ailleurs admissible puisqu'il n'affecte pas la qualité statistique des estimateurs paramétriques mais tend seulement à augmenter la valeur du paramètre autoregressif associé en confondant les interactions univariées et les effets régionaux.

Chapitre 6

MODELISATION SPATIO-DYNAMIQUE DE LA STRUCTURE URBAINE

6.1 INTRODUCTION

Ce chapitre ne s'inscrit pas dans aucun cadre théorique spécifique. On y traite de modèles de localisation, problèmes d'optimisation, multiplicateurs spatiaux, de systèmes d'équations différentielles, de systèmes spatio-temporels, des théories de catastrophe et de bifurcation et, enfin, des concepts d'auto-organisation de système. L'objectif principal n'est pas, comme on pourrait le croire, de faire un exercice énumératif de cadres théoriques pouvant se rapporter à la dynamique de la structure urbaine. Plutôt, il s'agit de proposer une formulation générale d'un modèle spatio-dynamique de localisation d'activités en employant et conciliant ces divers formalismes théoriques.

Les différentes théories sollicitées ici semblent souvent être "compartimentées" de sorte qu'un chercheur, dans l'analyse d'un phénomène, doit s'adresser à celle qui convient le mieux à ses aspirations et au type de problème confronté. Ainsi, selon le phénomène étudié, il existe un ou deux " tiroirs " dans lesquels se trouvent un certain nombre de définitions, de concepts et de méthodes appropriés.

L'approche de «conciliation» ne vise pas ou à tirer tous les " tiroirs " reliés à la dynamique de la structure urbaine et à choisir les éléments les plus pertinents. L'objectif implicite semble plutôt de démontrer la complémentarité des cadres théoriques et l'usage de cette complémentarité dans la construction et la justification d'un modèle complet de localisation d'activité.

La structure urbaine est vue ici sous l'angle de la localisation des activités. Ce choix est évidemment motivée par le raccordement entre la structure urbaine et les systèmes de transport. D'une part, les déplacements générés quotidiennement dépendent explicitement des activités urbaines, leurs types, leurs intensités et leurs localisations relatives. D'autre part, les processus de localisation spatiale sont conditionnées par

le patron d'accessibilité, lui-même tributaire des caractéristiques des réseaux de transport. Le sous-chapitre 6.2 décrit les formulations de modèles de localisation les plus conventionnelles et montre, par le fait même, que le processus en est un d'interaction. La structure fonctionnelle des modèles, gravitaire pour la plupart, rappelle celle des modèles de distribution spatiale des voyageurs.

Pourtant, la logique sous-jacente de ces modèles, nous apparaît peu manifeste. En préconisant une approche de problème d'optimisation, les principes d'allocation se démarquent. Ces principes relèvent tous plus ou moins d'une forme de maximisation des bénéfices de localisation. Les activités tendent donc à se distribuer géographiquement en fonction de certains avantages plus ou moins monnayables et de contraintes d'occupation. Les bénéfices procurés à une activité procèdent généralement des interactions avec les autres activités si bien que des «multiplicateurs spatiaux» traduisant les effets directs et indirects d'une activité vers une autre peuvent être calculés.

Les problèmes d'optimisation permettent de définir les conditions d'optimalité i.e. les conditions avec lesquelles la distribution spatiale des activités y trouve le plus son compte. Cet état d'équilibre implique que la totalité des effets inter-activités et spatiaux est acheminée et cumulée. En réalité, un tel état est rarement atteint. Il n'en reste pas moins que l'état d'équilibre demeure un état favorable et que le système tend à y parvenir. Cette convergence peut être assimilée à un processus temporel. Un système d'équations différentielles ou un système spatio-temporel (comme au chapitre 5) permet de représenter une telle dynamique. La solution du problème d'optimisation équivaut désormais à un état d'équilibre dynamique et les multiplicateurs spatiaux statiques se confondent à ceux issus d'un cadre spatio-temporel.

Si le problème d'optimisation comporte des relations non-linéaires au sein d'une même variable, il y a risque de multiplicité des états d'équilibre et, par ricochet, risque de «catastrophes» et de «bifurcations». Les discontinuités dynamiques qui en résultent participent sans doute à l'émergence de nouvelles structures spatiales ou, en d'autres mots, à l'auto-organisation de la distribution géographique des activités.

6.2 MODELES CLASSIQUES DE LOCALISATION

6.2.1 Un modèle urbain de synthèse; le modèle de Lowry

La publication de «Model of Metropolis» par Ira S. Lowry [1964] bouleversa passablement le domaine de la modélisation en géographie urbaine. Lowry proposa un modèle simple, déductif et opérationnel du développement urbain qui se fit beaucoup d'adeptes dans les organismes de planification. En effet, à l'origine, le modèle fut conçu pour prévoir l'impact des politiques d'aménagement sur la structure spatiale des activités d'une ville. Outre l'étude initiale portant sur Pittsburgh [Lowry, 1964], diverses versions furent appliquées ailleurs notamment à San Francisco [Goldner, 1968], dans des comtés fortement urbanisés du centre de l'Angleterre [Batty, 1976; Mackett, Mountcastle, 1977]. Mais sa vocation académique prit un essor encore plus remarquable si bien qu'une part inégale de la littérature méthodologique en géographie, en sciences régionales et en planification urbaine s'intéressa à son développement et à son raffinement. Les motifs de cet engouement tiennent surtout à sa souplesse, son caractère synthétique et sa facilité d'intégration avec d'autres modèles [Goldner, 1971].

Le modèle original de Lowry se présente comme un modèle de localisation spatiale d'activités, en particulier, des activités résidentielle et commerciale. La distribution spatiale de celles-ci y semble déterminée par celle de l'«emploi basique» i.e. exportateur de la région urbaine. Alors que l'activité commerciale est exprimée par l'«emploi non-basique», une distinction claire et précise entre «basique» et «non-basique» s'impose. Cette distinction apparaît un des principaux points faibles du modèle car, en pratique, elle reste difficile, arbitraire mais pourtant déterminante [Massey, 1971]. Lowry ne se préoccupa guère des processus expliquant la localisation de l'emploi exportateur et la considéra comme fixée exogènement. Toutefois, cette lacune peut être comblée sans difficulté comme le démontrèrent Putnam [1975: 1983] et Webber [1984].

Le modèle repose essentiellement sur la «théorie de la base économique» concernée par les propriétés multiplicatives de l'emploi basique sur l'emploi de service et sur la population [Samuelson, 1943]. Algébriquement, ce mécanisme se manifeste de la façon suivante. Soit P la population

totale, E l'emploi total, E^B l'emploi basique, S l'emploi commercial (de service). "a" l'inverse du taux d'activité et "o" un ratio "emploi commercial / services demandes per capita. Alors, on définit les relations numériques liant ces variables:

$$E = E^B + S \quad P = a \cdot E \quad S = o \cdot P$$

Les effets multiplicateurs de l'emploi basique sur l'emploi commercial sont dérivés à partir des relations précédentes:

$$\begin{aligned} S &= o \cdot P = o \cdot a \cdot E = o \cdot a \cdot (E^B + S) \\ (1 - oa) \cdot S &= o \cdot a \cdot E^B \\ S &= \frac{oa}{(1 - oa)} \cdot E^B \end{aligned} \quad (6.1)$$

et sur la population:

$$\begin{aligned} P &= a \cdot E = a \cdot (E^B + S) = a \cdot (E^B + o \cdot P) \\ (1 - oa) \cdot P &= a \cdot E^B \\ P &= \frac{a}{(1 - oa)} \cdot E^B \end{aligned} \quad (6.2)$$

Les coefficients de E^B dans (6.1) et (6.2) représente l'amplification moyenne générée par le secteur exportateur de la région. Le modèle de Lowry paraît en quelque sorte une version spatialement désagrégée de ce mécanisme multiplicateur. Les secteurs sont dorénavant présentés par zone et, conséquemment, l'effet multiplicatif se diffuse géographiquement.

La distribution spatiale des activités résidentielle et commerciale est assurée par deux sous-modèles d'interaction. Ceux-ci permettent effectivement de propager le mécanisme de base économique dans l'espace. La formulation initiale comprend aussi sept autres équations et trois inégalités qui décrivent tant contraintes que règles de consistance. La première équation concerne spécifiquement des restrictions quant à l'occupation du sol:

$$L_i = L_i^B + L_i^S + L_i^R + L_i^O \quad (6.3)$$

où L_i désigne la superficie totale de la zone i tandis que les suscrits B, S, P, et O identifient les parts accordées aux secteurs basique, commercial, résidentiel et inutilisable. Par ailleurs, L_i^B et L_i^P sont prédéterminées. L'allocation de la population zonale se base sur l'hypo-

thèse à l'effet que le lieu de domicile est fonction de la distance du lieu de travail:

$$P_i = g \cdot \sum_j E_j f^r(c_{ij}) \quad (6.4)$$

$$\text{sujet à } P_i \leq z^r L_i^r \quad (6.5)$$

où P_i : population de la zone i ; c_{ij} : coût de transport ou distance $i-j$;
 $f^r(\cdot)$: fonction atténuatrice de c_{ij} ; z^r : densité maximale résidentielle.
 La constante g est telle que:

$$\sum_i P_i = P \quad \text{où } P = \bar{a} \cdot E \quad (6.6)$$

L'importance totale du secteur commercial est fixée par la population totale selon l'équation $S = \bar{o} \cdot P$. La localisation commerciale considérée, à la fois, l'accessibilité à l'égard des lieux de résidence et des lieux de travail. Notons que des catégories de services peuvent être établies afin de tenir compte de la diversité des services. Ici, cependant, dans le but de faciliter la présentation, une version agrégée sera exposée. Ainsi:

$$S_j = \bar{o}^r \cdot \sum_i P_i f^r(c_{ij}) + \bar{o}^w \cdot E_j \quad (6.7)$$

$$\text{sujet à } S_j \geq S_j^{\min} \quad (6.8)$$

Les coefficients $\bar{o}^r$ et $\bar{o}^w$ désignent les poids relatifs de la consommation de service faite à partir de la résidence et du lieu d'emploi. Ils sont calculés tels que $\sum_j S_j = S$. La superficie vouée à l'activité commerciale dans une zone j est évaluée par:

$$L_j^c = z^c \cdot S_j \quad \text{sujet à } L_j^c \leq L_j - L_j^r + L_j^e \quad (6.9)$$

L'emploi total par zone reste tout simplement défini par:

$$E_j = S_j + E_j^e \quad (6.10)$$

L'ensemble des équations et des inégalités représente en fait un système non-linéaire simultané. En fait, la plupart d'entre elles s'avèrent des contraintes ou des conditions logiques qui offrent peu de substances théoriques quant à l'explication du processus. Seuls les sous-modèles (6.4) et (6.7) expriment un quelconque processus de localisation spatiale. Or, ils apparaissent un peu trop simplistes pour être adéquats. Une amélioration notable est apportée lorsqu'on concilie le formalisme de Lowry avec des modèles de localisation spatiale d'activités.

6.2.2 Modèles de localisation commerciale

Parmi les modèles de localisation d'une activité, celui concerné par le secteur commercial (biens et services) fut certes le plus sollicité. Un doit à Lakshmanan et Hansen [1965] la formulation la plus courante bien que sa forme s'inspire d'un modèle de probabilité de choix du consommateur de Huff [1964]. Le modèle alloue les dépenses des consommateurs des zones résidentielles vers des lieux d'achat (districts ou centres commerciaux). A chaque zone domiciliaire i , la population P_i et les dépenses per capita e_i sont connues. Le modèle se présente ainsi:

$$s_{ij} = a_i (e_i P_i) (W_j)^{\alpha} f^{\alpha}(c_{ij}) \quad (6.11)$$

où $a_i = 1/[\sum_j (W_j)^{\alpha} f^{\alpha}(c_{ij})]$ tel que $\sum_j s_{ij} = e_i P_i$

s_{ij} : dépenses faites dans le lieu j par les consommateurs de i ;

W_j : variable d'attraction commerciale (taille): $f^{\alpha}(\cdot)$: fonction inverse du coût de transport: α : paramètre d'économie d'échelle.

La variable de localisation devient la somme des dépenses effectuées dans la zone de consommation j et est obtenue par la sommation suivante:

$$S_j = \sum_i s_{ij} \quad (6.12)$$

Il s'agit ni plus ni moins d'une formulation gravitaire. le côté droit de l'équation (6.11) comportant un terme de «production» (le montant dépensé à la zone d'origine $\equiv e_i P_i$), un terme d'«attraction» (la taille des facilités commerciales $\equiv W_j$) et une fonction inverse de la distance ($f^{\alpha}(c_{ij})$). L'exposant α apparaît facultatif (les auteurs le fixèrent préalablement à un) et symbolise l'effet amplificateur des économies d'échelle pour le consommateur. Selon la terminologie de Wilson [1971a], le modèle d'achat de Lakshmanan et Hansen est un modèle d'interaction contraint à l'origine («attraction-constrained») du fait que les dépenses inter-zonales s_{ij} sont telles qu'elles correspondent exactement au montant moyen dépensé dans la zone résidentielle i .

Une telle structure fonctionnelle n'est pas réservée qu'à l'allocation spatiale des dépenses mais peut convenir à la localisation de l'emploi commerciale. Wilson [1971b] proposa une version analogique en substituant le terme $(e_i P_i)$ par $(\bar{o} P_i)$ et en donnant une signification différente au

terme d'interaction s_{ij} . En effet, si celui-ci est considéré non plus comme un flux monétaire mais comme une mesure arbitraire de demande, on peut assimiler S_j comme l'emploi requis pour satisfaire cette demande. Par conséquent, on obtient un modèle de localisation commerciale isomorphe à celui décrit par Lowry (eq. 6.7):

$$S_j = \sum_i s_{ij} = \sum_i a_i (\bar{e}P_i) W_j^{\alpha} f_{\theta}(c_{ij}) \quad (6.13)$$

où $a_i = 1/[\sum_j (W_j)^{\alpha} f_{\theta}(c_{ij})]$ tel que $\sum_j s_{ij} = \bar{e}P_i$

De nombreuses variantes des modèles d'achat (6.11) et (6.13) se révèlent possibles, notamment, en modifiant la fonction d'attraction. La fonction puissance:

$$g(W_j) = (W_j)^{\alpha}$$

reste la plus conventionnelle et implique que l'attraction n'est pas linéairement proportionnelle à la taille. Dépendant de la valeur du paramètre α , il peut y avoir économie d'échelle ($\alpha > 1$) ou diseconomie d'échelle ($\alpha < 1$). Toutefois, il n'y a pas de limite positive à l'attraction si bien qu'un consommateur est sans cesse plus attiré par de plus en plus grands centres commerciaux. Il y a sans doute une limite à l'attraction et elle peut être représentée à l'aide d'une fonction logistique:

$$g(W_j) = \frac{\kappa_1}{1 + \kappa_2 \exp\{-\kappa_3(W_j + \kappa_4)\}}$$

Celle-ci reproduit, pour certaines valeurs intermédiaires de W_j , une relation d'économie d'échelle et, à l'approche d'un certain seuil, une relation de diseconomie empêchant l'attraction de dépasser la limite κ_1 (Clarke, 1985).

Dans un modèle spatial mais statique comme (6.11-6.12), la modulation de la fonction de l'attraction n'a pas d'incidence théorique car elle n'affecte que l'évaluation du terme d'attraction. Ainsi:

$$G_j = g(W_j)$$

devient tout simplement une nouvelle variable calculée d'après W_j . La modification de la forme fonctionnelle et les variations paramétriques associées restent déterminantes si, d'une part, on introduit une relation non-linéaire entre la taille des facilités et la demande telle que:

$$W_j = f(S_j) \quad \text{et} \quad S_j = f(\bar{e}_1, P_1, W_j, c_{ij})$$

et si, d'autre part, on convertit le modèle d'achat en modèle dynamique. Un équilibre spatial existe pour S_j et W_j (qui devient endogène) (voir Poston, Wilson, 1977; Harris, Wilson, 1978; Clarke, 1985). Ce thème sera étudié spécifiquement dans la section 6.5.2.

Fotheringham [1985] remarqua que la formulation classique du modèle d'achat ne tenait compte que de la taille des facilités dans la zone et de la distance du marché mais négligeait les forces d'agglomération et de compétition. Pour un centre commercial, la proximité de d'autres centres peut à la fois être bénéfique en profitant du passage de la clientèle des facilités voisines (surtout entre des types de commerce complémentaires) mais aussi porter préjudice en présence de compétiteurs. Pour suppléer à cette lacune, il proposa d'insérer dans le modèle une nouvelle variable exprimant un indice d'accessibilité locale. Une mesure convenable de localisation relative reste la formule d'accessibilité potentielle de Hansen [1959], soit:

$$A_j = \sum_{k \neq j} \frac{W_k}{(C_{jk})^\eta}$$

Ordinairement, le paramètre η est choisi de manière à produire une fonction du coût de transport (distance) très sévère afin de ne considérer que le voisinage immédiat de la zone j . Le modèle (6.11) devient:

$$s_{1j} = a_1 (\bar{e}_1 P_1) (W_j)^\alpha (A_j)^\tau f_B(C_{1j}) \quad (6.14)$$

où $a_1 = 1/[\sum_j (W_j)^\alpha (A_j)^\tau f_B(C_{1j})]$ tel que $\sum_j s_{1j} = \bar{e}_1 P_1$

Lorsque l'exposant τ est positif, des forces agglomératives s'exercent dans le système étudié. Inversement, une valeur négative signale la présence de forces compétitives. Par ailleurs, l'introduction de la variable d'accessibilité locale convient également à la formulation (6.13) où une répartition de l'emploi non-basique est recherchée.

6.2.3 Modèles de localisation résidentielle

La localisation spatiale de l'activité résidentielle dans une région urbaine fut aussi un thème privilégié dans le domaine de la modélisation géographique. Le problème parut surtout sous l'angle de l'allocation d'individus de leurs lieux de travail (jugés fixes) vers des zones

domiciliaires. Le sous-modèle preconisé par Lowry [1964] (eq. 6.4) avait une structure de «modèle de potentiel» apparentée aux mesures de Hansen [1959]. Wilson [1969] vit plutôt le problème de localisation comme un phénomène d'interaction entre lieux de travail et lieux de résidence. En employant la méthode de maximisation de l'entropie, il suggéra un certain nombre de formulations apparentées. La forme la plus générale se présente comme un modèle de gravité contraint au lieu d'emploi:

$$p_{i,j} = Q_i b_j E_j f^p(c_{i,j}) \quad (6.15)$$

où $b_j = 1/[\sum_i Q_i f^p(c_{i,j})]$ tel que $\sum_i p_{i,j} = E_j$

Q_i : facteur d'attraction résidentielle de la zone i ; E_j : emploi total dans la zone j ; $f^p(\cdot)$: fonction inverse de la distance.

La population résidente estimée à la zone i devient une somme des travailleurs alloués $p_{i,j}$ multipliés par le taux inverse d'activité:

$$P_i = \bar{a} \cdot \sum_j p_{i,j} \quad (6.16)$$

Dans un contexte de planification, un tel procédé de localisation s'avère simpliste puisqu'il ne tient nullement compte des capacités d'accueil des zones résidentielles. En effet, les zones bien pourvues à l'égard de l'emploi ou celles à proximité de celles-ci se verront attribuer un nombre considérable d'habitants même si elles ne sont pas disposées à en loger tant. Alors, une formulation limitant l'accueil permettra d'allouer des résidents de façon plus réaliste en ne considérant non seulement l'attrait et l'accessibilité mais aussi la disponibilité. Ainsi (6.15-6.16) devient:

$$p_{i,j} = a_i Q_i b_j E_j f^p(c_{i,j}) \quad (6.17)$$

où $b_j = 1/[\sum_i a_i Q_i f^p(c_{i,j})]$ tel que $\sum_i p_{i,j} = E_j$

$$a_i = \begin{cases} 1/[\bar{a} \cdot \sum_j b_j E_j f^p(c_{i,j})] & \text{si } \bar{a} \cdot \sum_j p_{i,j} = Q_i \\ 1/\bar{a} & \text{si } \bar{a} \cdot \sum_j p_{i,j} < Q_i \end{cases}$$

Ici, Q_i apparaît une mesure de capacité résidentielle tout autant qu'une mesure d'attraction. Wilson [1969] proposa cette version semi-contrainte à l'attraction sans qu'il put la dériver en tenant compte des propriétés particulières des contraintes d'inégalité. Dans la prochaine section où sera abordée les méthodes de dérivation, on démontrera l'utilité de la théorie de Kuhn-Tucker en présence de telles contraintes de capacité.

Même si les cadres théoriques variaient, il y eut peu d'altération dans la forme gravitaire du modèle de localisation résidentielle. Les changements notables et bénéfiques ne tinrent qu'aux différentes désagregations possibles du modèle. Parmi celles-là, la plus intuitive reste sans doute celle subdivisant l'emploi en classes professionnelles et les logements en catégories de loyers [Senior, Wilson, 1974]. Examinons ici un modèle assujéti à des contraintes de capacité comme:

$$\bar{a} \cdot \sum_k p_{ij} \leq Q_i \quad \forall i, l \quad (6.18)$$

Alors:

$$p_{ij} = a_i Q_i b_j E_j f^k(c_{ij}) \quad (6.19)$$

où $b_j = 1 / [\sum_i \sum_l a_i Q_i f^k(c_{ij})]$ tel que $\sum_i \sum_l p_{ij} = E_j \quad \forall j, k$

et $a_i = 1 / [\bar{a} \cdot \sum_j \sum_l b_j E_j f^k(c_{ij})]$ si (6.18) satisfait en égalité et

$a_i = 1/\bar{a}$ si (6.18) satisfait qu'en inégalité

p_{ij} : travailleurs de type k dans j habitant un logement de type l dans i ;

Q_i : nombre de logements de type l dans i ; E_j : nombre de travailleurs de

type k dans j ; $f^k(\cdot)$: fonction de transport typique à la catégorie k .

Dans les trois dernières sections, divers modèles de localisation ont été successivement présentés sans qu'on traite de leurs fondements théoriques. On insista surtout sur leurs structures fonctionnelles et sur leurs composantes (variables, paramètres, sous-fonctions, exposants). La section suivante portera sur la composition de problèmes d'optimisation permettant de dériver ces modèles et avec lesquels on retrouve des justifications théoriques d'ordre statistique, économique ou relevant de l'aménagement.

6.3 PROBLEMES D'OPTIMISATION ET LOCALISATION D'ACTIVITE

6.3.1 Formulations normatives

Lorsqu'on parle de problèmes d'optimisation et des phénomènes de localisation, on fait immédiatement allusion aux problèmes d'allocation optimale de facilités. Ces problèmes, essentiellement normatifs, vise à une répartition idéale d'un équipement (hôpital, école, services publics, etc...) de manière à réduire les coûts et à accroître les bénéfices de toute nature. Le problème d'allocation quadratique de Koopmans et Beckmann [1957] figure parmi les exemples les plus classiques. Les programmes quadratiques s'avèrent appropriées pour le traitement des interactions (entre facilités), des interdépendances et des externalités. Le «modèle Koopmans-Beckmann» fut originalement conçu pour une allocation indivisible (0 ou 1) des facilités; le traitement binaire convient plutôt à des établissements de taille invariable mais cette restriction peut être omise [Hopkins, 1979]. Voici une formulation avec une variable continue:

$$\min_{\{x\}} \sum_{ik} X_k C_k - \sum_{i,j,k,l} g_{ijkl} X_i X_j X_k X_l \quad \text{où } g_{ijkl} = g_{jikl} \quad \forall i,j,k,l \quad (6.20)$$

$$\text{sujet à } \sum_i X_i = X^* \quad \forall k \quad X_i \geq 0 \quad \forall i,k \quad (6.21-6.22)$$

X_k : variable d'allocation (activités k et lieu i); C_k : coût d'établissement d'une activité k dans un lieu i ; X^* : volume d'activité requis pour le système; G_{ijkl} : coefficient d'interaction entre lieux et activités.

Le programme est tel que les interactions bénéfiques (externalités) entre lieux et activités sont maximisées alors que les coûts d'établissement sont minimisés. La signification exacte des coefficients d'interaction reste à être déterminée. Les conditions nécessaires d'optimalité de Kuhn-Tucker rappellent que:

$$C_k - 2 \cdot \sum_{j,l} g_{ijkl} X_j X_l + \lambda^k \begin{cases} = 0; & X_k \geq 0 \\ > 0; & X_k = 0 \end{cases} \quad (6.23)$$

$$\sum_{ik} X_i = X^* \quad \text{et } \lambda^k \text{ sans restriction} \quad (6.24)$$

Ces conditions impliquent que si les coûts d'établissement (plus une constante relative à chaque activité, λ^k) dépassent les effets d'interaction, il n'est pas judicieux d'établir l'activité k dans la zone i . Il y a allocation si on obtient un équilibre entre coûts et bénéfices d'interaction; à

l'optimalité, l'équilibre concerne d'ailleurs toutes les zones et les activités. Les multiplicateurs lagrangiens apparaissent ici comme des «prix-fantômes» lesquels indiquent un coût ou un escompte implicite pour chaque activité k . Ils servent alors à relativiser les dépenses vraies avec les externalités. De par sa nature quadratique, le problème est convexe et les conditions nécessaires K-T sont également suffisantes.

La programmation linéaire offre une alternative valable dans la modélisation de processus d'allocation gouvernés par des principes optimisants. Un des problèmes typiques en micro-économie est la résolution d'un équilibre de marché dans lequel les individus sont vus comme des compétiteurs dans leurs choix résidentiels. Décrit par Alonso [1964], cet équilibre se rapporte aux prix d'enchère que les travailleurs sont prêts à payer pour un logement d'une certaine catégorie dans une zone spécifique vis-à-vis le coût de transport entre domicile et lieu d'emploi. Herbert et Stevens [1960] ont repris ce problème sous une forme de problème de transport en programmation linéaire. La variable d'allocation, p_{kjl} , désigne le nombre de travailleurs du type k travaillant dans la zone j mais optant pour une résidence de type l dans une zone i :

$$\max_{\{p\}} = \sum_i \sum_j \sum_k \sum_l p_{kjl} (b_{kjl} - c_{ij}) \quad (6.25)$$

$$\text{sujet à} \quad \sum_k p_{kjl} \leq Q_l \quad (6.26)$$

$$\sum_l p_{kjl} = E_j \quad p_{kjl} \geq 0 \quad \forall ijk \quad (6.27-6.28)$$

Les variables ont la même signification que dans le modèle (6.18-6.19) alors que la variable b_{kjl} symbolise le prix d'enchère qu'un travailleur de type k dans j peut offrir pour un logement de type l dans i . La fonction objective montre que les offres moins les coûts de transport sont maximisées signifiant ainsi la présence d'un marché compétitif. Ici, les conditions K-T apparaissent très révélatrices:

$$b_{kjl} - c_{ij} - \eta_{il} - v_{jk} \begin{cases} = 0; & p_{kjl} \geq 0 \\ < 0; & p_{kjl} = 0 \end{cases} \quad (6.29)$$

$$\eta_{il} \cdot (\sum_j \sum_k p_{kjl} - Q_l) = 0; \quad \eta_{il} \geq 0 \quad (6.30)$$

$$\sum_i \sum_l p_{kjl} - E_j = 0; \quad v_{jk} \text{ sans restriction} \quad (6.31)$$

La première condition signale qu'il n'y a pas d'allocation si le prix offert est excédé par le coût de transport, le prix-fantôme d'un logis l de la zone i et le prix-fantôme des travailleurs k de la zone j . Ce dernier est sans restriction mais en se référant à la définition des multiplicateurs lagrangien de la section 2.3.2, on sait qu'il dépend de E_{ij} i.e. du marché des acheteurs de type k et lieu j . L'augmentation de ceux-ci pousse à la baisse v_{jk} et, par conséquent, favorise leur position d'acheteur tout en pressant à la hausse le prix offert net. Par contre, le «marché des vendeurs» Q_i diffère du fait que si l'allocation est inférieure (disponibilité de logement), le prix-fantôme est nul. Si l'allocation égale le stock disponible (saturation de ce sous-marché l,i), le prix est nécessairement positif et réduit les possibilités d'achat.

Dans le modèle Herbert-Stevens, le coût de transport n'apparaît pas un facteur d'interaction important mais plutôt comme une dépense particulière. Par surcroît, le problème linéaire reste très peu descriptif, les comportements individuels étant invariablement rationnels. On retrouve dans le modèle «TOPAZ» (Techniques for the Optimal Placement of Activities in Zones) [Brotchie, 1969; Brotchie, Toakley, Sharpe, 1971] cette volonté de considérer la variabilité des comportements dans un programme d'optimisation. La formulation mise de l'avant par TOPAZ s'accorde bien avec les théories en vigueur en géographie et concerne la localisation d'activité au sens large. La fonction objective et quelques contraintes s'apparentent beaucoup au problème Koopmans-Beckmann (version continue). Le modèle consiste à allouer des activités de manière à minimiser les coûts d'établissement et les coûts de transport en satisfaisant des quantités requises par secteur. Toutefois, le terme quadratique d'interaction de (6.20) est substitué par le terme des coûts de transport comprenant une matrice O-D désignant les interactions générées par l'allocation courante. La matrice O-D est contruite à partir d'un modèle de gravité doublement contraint. Le modèle TOPAZ comprend aussi des contraintes d'occupation du sol:

$$\min_{\{x\}} \sum_{i,k} x_k^i C_k^i + \sum_{i,j} x_{i,j} c_{i,j} \quad (6.32)$$

$$\text{où } x_{i,j} = a_i O_i b_j D_j f(c_{i,j}) \quad O_i = \sum_k g^k x_k^i \quad D_j = \sum_k h^k x_j^k \quad (6.33)$$

$$\text{sujet à } \sum_i x_k^i = X^k \quad \sum_k z^k x_k^i \leq L_i \quad x_k^i \geq 0 \quad (6.34-6.36)$$

Les variables gardent les mêmes définitions que dans le modèle Koopmans-Beckmann (6.20-6.22), z^k désigne la densité de l'activité k et L_i mesure la superficie de la zone i . Les coefficients a_i et b_j sont tels que la matrice $\{x_{ij}\}$ respecte les totaux marginaux d'origine O_i et de destination D_j . Le problème est solutionné en résolvant itérativement le programme apparemment linéaire que constituent la fonction objective (6.32) et les contraintes (6.34-6.36) en redéfinissant à chaque séquence la configuration migratoire $\{x_{ij}\}$. Il n'est pourtant plus linéaire ni même quadratique de par l'inclusion implicite d'une équation multiplicative comme (6.35).

Dickey et Najafi [1973], en essayant diverses solutions, remarquèrent le peu de sensibilité des coûts d'établissement en comparaison à la très haute variabilité causée par les interactions spatiales. La stabilité du terme linéaire est probablement due à l'exercice des contraintes (6.34) et (6.35). Sachant le rôle déterminant des interactions, l'utilisateur est confronté aux choix de la fonction de transport $f(c_{ij})$ et de la valeur de son paramètre. Des choix arbitraires impliquent par conséquent des solutions arbitrairement distinctes et donc peu révélatrices. Cette ambivalence tient à la double provenance du modèle, l'une rationnelle (problème d'optimisation), l'autre heuristique (sous-modèle de gravité).

6.3.2 Intégration de l'entropie dans un problème de localisation

Les possibilités d'intégration d'un cadre de programmation mathématique peuvent être mises à contribution dans l'établissement de liens entre un problème optimisant et un phénomène d'interaction. Evans [1973] démontra que le fameux «problème de transport en programmation linéaire» est un cas extrême d'un problème d'optimisation duquel dérive un modèle d'interaction spatiale. Elle mit aussi en relief les formulations alternatives du problème de distribution spatiale en intervertissant contraintes en termes de la fonction objective et vice et versa. La section 3.3.3 aborde partiellement ces aspects. En exploitant ces propriétés, on exposa l'équivalence d'un problème linéaire sollicitant un terme d'interaction avec un problème non-linéaire dotée d'une fonction d'entropie [Coelho, Wilson, 1977; Coelho, Williams, Wilson, 1978].

Considérons un problème de localisation de deux activités X^A et X^B qui suscitent des interactions spatiales x_{ij} . La fonction objective comprendra, tout comme (6.32), un terme mesurant les coûts d'établissement et un autre mesurant les coûts d'interaction (disons de transport). Cependant, au lieu de définir l'interaction comme un produit d'un modèle gravitaire, on cherchera plutôt la matrice d'interaction qui maximisera l'entropie. En effet, une telle démarche permet, comme on sait, de dériver un modèle de type gravitaire. Voici le programme non-linéaire conséquent:

$$\min_{\{x, X\}} \sum_i (X_i^A C_i^A + X_i^B C_i^B) + \sum_{i,j} x_{ij} c_{ij} + \frac{1}{\psi} \sum_{i,j} x_{ij} (\ln x_{ij} - 1) \quad (6.37)$$

$$\text{sujet à} \quad \sum_i X_i^A \geq X^A \quad \sum_i X_i^B \geq X^B \quad (6.38-6.39)$$

$$z^A X_i^A + z^B X_i^B \leq L_i \quad X_i^A, X_i^B \geq 0 \quad (6.40-6.41)$$

$$X_i^A - \sum_j x_{ij} = 0 \quad X_i^B - \sum_j x_{ij} = 0 \quad (6.42-6.43)$$

Les contraintes (6.38-6.41) sont identiques à celles qu'on retrouve dans le modèle TOPAZ (6.34-6.36). On y a rajouté des contraintes marginales de la matrice d'interaction (6.42-6.43). Ici, il n'y a pas de restriction de non-négativité de $\{x_{ij}\}$ puisque la fonction d'entropie, requérant le logarithme de x_{ij} , exclue toute valeur négative ou nulle. On est en présence d'un programme convexe car la minimisation s'exerce sur une fonction objective strictement convexe (les deux premiers termes sont linéaires tandis que le dernier est naturellement convexe) et dans un ensemble faisable convexe puisque toutes les contraintes sont linéaires (voir section 2.2.3). Dans ce cas, les conditions nécessaires d'optimalité de Kuhn-Tucker sont suffisantes car tout minimum local est le minimum global. Ces conditions s'écrivent comme suit; pour x_{ij} :

$$c_{ij} + \frac{1}{\psi} \ln x_{ij} + \eta_i + v_j = 0; \quad x_{ij} > 0 \quad (6.44)$$

$$x_{ij} = \exp(-\psi c_{ij} + \psi \eta_i + \psi v_j) \quad (6.45)$$

$$\text{où } \eta_i = \frac{1}{\psi} \ln X_i^A - \frac{1}{\psi} \ln \sum_j \exp(-\psi c_{ij} + \psi v_j)$$

$$\text{où } v_j = \frac{1}{\psi} \ln X_j^B - \frac{1}{\psi} \ln \sum_i \exp(-\psi c_{ij} + \psi \eta_i)$$

Pour X_i^A et X_j^B :

$$C_i^A + \lambda^A + \zeta_i z^A + \eta_i = 0; \quad X_i^A > 0 \quad (6.46)$$

$$C_j^B + \lambda^B + \zeta_j z^B + v_j = 0; \quad X_j^B > 0 \quad (6.47)$$

On ajoutera que $\lambda^A, \lambda^B \leq 0$ et que $\tau_i \geq 0$ ($\forall i$). La positivité des variables de localisation X_i^A et X_i^B est justifiée par leur apparition sous une forme logarithmique dans la définition des multiplicateurs lagrangiens η_i et ν_j . De plus, les contraintes marginales (6.42-6.43) imposaient déjà une stricte positivité car la somme de nombres positifs reste positive. Cette positivité implique que les dérivées partielles (6.46) et (6.47) s'annulent systématiquement car il n'y a plus de solution frontière pour les variables de localisation. La présence de η_i et de ν_j dans les conditions respectives de X_i^A et de X_i^B facilite la recherche de la solution car les multiplicateurs sont des fonctions explicites de ces variables d'allocation. Ainsi, on s'apercevra que X_i^A et X_i^B deviennent des fonctions exponentielles en réorganisant (6.46) et (6.47). Ce dénouement particulier s'explique aisément du fait que les variables de localisation sont des sommes des variables d'interaction de par les contraintes marginales (6.42-6.43). Par conséquent, un programme destiné à dériver une distribution d'activité peut se contenter du traitement de la variable d'interaction à condition que celle-ci soit liée numériquement aux activités. Le problème (6.37-6.43) à deux variables décisionnelles est convertible en un comme celui-ci :

$$\min_{\{X\}} \sum_{i,j} x_{i,j} (c_{i,j} + C_i^A + C_j^B) + \frac{1}{\psi} \sum_{i,j} x_{i,j} (\ln x_{i,j} - 1) \quad (6.48)$$

$$\text{sujet à } \sum_{i,j} x_{i,j} \geq \max(X^A, X^B) = X \quad \sum_j (x_{i,j} z^A + x_{j,i} z^B) \leq L_i \quad (6.49-6.50)$$

On établira la distribution d'activités avec la matrice solution $\{x_{i,j}\}$:

$$X_i^A = \sum_j x_{i,j} \quad X_j^B = \sum_i x_{i,j}$$

Le modèle présenté précédemment est équivalent au modèle TOPAZ puisque le terme d'interaction provient d'un modèle d'interaction de type gravitaire. On verra ultérieurement que cette équivalence est fondamentale dans l'intégration des modèles de localisation car elle sert de charnière entre différentes formulations issues de cadres théoriques distincts. Ainsi, en s'appuyant sur le théorème d'équivalence de Coelho et Wilson [1977], on a vu qu'un modèle d'allocation optimale des activités tenant compte des interactions peut être substitué par un problème convexe de minimisation des coûts comprenant un terme de dispersion.

6.3.3 Bénéfices de localisation

La formulation (6.48-6.50) s'apparente à un problème de maximisation des bénéfices de localisation [voir Williams, Senior, 1978]. Au lieu d'utiliser des coûts d'établissement, on préférera une mesure de bénéfices (qui peut pourtant être négative) rattachées au choix de localisation. Le coût de transport, toujours présent, symbolisera la désutilité de se localiser dans un lieu à l'égard d'une autre activité. Par exemple, définissons un problème de localisation résidentielle:

$$\max_{\{P\}} \frac{-1}{\psi^P} \sum_i \sum_j p_{ij} \{ \ln(p_{ij}/Q_i) - 1 \} - \sum_i \sum_j p_{ij} c_{ij} \quad (6.51)$$

$$\text{sujet à} \quad \sum_i p_{ij} = E_j \quad (6.52)$$

où Q_i : variable d'attraction résidentielle de la zone i ; E_j : emploi totale de la zone j ; p_{ij} : nombre de travailleurs de j habitant i .

Ce programme vise à distribuer les travailleurs à partir de leurs lieux d'emploi vers des zones de résidence de manière à minimiser leurs coûts de déplacement et à maximiser leurs bénéfices de localisation symbolisés par:

$$\ln(Q_i)/\psi^P$$

Cette mesure est comprise dans le terme d'entropie de (6.51) lequel s'apparente à une mesure d'entropie conditionnelle de Kullback [1959]. Dépourvu d'inégalité, ce programme est solutionné lorsque:

$$\frac{-1}{\psi^P} \ln p_{ij} + \frac{1}{\psi^P} \ln Q_i - c_{ij} + v_j = 0 \quad (6.53)$$

$$p_{ij} = Q_i \exp(-\psi^P c_{ij} + \psi^P v_j) \equiv Q_i b_j E_j \exp(-\psi^P c_{ij}) \quad (6.54)$$

où $b_j = 1/[\sum_i Q_i \exp(-\psi^P c_{ij})]$

On notera que du programme convexe (6.51-6.52), on dérive un modèle gravitaire contraint au lieu d'emploi identique à celui présenté dans la section 6.2.3 (voir 6.15-6.16). La forme de la fonction de transport est désormais spécifiée:

$$f^P(c_{ij}) = \exp(-\psi^P c_{ij})$$

Le multiplicateur de Lagrange v_j lié à la contrainte marginale du lieu d'emploi s'interprète communément comme un indice d'accessibilité [Wilson, 1974]. Ainsi, d'après (6.52) et (6.54):

$$v_j = \frac{1}{\psi^P} \ln E_j - \frac{1}{\psi^P} \ln \sum_i Q_i \exp(-\psi^P c_{ij}) \equiv \frac{1}{\psi^P} \ln E_j + \eta_j \quad (6.55)$$

Le terme η_j désigne une mesure d'utilité composée du lieu j en relation avec l'ensemble des zones résidentielles. La forme fonctionnelle reste parfaitement isomorphe avec les indices d'accessibilité d'Hansen [1959].

Dans le cas où Q_i désigne le stock de logement dans la zone i , il convient d'ajouter au problème (6.51-6.52) la contrainte d'inégalité suivante:

$$\bar{a} \cdot \sum_j p_{ij} = P_i \leq Q_i \quad (6.56)$$

La solution contient alors une série supplémentaire de multiplicateurs:

$$\frac{-1}{\psi^P} \ln p_{ij} + \frac{1}{\psi^P} \ln Q_i - c_{ij} + \bar{a} \eta_i + v_i = 0 \quad (6.57)$$

$$\bar{a} \cdot \sum_j p_{ij} - Q_i \begin{cases} = 0; & \eta_i \leq 0 \\ < 0; & \eta_i = 0 \end{cases} \quad (6.58)$$

On obtient alors le modèle doublement contraint:

$$p_{ij} = a_i Q_i b_j E_j \exp(-\psi^P c_{ij}) \quad (6.59)$$

$$\text{où } a_i \begin{cases} = 1/[\bar{a} \cdot \sum_j b_j E_j \exp(-\psi^P c_{ij})]; & \bar{a} \cdot \sum_j p_{ij} = Q_i \\ = 1/\bar{a}; & \bar{a} \cdot \sum_j p_{ij} < Q_i \end{cases}$$

$$\text{car } a_i = \exp(\psi^P \bar{a} \eta_i)$$

Cette démonstration basée sur la théorie de Kuhn-Tucker vaut pour toute contrainte marginale d'inégalité. Elle explique la définition particulière des coefficients de balancement dans le modèle de localisation résidentielle (6.17).

Senior et Wilson [1974] proposèrent une version du modèle Herbert-Stevens qui comportait toutefois un terme d'entropie. Des lors, il ne s'agit plus d'un problème de programmation linéaire car l'introduction d'une variabilité dans le processus de localisation assure une solution de type exponentiel. Selon le formalisme du «bénéfice de localisation», la variable s'identifiant au prix d'enchère (b_{kj}) devient une mesure additionnelle de l'utilité qu'offre un choix " $i \cdot l$ " pour un travailleur " $j \cdot k$ ". Le modèle solution ressemble à (6.59) car une contrainte d'égalité concerne le lieu d'emploi et une d'inégalité pour le lieu de résidence (voir 6.26-6.27). Mais, les variables sont davantage désagrégées et

l'équation contient une fonction exponentielle du prix d'enchère. Los [1979] montra que la variable du coût de transport n'a pas à être fixée exogènement mais peut émaner d'un trafic d'équilibre; l'intégration entre un modèle résidentiel un modèle d'équilibre de réseau se fait analogiquement à celle concernant la distribution spatiale des voyageurs et l'affectation routière (voir sous-chapitre 3.7).

Un modèle de localisation commerciale issu du modèle d'achat de Lakshmanan et Hansen [1965] peut également être dérivé par maximisation des bénéfices de localisation. Wilson [1976] définit une mesure de «bien-être du consommateur» comme étant déterminée par le coût de transport vers le lieu d'achat et par les «bénéfices de la taille» du lieu d'achat. Wilson proposa un programme non-linéaire servant à évaluer une distribution optimale des facilités d'achat et maximisant ce soi-disant «bien-être du consommateur»:

$$\max_{\{W\}} \sum_{i,j} s_{i,j} \{(\alpha/\psi^\alpha) \ln W_j - c_{i,j}\} \quad (6.60)$$

$$\text{sujet à (6.11) et } \sum_j W_j = W \quad (6.61)$$

où W_j : taille des facilités commerciales dans la zone j ; W : total admissible; " $(\alpha/\psi^\alpha) \ln W_j$ ": bénéfices de la taille.

Ce problème reste intrinsèquement non-linéaire car W_j se retrouve dans la mesure de bien-être et dans le terme d'interaction $s_{i,j}$ évalué d'après (6.11). A cause de cela, il apparaît difficile à résoudre. On remarque cependant son affinité avec le modèle TOPAZ car tous deux impliquent un terme généré par un modèle d'interaction spécifié exogènement. Coelho et Wilson [1976] avancèrent une formulation équivalente en référant au théorème d'équivalence qu'ils exposèrent dans un autre article [Coelho, Wilson, 1977]. Ainsi, (6.60-6.61, 6.11) devient:

$$\max_{\{W,S\}} \sum_{i,j} s_{i,j} \{(\alpha/\psi^\alpha) \ln W_j - c_{i,j}\} - \frac{1}{\psi^\alpha} \sum_{i,j} s_{i,j} (\ln s_{i,j} - 1) \quad (6.62)$$

$$\text{sujet à } \sum_j s_{i,j} = \bar{e}_i P_i \quad \sum_j W_j = W \quad (6.63-6.64)$$

avec comme solution:

$$\frac{\alpha}{\psi^\alpha} \ln W_j - c_{i,j} - \frac{1}{\psi^\alpha} \ln s_{i,j} + \eta_i = 0 \quad (6.65)$$

$$\text{i.e. } s_{i,j} = W_j^\alpha \exp(-\psi^\alpha c_{i,j} + \psi^\alpha \eta_i)$$

$$\text{où } \eta_1 = \frac{1}{\psi^\theta} \ln(\bar{e}_1 P_1) - \bar{\eta}_1 \quad \text{et} \quad \bar{\eta}_1 = \frac{1}{\psi^\theta} \ln \sum_j W_j^\alpha \exp(-\psi^\theta c_{1,j})$$

$$\text{donc } s_{1,j} = (\bar{e}_1 P_1) (W_j)^\alpha \exp(-\psi^\theta c_{1,j} - \psi^\theta \bar{\eta}_1) \quad (6.66)$$

$$\frac{\alpha \cdot \sum_1 s_{1,j}}{\psi^\theta \cdot W_j} - \lambda = 0 \quad \text{soit} \quad W_j = \frac{\alpha \cdot \sum_1 s_{1,j}}{\psi^\theta \lambda} \quad (6.67)$$

Comme prévu, $s_{1,j}$ provient d'un modèle gravitaire identique à celui de Lakshmanan et Hansen. Par ailleurs, la taille optimale W_j devient une fonction linéaire explicite de la demande totale à j ($S_j = \sum_1 s_{1,j}$). Ces résultats corroborent avec ceux du modèle de localisation résidentielle (6.15-6.16) où la variable de localisation est une somme des interactions. Par contre, W_j étant à la fois variable de localisation et d'attraction, une interdépendance entre les équations (6.66) et (6.67) s'ensuit. La solution suscite alors un état d'équilibre.

Dans le problème de localisation résidentielle, le terme $(\ln Q_1 / \psi^\theta)$ est vu comme la mesure brute de bénéfice de résidence dans la zone i . Le terme de bénéfice brut de consommation est quant à lui lié à la taille des facilités commerciales. Dans le programme (6.62-6.64), on semble rechercher la structure spatiale de W_j qui avantagera le plus les consommateurs i.e. leur offrira le plus de facilités aux moindres coûts de transport. Le modèle apparaît «normatif» car il semble démontrer une problématique d'aménagement en faveur des consommateurs. En fait, il est bien connu qu'un tel modèle fournit une situation d'équilibre entre les tailles et la demande du type $W_j - \kappa \cdot S_j = 0$ [Harris, Wilson, 1978; Harris, Choukroun, Wilson, 1982]. Cet équilibre n'est pourtant pas celui envisagé par les consommateurs mais maximise plutôt les profits nets du secteur commercial. En effet, la distribution optimale de W_j est celle qui ajuste la demande S_j aux coûts implicites d'exploitation (présupposés proportionnels à W_j). Le critère de «bien-être du consommateur» de Wilson [1976] correspond donc ironiquement à un critère de rentabilité des facilités. En ce sens, le modèle apparaît «descriptif» car il produit une distribution commerciale à la fois tributaire des décisions des commerçants et s'accordant avec les impératifs du marché.

Leonardi [1978] posa le problème de localisation commerciale sous un angle différent. En effet, sa formulation exclut le terme d'interaction et concerne la maximisation de l'accessibilité commerciale à l'égard du

marché. Une telle forme implique une repartition spatiale des commerces qui favorise le maximum de contacts (à motif de consommation) ou, selon le côté où l'on se situe, permet aux consommateurs d'accéder au maximum de facilités. Voici le problème de Leonardi:

$$\max_{\{W\}} \sum_i (\bar{e}_i P_i) \cdot \ln \sum_j W_j \exp(-\psi^{\alpha} c_{i,j}) \quad (6.68)$$

$$\text{sujet à} \quad \sum_j W_j = W \quad (6.69)$$

La solution optimale stipule que le rapport entre la demande générée et la taille doit être égale à travers le système spatial:

$$\frac{\sum_i s_{i,j}}{W_j} = \lambda \quad (6.70)$$

où λ est le multiplicateur lié à la contrainte (6.69)

$$\text{et } s_{i,j} = \frac{(\bar{e}_i P_i) W_j \exp(-\psi^{\alpha} c_{i,j})}{\sum_j W_j \exp(-\psi^{\alpha} c_{i,j})} \quad (6.71)$$

On retrouve des conditions d'optimalité équivalentes à celles fournies par le problème de maximisation des bénéfices de localisation lequel sollicite explicitement un terme de consommation. La raison est que le terme d'accessibilité inclus dans la fonction objective (6.68) équivaut au multiplicateur η_i lié à la contrainte marginale du marché (6.63). Par ricochet, le modèle de Leonardi est la formulation «duale» du problème «primal» (6.62-6.64) mais sans exposant pour W_j (cette distinction est facultative). En effet, le problème (6.68-6.69) est équivalent à:

$$\min_{\{\eta, \lambda\}} \frac{1-\alpha}{\psi^{\alpha}} \sum_{i,j} s_{i,j} / \psi^{\alpha} - \sum_i (\bar{e}_i P_i) \cdot \eta_i + W \cdot \lambda \quad (6.72)$$

sujet à (6.65) et (6.67)

Sachant que la variable duale η_i est définie à l'optimalité comme le multiplicateur de Lagrange associé à la contrainte (6.63) du problème (6.62-6.64), la fonction objective duale devient:

$$\min_{\{\eta, \lambda\}} \frac{1-\alpha}{\psi^{\alpha}} \sum_{i,j} s_{i,j} / \psi^{\alpha} - \sum_i (\bar{e}_i P_i) [\ln(\bar{e}_i P_i) / \psi^{\alpha} - \eta_i] + W \cdot \lambda \quad (6.73)$$

A l'optimalité, la somme des $s_{i,j}$ est constante de même que l'entropie de $(\bar{e}_i P_i)$. Alors:

$$\min_{\{\eta, \lambda\}} \sum_i (\bar{e}_i P_i) \cdot \eta_i + W \cdot \lambda \quad (6.74)$$

Par rapport à la formulation de Leonardi (6.68), il s'agit dorénavant d'une minimisation au lieu d'une maximisation. Pourtant, le problème est équivalent car les variables décisionnelles deviennent des variables duales.

Le concept de bénéfice de localisation apparaît comme une variante du concept d'utilité. Ici, le processus de maximisation est déterministe (au sens donné par Beckmann et Golob [1971]) car la mesure d'utilité est spécifiée a priori et il n'y a aucun terme stochastique d'erreur. Pourtant, il est bien clair que les deux formes de maximisation conduisent à la dérivation de modèles isomorphes. L'approche basée sur un terme d'erreur aléatoire s'appuie sur un comportement rationel des individus mais presume une distribution d'erreur de perception. L'approche dite déterministe incorpore une mesure de variabilité dans la maximisation de l'utilité. Le sous-chapitre 3.4 démontra cette équivalence.

L'équifinalité des deux approches permet de recourir alternativement aux concepts issus de l'une ou l'autre approche. Adaptant la théorie de l'erreur aléatoire aux problèmes d'aménagement, Neuberger [1971] deriva des mesures de bénéfice des citoyens pouvant servir à l'évaluation de plan d'aménagement. Ce type de mesure s'appelle «surplus collectif» ("group surplus") car il évoque une différence globale entre un état initial et un état final. La maximisation de ce surplus permet d'appréhender des problèmes de localisation dotés d'options hiérarchisées [Williams, 1977; Williams, Senior, 1978]. En outre, cette conceptualisation offre une justification théorique au modèle de Lowry.

Le modèle de Lowry est, comme on sait, basé sur les concepts d'interaction spatiale et d'accessibilité. Il conduit à une situation d'équilibre telle qu'activités résidentielle et commerciale sont ajustées à la distribution spatiale de l'emploi basique. On critiqua abondamment le mécanisme liant les activités ainsi la paire de sous-modèles d'interaction effectuant l'allocation zonale. Wilson [1974, chap.11] montra que ces modèles de localisation, pris séparément, sont dérivables par la méthode d'entropie maximale. Par ailleurs, ces dérivations peuvent se faire de façon équivalente selon un cadre théorique de l'utilité aléatoire (leur donnant une justification béhavioriste) ou selon une problématique de programmation mathématique (maximisation des bénéfices de localisation).

Le processus conjoint d'allocations résidentielle et commerciale a très peu de fondement formel. Williams et Senior [1978], en préconisant une théorie de choix probabiliste, présument un schéma conditionnel de localisation. En effet, le choix du lieu de consommation est conditionné par celui, plus important, du lieu de résidence lequel dépend de l'accessibilité du lieu d'emploi et des qualités de la zone domiciliaire. Cette structure conditionnelle est assurée par la détermination appropriée de mesure d'utilité consistante avec les processus de choix. Ici, on reprendra la démarche de Williams et Senior en privilégiant toutefois le formalisme de l'optimisation de l'utilité.

Le chapitre 3 concernant la modélisation du transport de même que ce sous-chapitre nous ont appris que, dans les programmes munis d'un terme d'entropie, les multiplicateurs de Lagrange associés à des contraintes de total se révélaient de véritables mesures d'utilité moyennant les attributs des alternatives qui composaient ce total. En outre, des mesures de coût composé et d'accessibilité à l'égard d'une activité sont évaluables. Dans un premier temps, on calculera un indice d'accessibilité aux facilités commerciales en recourant à un modèle spatial de consommation. On reprend ici la mesure de «bénéfices de la taille» de Wilson [1976], soit:

$$w_k = \frac{\alpha}{\psi^\beta} \ln W_k$$

Un problème de localisation maximisant un bénéfice de consommation au lieu k s'écrira comme suit:

$$\max_{\{S\}} \sum_k s_{1k} \left[(-\ln s_{1k} + 1) / \psi^\beta + w_k - c_{1k} \right] \quad \text{sujet à} \quad \sum_k s_{1k} = \bar{o} \cdot P_1 \quad (6.75-6.76)$$

La solution est donnée par:

$$\frac{-1}{\psi^\beta} \ln s_{1k} + w_k - c_{1k} + \frac{1}{\psi^\beta} \ln(\bar{o} P_1) - \bar{Q}_1^\beta = 0 \quad (6.77)$$

$$s_{1k} = \bar{o} P_1 \exp\{\psi^\beta (w_k - c_{1k} - \bar{Q}_1^\beta)\} \quad (6.78)$$

$$\bar{Q}_1^\beta = \frac{1}{\psi^\beta} \ln \sum_k \exp\{\psi^\beta (w_k - c_{1k})\} \quad (6.79)$$

La variable $\bar{Q}_1^\beta$ définit le bénéfice de localisation associé à la zone résidentielle à l'égard des facilités commerciales. En fait, elle désigne un attribut de la zone i (accessibilité commerciale) qui peut être intégré dans un modèle de choix résidentiel au même titre qu'un facteur d'attraction. Une telle insertion garantit la structure conditionnelle des proces-

sus. Voici d'abord une mesure d'utilité calculée d'après un facteur d'attraction résidentielle spécifique à la zone i .

$$q_i = \frac{1}{\psi^P} \ln Q_i$$

Le problème d'allocation résidentielle en conséquence s'écrit :

$$\max_{\{P\}} \sum_j \sum_i p_{ij} \left[(-\ln p_{ij} + 1) / \psi^P + q_i + \hat{U}_i^P - c_{ij} \right] \quad \text{sujet à} \quad \sum_i p_{ij} = E_j \quad (6.80-6.81)$$

Avec comme conditions d'optimalité :

$$\frac{-1}{\psi^P} \ln p_{ij} + q_i + \hat{U}_i^P - c_{ij} + \frac{1}{\psi^S} \ln E_j - \hat{U}_j^S = 0 \quad (6.82)$$

$$p_{ij} = E_j \exp \{ \psi^P (q_i + \hat{U}_i^P - c_{ij}) - \psi^S \} \quad (6.83)$$

$$\hat{U}_j^S = \frac{1}{\psi^S} \ln \sum_i \exp \{ \psi^P (q_i + \hat{U}_i^P - c_{ij}) \} \quad (6.84)$$

La mesure d'utilité composée $\hat{U}_j^S$ reste conforme au formalisme de notre problématique d'optimisation et évalue le bénéfice moyen de travailler dans une zone j en rapport avec la localisation générale de l'activité résidentielle.

La mesure globale de bénéfice du système spatial d'occupation du sol (i.e. le «surplus collectif») est dérivée du problème dual associé au choix résidentiel (celui-ci étant prioritaire au choix du lieu de consommation). On sait que le dual de (6.80-6.81) est :

$$\min_{\{v\}} \sum_j \sum_i p_{ij} / \psi^P - \sum_j E_j v_j \quad \text{où } v_j \text{ est rattaché à la contrainte (6.81)} \quad (6.85)$$

sujet à (6.82)

La variable duale v_j peut être établie en fonction du coût composé $\hat{U}_j^S$:

$$v_j = \frac{1}{\psi^S} \ln E_j - \hat{U}_j^S \quad (6.86)$$

On sait que la multiplication du premier terme de (6.86) dans (6.85) est constante car elle concerne l'entropie de la distribution d'emploi qui est fixe. Le premier terme de la fonction objective duale dépend implicitement de v_j mais, à l'optimalité, il devient constant. On établit alors un nouveau problème, la maximisation du surplus collectif :

$$\min_{\{v\}} \sum_j E_j \hat{U}_j^S \equiv \max_{\{P, S\}} \sum_j E_j \hat{U}_j^S \quad (6.87)$$

sujet à (6.84)

Cette dernière fonction objective apparaît comme un indice approprié d'évaluation d'un plan d'aménagement. En effet, la multiplication successive de l'emploi avec l'utilité composée $\bar{U}_j^P$ considère la distribution spatiale de l'emploi en rapport avec l'activité domiciliaire de même que la répartition de celle-ci en rapport avec le secteur commercial. La forme fonctionnelle de cet indice reste consistante avec la mesure de «surplus du consommateur» d'Hotelling [1938; voir Williams, 1976]. Ici, le programme de maximisation de surplus collectif devient soluble en développant le terme $\bar{U}_j^P$ à l'aide de la contrainte (6.84) autour de la fonction objective:

$$\begin{aligned} \max_{\{P, S\}} \sum_j E_j \sum_i \left[-\ln p_{1j} / \psi^P + q_i + \bar{U}_j^P - c_{1j} \right] \\ \equiv \sum_{1j} p_{1j} \left[-\ln p_{1j} / \psi^P + q_i - c_{1j} \right] + \sum_{1j} p_{1j} \bar{U}_j^P \end{aligned} \quad (6.88)$$

du fait que $\sum_j p_{1j} = P_1$ et que $\sum_k s_{1k} = \partial P_1$, alors:

$$\sum_{1j} p_{1j} \bar{U}_j^P \equiv \sum_{1k} s_{1k} \bar{U}_j^P / \partial$$

En employant la condition d'optimalité (6.82) pour la définition de l'utilité composée $\bar{U}_j^P$, la fonction objective devient par conséquent:

$$\begin{aligned} \max_{\{P, S\}} \sum_{1j} p_{1j} \left[-\ln p_{1j} / \psi^P + q_i - c_{1j} + \ln E_j / \psi^P \right] \\ + \sum_{1k} s_{1k} \left[(-\ln s_{1k} + 1) / \psi^S + w_k - c_{1k} + \ln(\partial P_1) / \psi^S \right] \end{aligned} \quad (6.89)$$

$$\text{sujet à} \quad \sum_1 p_{1j} = E_j \quad \sum_k s_{1k} = \partial P_1 \quad (6.90-6.91)$$

Ici, ces deux dernières contraintes sont associées aux multiplicateurs correspondant aux mesures composées $\bar{U}_j^P$ et $\bar{U}_j^S$. Cette dernière démonstration exposa l'équivalence de la maximisation du surplus collectif (calculé selon une séquence de choix conditionnels) avec la celle des bénéfices de localisation des processus séparés. Ainsi, les programmes exposés dans la section 6.3.2 qui traitaient indépendamment de chaque processus peuvent être mis à contribution dans la détermination d'un modèle-synthèse regroupant plusieurs activités plus ou moins liées ensemble.

6.3.4 Le modèle de Lowry en tant que modèle de bénéfice de localisation

Le processus conditionnel de localisation résidentielle et commerciale peut être formulé sous la forme d'un problème de maximisation des bénéfices. Le système d'équations non-linéaires et d'inégalités que symbolise le modèle de Lowry [1964] apparaît sans doute complexe car il ne possède aucune rationalité explicite. L'une des difficultés, mais aussi l'un des grands intérêts, réside dans l'interdépendance de l'emploi de service (ou commercial) et de la population. Néanmoins, il a été démontré, peu de temps après son apparition, que le modèle conduit à un état d'équilibre des distributions [Garin, 1966]. L'existence d'un état d'équilibre spatial signifie, selon l'idée originale, un ajustement de l'occupation du sol avec des critères d'accessibilité. Dans le cas où le modèle de Lowry émanerait d'un principe d'optimisation, cela indique que les différents bénéfices de localisation sont maximisés.

Nous reprenons ici la formulation complète du problème de maximisation de l'utilité (6.89) avec les contraintes marginales et les relations de consistance. Nous avons opté pour une fonction d'attraction générale de la variable W_j au lieu de la fonction puissance habituelle:

$$\begin{aligned} \max_{\{P, S\}} G = G^r + G^c = & \frac{-1}{\psi^r} \sum_{i,j} p_{i,j} \left[\ln \frac{p_{i,j}}{Q_i} - 1 \right] - \sum_{i,j} p_{i,j} c_{i,j} \\ & - \frac{1}{\psi^c} \sum_{i,j} s_{i,j} \left[\ln \frac{s_{i,j}}{g(W_j)} - 1 \right] - \sum_{i,j} s_{i,j} c_{i,j} \end{aligned} \quad (6.92)$$

$$\text{soit à} \quad \sum_i p_{i,j} = E_j \quad \sum_i s_{i,j} = \delta \cdot P_j \quad (6.93-6.94)$$

$$\text{et les relations} \quad P_i = \alpha \cdot \sum_j p_{i,j} \quad S_j = \sum_i s_{i,j} \quad (6.95-6.96)$$

$$E_j = E_j^r + S_j \quad (6.97)$$

Le seul problème G assujéti aux contraintes linéaires (6.93-6.94) est équivalent à la somme des problèmes G^r et G^c ; on obtiendrait comme solution un modèle de localisation résidentielle comme (6.15) et un modèle de localisation commerciale comme (6.13). Toutefois, les équations (6.95-6.97) définissant les relations entre variables introduisent des interdépendances internes qui empêchent la formation de modèles d'allocation distincts. Ainsi, on obtient la paire d'équations gravitaires suivante:

$$p_{1j} = \frac{E_j Q_1 \exp(-\psi^0 c_{1j})}{\sum_1 Q_1 \exp(-\psi^0 c_{1j})} \quad \text{où } E_j = E_j^0 + \sum_1 s_{1j} \quad (6.98)$$

$$s_{1j} = \frac{\partial P_1 g(W_j) \exp(-\psi^0 c_{1j})}{\sum_1 g(W_j) \exp(-\psi^0 c_{1j})} \quad \text{où } P_1 = \bar{a} \cdot \sum_1 p_{1j} \quad (6.99)$$

Ce système d'équations apparaît, de par le type de fonctions utilisées, non-linéaire mais il reste linéaire pour les séries $\{E_j\}$, $\{S_j\}$ et $\{P_j\}$. En effet, les fonctions non-linéaires ne concernent que les variables d'attraction et de coût. La solubilité du système d'équation n'est démontrée que par l'existence d'un état d'équilibre dans lequel les distributions résidentielles et commerciales sont concomitantes. Cet état d'équilibre se révèle la seule solution du problème 6. Une preuve de convergence de la paire d'équations spatiales est maintenant requise.

6.3.5 Etat d'équilibre de la structure urbaine

La preuve de convergence vers un état d'équilibre est donnée en convertissant le système (6.98-6.99) dans un format matriciel:

$$e = \{E_j\} \text{ N} \cdot 1 \quad e^0 = \{E_j^0\} \text{ N} \cdot 1 \quad s = \{S_j\} \text{ N} \cdot 1 \quad p = \{P_j\} \text{ N} \cdot 1$$

$$\Pi = \{\pi_{1j}\} \text{ N} \cdot \text{N}$$

$$\Sigma = \{\sigma_{1j}\} \text{ N} \cdot \text{N}$$

$$\pi_{1j} = \frac{\bar{a} \cdot Q_1 \exp(-\psi^0 c_{1j})}{\sum_1 Q_1 \exp(-\psi^0 c_{1j})} \quad (6.100) \quad \sigma_{1j} = \frac{\partial \cdot g(W_j) \exp(-\psi^0 c_{1j})}{\sum_1 g(W_j) \exp(-\psi^0 c_{1j})} \quad (6.101)$$

Les entrées des matrices Π et Σ sont invariables car elles sont définies par des variables exogènes. Les vecteurs et matrices peuvent représenter parcimonieusement le mécanisme d'allocation spatiale de Lowry:

$$p = \Pi e \equiv P_1 = \sum_1 \pi_{1j} E_j \quad (6.102)$$

$$s = \Sigma p \equiv S_1 = \sum_1 \sigma_{1j} P_j \quad (6.103)$$

$$e = e^0 + s \quad (6.104)$$

La preuve de convergence est donnée par la dérivation de multiplicateurs spatiaux associés à la distribution de l'emploi basique:

$$p = \Pi e = \Pi [e^0 + s] = \Pi e^0 + \Pi \Sigma p$$

$$p - \Pi \Sigma p = \Pi e^0$$

$$p = [I - \Pi \Sigma]^{-1} \Pi e^0 \quad (6.105)$$

La matrice $[I - \Pi\Sigma]^{-1}\Pi$ est un multiplicateur spatial convoluant le vecteur d'emploi basique en un vecteur de population. La convolution contient l'interaction spatiale directe de e^b vers p mais aussi l'interaction indirecte qui passe par la distribution de l'emploi commerciale s elle-même déterminée par le patron spatial résidentiel. De plus, l'effet du multiplicateur scalaire (6.2) y est inclus. Pour la distribution s , on a:

$$s = \Sigma p = \Sigma [I - \Pi\Sigma]^{-1} \Pi e^b \quad (6.106)$$

Cette fois, la matrice $\Sigma [I - \Pi\Sigma]^{-1} \Pi$ mesure l'effet spatial total entre l'emploi basique et l'emploi commercial via la distribution résidentielle. On peut aussi démontrer que:

$$s = \Sigma p = \Sigma \Pi e^b + \Sigma \Pi s$$

$$s - \Sigma \Pi s = \Sigma \Pi e^b$$

$$s = [I - \Sigma \Pi]^{-1} \Sigma \Pi e^b \quad (6.107)$$

La matrice $[I - \Sigma \Pi]^{-1} \Sigma \Pi$ est une forme alternative mais équivalente qui met en relief la conversion directe du vecteur e^b en un vecteur s initial lequel sera convolué par la matrice inversée. La différence entre les matrices $[I - \Pi\Sigma]^{-1}$ et $[I - \Sigma \Pi]^{-1}$ réside le vecteur de droite qui est multiplié: le premier concerne un vecteur-population et le second, un vecteur-emploi. On peut également représenter (6.105) par:

$$p = \Pi [I - \Sigma \Pi]^{-1} e^b \quad (6.108)$$

Sheppard [1979a; 1979b] exposa les propriétés intéressantes des multiplicateurs spatiaux. En fait les multiplicateurs spatiaux s'avèrent des versions géographiques des multiplicateurs économiques typiques de l'«analyse input-output» liant les groupes industriels [voir Leontief, Strout, 1963]. Les éléments d'un multiplicateur économique matriciel calculent l'effet total qu'a un secteur industriel sur un autre. Outre les échanges directs, on considère l'infinité des apports indirects (i.e. passant par d'autres secteurs industriels) comme l'amplification totale générée par un secteur sur un autre.

Dans le cas du modèle de Lowry, on sait que l'emploi basique dans une zone j affecte l'activité résidentielle dans une zone i par un processus d'interaction (contenu dans un élément π_{ij}). Par ailleurs, cette activité

résidentielle sollicite des services à proximité et, probablement, dans la zone d'emploi j . L'emploi de service ainsi requis dans j affectera de la même manière la population (via π_{ij}) et ainsi de suite. Théoriquement, les effets sont infinis du fait que chaque zone influence l'ensemble du système dans lequel chaque zone, via l'autre processus, agit sur l'ensemble.

Sheppard [1979a] montra que les multiplicateurs spatiaux peuvent servir à mesurer l'«influence» ou le «potentiel géographique» selon un processus d'interaction. On remarquera que les multiplicateurs matriciels comprennent usuellement une inversion d'une matrice laquelle est le solde d'une matrice d'identité (I) avec une autre matrice. Sous une forme scalaire, un multiplicateur à la forme suivante:

$$\frac{a}{1-h} \quad |h| < 1 \quad (6.109a)$$

Cette fonction correspond à la somme d'une progression géométrique infinie:

$$\frac{a}{1-h} = a \cdot (1 + h + h^2 + h^3 + \dots) = a \cdot \sum_{r=0}^{\infty} h^r \quad (6.109b)$$

Le multiplicateur de base économique présenté dans la section 6.2.1 est du même type que (6.109). Pour un multiplicateur matriciel, l'analogie tient. En effet, considérons une matrice V avec éléments v_{ij} traduisant l'interaction $j \rightarrow i$ telle que $\sum_i v_{ij} \leq 1$; $\forall j$. Les effets indirects $j \rightarrow i$ du deuxième degré égalent donc la somme des interactions:

$$v_{ij}^{(2)} = \sum_k v_{ik} v_{kj}$$

Tandis que les effets indirects du troisième degré:

$$v_{ij}^{(3)} = \sum_k \sum_c v_{ik} v_{kc} v_{cj}$$

Et ainsi de suite. Ces effets de deuxième et troisième degré sont les éléments des matrices élevées aux puissances correspondantes [Tinkler, 1974]:

$$V^2 = VV = \{v_{ij}^{(2)}\} \text{ N} \cdot \text{N} \quad V^3 = V^2V = \{v_{ij}^{(3)}\} \text{ N} \cdot \text{N}$$

Alors si on veut composer une matrice de l'influence totale, on sommerait:

$$M = \sum_{r=0}^{\infty} V^r = I + V + V^2 + V^3 + \dots \quad (6.110)$$

Une condition nécessaire et suffisante pour que M soit finie est que tous les arguments réels des valeurs propres (ou racines caractéristiques) de M soient inférieurs à un en valeur absolue [Sheppard, 1979a]. Inversement,

il faut qu'au moins une somme de rangée (ou de colonnes) soit inférieure à un sans qu'aucune puisse dépasser l'unité. Dans un tel cas:

$$M = \sum_{r=0}^{\infty} V^r = [I-V]^{-1} \quad (6.111)$$

Ces restrictions sur la matrice d'interaction V permettent de faire converger la série géométrique (6.111). Toutefois, V ne peut être une matrice stochastique de probabilité de transition car, dans ce cas, les sommes de toute rangée (ou de toute colonne) égaleraient systématiquement l'unité: La matrice $[I-V]$ deviendrait singulière donc non-invertible. Sheppard [1979a] note néanmoins qu'une matrice d'interaction augmentée d'une rangée et d'une colonne pour incorporer une région externe (passant ainsi de système fermé à système ouvert) pourra se prévaloir du standard de matrice stochastique. L'analyse de l'influence portera alors sur la sous-matrice incluse dans la matrice augmentée i.e. sur la région initiale. Si le système est ouvert ou, entre d'autres termes, la colonne et la rangée supplémentaire ne contiennent pas que des zéros, alors la sous-matrice satisfait les conditions d'existence de la matrice d'influence M . Cela réfère aux processus markoviens transients où l'on constate que seuls les systèmes ouverts peuvent converger vers un état d'équilibre fini.

Maintenant, revenons au modèle de Lowry. Il faut s'assurer que les multiplicateurs spatiaux sont finis et, en particulier examiner les propriétés d'invertibilité des matrices. Dans les cas exposés, l'inversion concerne une matrice $[I-\Pi\Sigma]$ ou $[I-\Sigma\Pi]$. D'après la définition des éléments π_{1j} et σ_{1j} , on sait que:

$$\sum_i \pi_{1j} = \bar{a} \quad \forall j \quad \sum_i \sigma_{1j} = \bar{o} \quad \forall j$$

Le produit matriciel $\Pi\Sigma$ ou $\Sigma\Pi$ permet d'obtenir des éléments x_{1j} ou y_{1j} dont les sommes de rangée sont nécessairement constantes:

$$\sum_i x_{1j} = \sum_i y_{1j} = \bar{a}\bar{o} \quad \forall j \quad \text{où } \Pi\Sigma = \{x_{1j}\} \text{ } N \cdot N \text{ et } \Sigma\Pi = \{y_{1j}\} \text{ } N \cdot N$$

Le produit des coefficients scalaires $\bar{a}$ et $\bar{o}$ doit être inférieur à 1 si l'on veut que les matrices soient invertibles. Le système considéré étant ouvert car incluant un vecteur exogène e^P , ce produit ne peut dépasser l'unité. En effet, en reprenant les définitions de $\bar{a}$ et $\bar{o}$ selon les totaux d'emploi total E , de service S et la population totale P , on a:

$$S = \bar{o} \cdot P \quad \text{et} \quad P = \bar{a} \cdot E$$

$$S = \alpha \cdot (\bar{\alpha} \cdot E)$$

Alors si $\alpha \bar{\alpha} > 1$, $S > E$, ce qui est impossible puisque S est compris dans E ;
 Si $\alpha \bar{\alpha} = 1$, $S = E$, ce qui implique que $E - S = E^* = 0$, donc il n'y a plus de variable exogène et le système est fermé;
 Donc $\alpha \bar{\alpha} < 1$, $S < E$, garantis selon les définitions de l'emploi total, l'emploi basique et de l'emploi commercial.

Maintenant, complexifions le processus de localisation. Le mécanisme d'allocation spatiale des services qui a été prôné jusqu'ici ne considère que la distribution résidentielle comme seul et unique marché. Dans le modèle original de Lowry, le sous-modèle de localisation commerciale possède deux termes additifs distincts; l'un comprend une fonction d'accessibilité de la population tandis que l'autre a comme argument l'emploi total de la zone. Ce dernier terme implique que les travailleurs, à partir de leurs lieux d'emploi, peuvent aussi être consommateurs. Lowry [1964] ne spécifia pas d'interaction spatiale dans ce cas car il jugea, qu'étant donnée la faible mobilité des déplacements à motif de consommation, le marché se limitait à la zone. Cette hypothèse paraît restrictive car la mobilité varie d'un système à l'autre et que la taille moyenne des zones affecte la proportion d'interactions inter-zonales. Wilson [1974] ne vit pas d'obstacle à ce qu'on mette à contribution les zones d'emploi voisines en convertissant le terme aspatial de (6.7) en un sous-modèle d'interaction. Une fonction de transport très sévère exprimera la mobilité résulte dont fait preuve ce type d'interaction.

Nous justifierons l'insertion d'un nouveau type d'interaction dans le processus d'allocation en ajoutant à la fonction objective (6.92) un terme s'identifiant aux bénéfices de consommation encourus à partir du lieu d'emploi. Soit:

$$\max_{\{P, S, F\}} G' = G^P + G^S - \frac{1}{\psi^P} \sum_j \sum_k f_{jk} \left[\ln \frac{f_{jk}}{h(W_j)} - 1 \right] - \sum_j \sum_k f_{jk} C_{jk} \quad (6.112)$$

$$\text{sujet à} \quad \sum_j f_{jk} = iE_k \quad \text{et (6.93-6.97)} \quad (6.113)$$

où f_{jk} : interactions à motif de consommation du lieu d'emploi k vers lieu de service j ; $h(\cdot)$: fonction d'attraction de j ; i : coefficient convertissant les interactions en emploi commercial.

L'emploi commercial est dorénavant calculé selon l'équation d'allocation suivante:

$$S_j = \bar{o} \cdot \sum_k S_{kj} + \bar{i} \cdot \sum_k f_{jk} \quad (6.114)$$

La solution du programme ci-haut de maximisation pour $\{f_{jk}\}$ est:

$$f_{jk} = \frac{\bar{i} E_k h(W_j) \exp(-\psi^r c_{jk})}{\sum_k h(W_j) \exp(-\psi^r c_{jk})} \quad (6.115)$$

Pour être réaliste, la paramètre de transport ψ^r devrait être supérieur à ψ^s car il désigne une mobilité plus restreinte. On présume également que la fonction d'attraction $h(W_j)$ diffère de $g(W_j)$ parce que la consommation à partir d'un lieu de travail se distingue de celle partant du lieu de résidence. Enfin, les coefficients $\bar{o}$ et $\bar{i}$ sont tels qu'ils traduisent la part moyenne attribuable aux consommateurs-résidents et aux consommateurs-travailleurs, soit:

$$S = \bar{o} \cdot P + \bar{i} \cdot E$$

Puisque E comprend S , $\bar{o}$ et $\bar{i}$ sont restreints de la façon suivante:

$$S = \bar{o} \cdot (\bar{a}E^s + \bar{a}S) + \bar{i} \cdot E^s + \bar{i} \cdot S \equiv (\bar{a}\bar{o} + \bar{i})E^s + (\bar{a}\bar{o} + \bar{i})S$$

Alors, $0 < \bar{a}\bar{o} + \bar{i} < 1$ ou bien

$$0 < \bar{i} < 1 - \bar{a}\bar{o} \quad \text{et} \quad 0 < \bar{o} < (1 - \bar{i})/\bar{a}$$

On convient que l'incorporation d'un nouveau processus d'interaction introduit encore une fois une dépendance non-linéaire chez l'emploi de service. Dans notre formulation initiale, la distribution commerciale $\{S_j\}$ dépendait uniquement de celle des résidences $\{P_i\}$ laquelle était conditionnée par le patron spatial des emplois en partie constitué du secteur commercial. Cette boucle rétroactive (sans référence temporelle toutefois) semble désormais assistée par une autre boucle manifestée par la contribution de l'emploi total dans la formation du secteur des commerces. La complexité additionnelle ne compromettra pas l'existence d'une situation d'équilibre car la dérivation d'un multiplicateur spatial transformant un vecteur d'emploi basique en celui d'emploi non-basique ou de population reste possible.

Tout comme la démonstration précédente, il suffit de définir une matrice d'interaction mesurant l'effet inter-zonal de façon standard. Alors:

$$\tilde{\varphi} = \{\varphi_{ij}\} N \cdot N \quad \text{où} \quad \varphi_{ij} = \frac{\bar{i} \cdot h(W_i) \exp(-\psi^r c_{ij})}{\sum_j h(W_i) \exp(-\psi^r c_{ij})} \quad (6.116)$$

Le système est défini par les relations inter-activités suivantes:

$$(6.102), (6.104) \text{ et } s = \Sigma p + \tilde{x}e \quad (6.117)$$

Il faut développer à partir de l'équation relative à s car les nouveaux effets interactifs se situent à ce niveau. Commencer par l'équation de p résulterait en un développement sans fin. Ainsi:

$$\begin{aligned} s &= \Sigma p + \tilde{x}e = \Sigma \Pi e^s + \Sigma \Pi s + \tilde{x}e^s + \tilde{x}s \\ s - \Sigma \Pi s - \tilde{x}s &= \Sigma \Pi e^s + \tilde{x}e^s \\ s &= [I - (\Sigma \Pi + \tilde{x})]^{-1} (\Sigma \Pi + \tilde{x}) e^s \end{aligned} \quad (6.118)$$

On notera l'isomorphisme entre le multiplicateur (6.107) et ce dernier. En effet, la matrice composée $\Sigma \Pi$ est substituée par $(\Sigma \Pi + \tilde{x})$, signifiant ainsi que la distribution spatiale du secteur commercial est établie, d'une part, par le patron résidentiel lui-même conditionné par les services et, d'autre part, par la consommation des travailleurs dont une partie provient du secteur commercial. Le multiplicateur associé à p est déduit en reprenant s défini par l'équation (6.118):

$$\begin{aligned} p &= \Pi e = \Pi e^s + \Pi s = \Pi e^s + \Pi [I - (\Sigma \Pi + \tilde{x})]^{-1} (\Sigma \Pi + \tilde{x}) e^s \\ p &= \Pi \left[I + [I - (\Sigma \Pi + \tilde{x})]^{-1} (\Sigma \Pi + \tilde{x}) \right] e^s \end{aligned} \quad (6.119)$$

En pré-multipliant par $[I - (\Sigma \Pi + \tilde{x})]$ le terme entre les grands crochets, on le simplifie considérablement:

$$\left[[I - (\Sigma \Pi + \tilde{x})] + (\Sigma \Pi + \tilde{x}) \right] = I$$

Maintenant, pré-multiplions par $[I - (\Sigma \Pi + \tilde{x})]^{-1}$ pour revenir à la forme initiale. Alors, (6.119) devient:

$$p = \Pi [I - (\Sigma \Pi + \tilde{x})]^{-1} e^s \quad (6.120)$$

Encore une fois, on remarquera la ressemblance entre cette équation et l'équation (6.108). Pour s'assurer que les multiplicateurs spatiaux (6.118) et (6.120) sont mathématiquement finis (ou de l'invertibilité de $[I - (\Sigma \Pi + \tilde{x})]$), il faut que toutes les sommes des colonnes de $(\Sigma \Pi + \tilde{x})$ soient, en termes absolus, inférieures à l'unité. On a déjà prouvé que les sommes des colonnes de $\Sigma \Pi$ sont constantes et égales à $\bar{\alpha} \bar{o}$. D'après la définition des éléments $\varphi_{1,j}$, toute somme égale constamment à:

$$\sum_i \varphi_{1,j} = \bar{i} \quad \forall j$$

De ce fait, toute somme de colonne de $(\Sigma \Pi + \bar{Q})$ égale $(\bar{a} \bar{o} + i)$. On sait par ailleurs que pour être vraisemblable, le scalaire $(\bar{a} \bar{o} + i)$ est nécessairement positif et inférieur à un.

Dans ce sous-chapitre, on a démontré un certain nombre de possibilités d'encadrement des modèles de localisation d'activités dans un formalisme sollicitant la programmation mathématique. On a incidemment préconiser des formulations de maximisation de «bénéfices de localisation» car elles apparaissent compatibles avec de nombreux cadres théoriques. Cette approche dérive explicitement de la maximisation de l'utilité déterministe et se concilie très bien avec celles de l'entropie, de l'utilité aléatoire, de l'accessibilité et du surplus collectif. En surcroît, les problèmes de type «normatif» peuvent se considérer comme des cas particuliers dans lesquels aucune variabilité n'est inscrite.

A titre d'exemple, on a d'abord présente des formulations intéressées à l'allocation spatiale d'une seule activité, résidentielle ou commerciale. La solution de ces problèmes correspondait habituellement à un modèle d'interaction de type gravitaire. En adoptant un schème conditionnel, on a reconnu qu'un problème de maximisation des bénéfices à la capacité de traiter deux allocations distinctes mais imbriquées. Les conditions d'optimalité commandent la résolution d'un système d'équations linéaires qui mène à un état d'équilibre des distributions spatiales. L'existence d'un tel état a été démontré précédemment selon une notation matricielle. Mathématiquement, le système spatial mais intemporel est dit convergent même si ce qualificatif semble souvent réservé aux analyses dynamiques. Le sous-chapitre suivant exploitera cette propriété de convergence au profit d'une analyse spatio-temporelle. Le format matriciel sera davantage employé car il est conforme aux représentations usuelles de système qui ont été illustrées dans le sous-chapitre 5.3.

6.4 CONVERSION SPATIO-DYNAMIQUE DU MODELE DE LOWRY

6.4.1 Tentatives d'incorporation de la dimension du temps

Les modèles de localisation urbaine de type «descriptif» visent avant tout à établir une distribution probable d'une ou de plusieurs activités dans l'espace urbain. Les principes évoqués dans le sous-chapitre précédent relèvent tous d'une maximisation d'une forme de bénéfices de localisation (utilités maximales, coûts de déplacement minimaux). En fait, les principes d'allocation d'activité tiennent généralement compte d'une répartition spatiale d'une ou des activités «fixes». L'activité à allouer s'ajuste donc en fonction des interactions avec les autres activités en profitant des meilleurs sites et en réduisant les coûts de transport. Toutefois, les modèles comptent une composante de dispersion afin de représenter la variabilité des comportements de localisation.

En réalité, la solution d'un problème mathématique maximisant les bénéfices de localisation fournit les conditions d'«optimalité». Il s'agit d'un état d'équilibre où la distribution spatiale de l'activité en question s'accorde le mieux avec l'environnement (considérant le niveau de variabilité donné). Par conséquent, ces modèles s'avèrent de piètres modèles de «développement urbain» car la localisation des activités exogènes semble constante et que le processus d'allocation est immédiat.

La grande majorité des applications temporelles des modèles de localisation a un caractère "pseudo-dynamique". Les formulations les plus élaborées permettent de définir une structure urbaine d'équilibre déterminée par les conditions actuelles de l'environnement (réseau de transport, activités exogènes, facteurs d'attraction). Dans un contexte prédictif, ces modèles, calibrés pour le présent, sont alimentés par des données hypothétiques futures; la méthode se révèle «statique-comparative» car elle ne fait que remplacer les conditions externes. Ce cadre, apparemment le plus utilisé en planification, paraît valide si les mécanismes de développement urbain restent constants durant la période étudiée et si les distributions d'activité reagissent promptement aux modifications de l'environnement. Les auteurs du modèle économétrique de localisation résidentielle «NBER» [Ingram et al., 1972] soulignèrent avec humour la déficience du cadre

statique-comparatif:

«... in existing theories of location it is assumed that either cities are destroyed every night and rebuilt the next morning or that households live in house trailers that are relocated daily.» [p.8]

La complexité de l'ajustement réel entre les distributions urbaines d'activités ainsi que les incessantes perturbations de l'environnement font que l'état le plus commun du système urbain en est un de déséquilibre [Harris, 1970]. L'inscription d'éléments dynamiques dans les modèles de localisation semble la meilleure alternative quant à la considération des états de déséquilibre. Le principe d'optimisation des bénéfices est certes conservée mais il devient un état de convergence qu'on atteindra avec certains retards. La permanence d'une structure urbaine, les délais de réaction dans le processus de relocalisation et la persistance d'un effet pourtant éphémère peuvent être modélisés, comme on sait (voir chapitre 5), à l'aide d'une structure de décalage.

Plusieurs furent tentés d'insérer une dimension temporelle au modèle de Lowry et à ses variantes. Czamanski [1965] inscrivit une simple structure de décalage aux équations contraignant les variables zonales à leurs totaux respectifs, ralentissant ainsi le réajustement de la population totale et de l'emploi total de service à l'égard des variations temporelles de l'emploi basique.

Le fameux modèle «EMPIRIC» [Hill, 1965; Irwin, Brand, 1965] fait usage de l'opérateur de différence ∇ de sorte que le principe d'allocation est incrémentiel dans le temps. Chacune des variables endogènes ou «variables localisées» sont des fonctions linéaires des autres variables localisées (calculées dans la zone) et de plusieurs «variables localisantes» (sous formes incrémentielles ou non). On y remarque l'absence d'interaction spatiale, les relations linéaires ne prévoyant pas la contribution des lieux voisins. L'usage des différences n'instaurent pas nécessairement des qualités dynamiques du fait que les incréments sont des fonctions simultanées des incréments des variables externes. De plus, la structure fonctionnelle apparaît purement "empirique" car elle ne dérive pas d'un principe théorique formel.

Crecine [1968] ajouta des éléments dynamiques au modèle de Lowry; le modèle «TOMM» ("Time Oriented Metropolitan Model") préserve les propriétés de délai chez les totaux et réajuste à chaque période des contraintes d'utilisation du sol. Ainsi, la croissance des activités dans une zone tend à réduire progressivement le rythme d'allocation de chacune d'elles de par l'indisponibilité croissante du sol. Implicitement, ce phénomène de capacité d'occupation du sol s'assimile à un processus autorégressif temporel négatif (bien que l'auteur ne fit point ce rapprochement).

Batty [1976 chap.11; 1978] préconisa une version «différenciée» du modèle de Lowry dotée de sous-modèles de localisation (résidences et services) qui portaient uniquement sur une part "relocalisable" de l'activité). Autrement dit, seules les fractions mobiles (qu'on appela "mover pool") et additionnelles des activités sont réallouées en fonction des principes d'interaction. Implicitement, il s'agit d'un schéma autorégressif puisqu'il y a permanence des structures urbaines antérieures.

Néanmoins, ces extensions temporelles du modèle de Lowry n'exploitent pas les possibilités d'une véritable modélisation spatio-dynamique; identification des structures de décalage et des structures spatiales d'interaction, estimation paramétrique. De plus, leurs auteurs ne conjuguèrent pas le formalisme des modèles de localisation spatiale avec celui entourant la dynamique de localisation. Ces deux facettes d'un même phénomène seront conciliées dans la section suivante.

6.4.2 Dynamique d'un système gradient

Un programme de maximisation des bénéfices de localisations résidentielle et commerciale a été formulé dans la section 6.3.4 (6.92-6.97). Ses conditions d'optimalité (ou sa solution) équivalent à une paire de modèles d'interaction isomorphe au modèle synthèse proposé par Lowry [1964]. Cette paire constitue un système d'équations avec deux distributions inconnues (population $\{P_i\}$ et emploi commercial $\{S_i\}$) et une connue (emploi basique $\{EP_i\}$). Une application itérative des deux équations avec comme distribution initiale de l'emploi totale, celle de l'emploi basique, convergera vers un état d'équilibre stable où chacune des activités est définitivement constante. Cette propriété de convergence a été démontrée dans la section

6.3.5 où des multiplicateurs spatiaux de l'emploi basique y ont été dérivés. L'état d'équilibre reste la solution globale du problème G ($=G^P+G^B$) (6.92-6.97).

Considérons une application où la distribution de l'emploi de service est fixée exogènement. La partie G^B devient alors sans pertinence car constante. Le programme équivaudra à un problème G^P dont la solution nous fournit le modèle de localisation résidentielle. La série spatiale $\{P_i\}$ optimale est donc celle découlant de la somme des interactions originant des lieux d'emploi (6.98). D'un point de vue dynamique, cette série s'avère un état d'équilibre univarié. Dans le langage des systèmes spatio-temporels (voir ss-chap. 5.3), les emplois commercial et basique sont des variables d'"input" qui sont, dans ce cas-ci, tenues constants.

La modélisation proposée ici veut exploiter les propriétés de convergences des distributions spatiales endogènes de manière à les assimiler à des dynamiques. Dans nombre de systèmes non-géographiques, la dynamique des composantes est orientée vers un état d'équilibre stable. Par exemple, en absence d'énergie externe, les écarts thermiques d'un gaz dans un volume donné évoluent régulièrement vers l'état d'«entropie maximale» i.e. vers une température uniforme (équilibre thermodynamique). Dans un champs électromagnétique, un corps attiré empruntera un chemin maximisant la différence de potentiel au point initial; le vecteur en question correspond au gradient du point initial. On appelle ces systèmes des «systèmes gradients» car leurs dynamiques sont guidées par un gradient d'une fonction scalaire identifiée à un quelconque état équilibré et stable. Le concept reste parfaitement envisageable pour un système socio-économique dynamique. Isard et Liossatos [1979 chap.11] notent que les systèmes gradients ne sont pas nécessairement orientés vers un état unique mais qu'en tout point, ils sont sous l'influence d'un seul «attracteur».

L'optimum global de G^P+G^B représente l'équilibre bivarié du système; l'optimum seul de G^P est un état d'équilibre de $\{P_i\}$ quand l'emploi commercial est tenu constant tandis que l'optimum seul de G^B est un état d'équilibre de $\{S_i\}$ si l'activité résidentielle est constante. Si notre système est gradient, les distributions $\{P_i\}$ et $\{S_i\}$ évolueront vers l'équilibre global mais, prises séparément, elles sembleront converger vers

leurs propres équilibres. Mathématiquement, la dynamique se présente comme une équation différentielle du temps (continu) proportionnelle au gradient de la fonction d'optimisation:

$$\frac{dP_1}{dt} \propto \frac{\partial G}{\partial P_1} = \frac{\partial G^P}{\partial P_1} \quad \frac{dS_1}{dt} \propto \frac{\partial G}{\partial S_1} = \frac{\partial G^S}{\partial S_1} \quad (6.121-6.122)$$

où $G = G^P + G^S$ (voir problème 6.92-6.97)

Par ailleurs, on sait qu'à l'optimalité:

$$\partial G^P / \partial P_1 = 0 \quad \text{où alors} \quad \underline{P}_1 = \sum_j \pi_{1j} E_j \quad (6.123)$$

$$\partial G^S / \partial S_1 = 0 \quad \text{où alors} \quad \underline{S}_1 = \sum_j \sigma_{1j} P_j \quad (6.124)$$

Les termes π_{1j} et σ_{1j} sont définis à (6.101) et (6.102). Ainsi dit, $dP_1/dt = 0$ si $P_1(t) = \underline{P}_1$ et $dS_1/dt = 0$ si $S_1(t) = \underline{S}_1$. Rendues à leurs valeurs d'équilibre, les distributions $\{P_1\}$ et $\{S_1\}$ deviennent immuables. Une dynamique conséquente, assurant l'immuabilité à l'équilibre, serait:

$$dP_1/dt = v^P(\underline{P}_1 - P_1) \quad (6.125)$$

$$dS_1/dt = v^S(\underline{S}_1 - S_1) \quad (6.126)$$

Pour être cohérent, les paramètres v^P et v^S sont positifs mais inférieurs à un sinon des oscillations se produisent. Ici, les conditions d'équilibre dynamique ($dP_1/dt = 0$; $dS_1/dt = 0$) correspondent aux modèles-solutions du programme G. Afin d'examiner la forme de trajectoire dans le temps que supposent ces équations différentielles, solutionnons l'une d'elles en tenant une distribution d'activité constante. Soit:

$$\frac{dP_1}{dt} = v^P(\underline{P}_1 - P_1)$$

Multiplions par le différentiel dt:

$$dP_1 = v^P(\underline{P}_1 - P_1)dt \quad \text{ou} \quad \frac{dP_1}{v^P(\underline{P}_1 - P_1)} = dt$$

L'intégration permet de trouver:

$$\int \frac{dP_1}{v^P(\underline{P}_1 - P_1)} = \int dt = t$$

$$-\ln(\underline{P}_1 - P_1)/v^P + C_1 = t$$

La constante d'intégration C_1 est relative aux conditions initiales, disons $P_1(0)$ où $t=0$. Par conséquent:

$$C_1 = \ln(\underline{P}_1 - P_1(0))/u^P$$

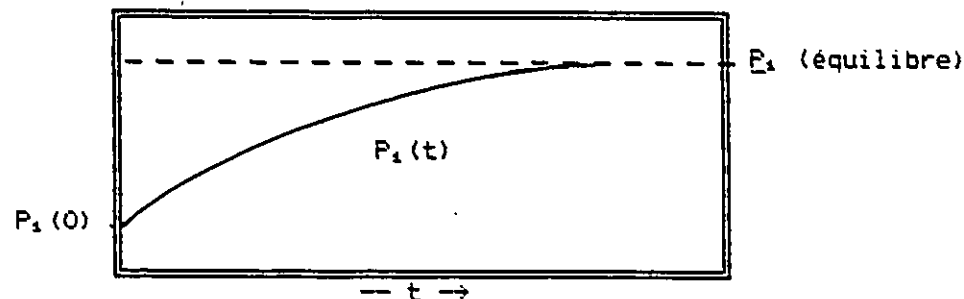
Soit une solution comme fonction du temps:

$$\ln(\underline{P}_1 - P_1) = -u^P t + u^P C_1$$

$$P_1(t) = \underline{P}_1 - \exp(-u^P t + u^P C_1) = \underline{P}_1 - (\underline{P}_1 - P_1(0))\exp(-u^P t) \quad (6.127)$$

La trajectoire dynamique de $P_1(t)$ est une exponentielle-négative convergeant vers une valeur stable $\underline{P}_1$. Le terme constant $(\underline{P}_1 - P_1(0))$ mesurant l'écart entre les conditions initiales et finales détermine le rythme de l'évolution. Le graphique 6.1 illustre la forme trajectorielle de $P_1(t)$ lorsque les distributions $\{S_i\}$ et $\{E_i^P\}$ sont maintenues dans un état soutenu:

Figure 6.1 Trajectoire d'un système dynamique convergent



En réalité, il y a N trajectoires $P_i(t)$ et N trajectoires $S_i(t)$ qui sont toutes interdépendantes. Un «espace-état» à $2N$ dimensions serait nécessaire pour représenter la co-évolution du système spatial [voir Dendrinos, Mullaly, 1985]. Deux variables ayant un comportement convergent du type exponentiel-négatif tracerait une ligne projectile (non-spirale) dirigée vers un noyau «attracteur» (co-équilibre). Dans la section 6.5.3, une approximation d'une trajectoire à $2N$ dimensions sera faite; la forme "projectile" est une forme possible mais pas unique.

Dans une application pratique, les équations différentielles restent de peu d'utilité, étant donné que les observations temporelles sont discrètes. Elles peuvent être substituées par des équations de différences du premier ordre. Alors (6.125) et (6.126) devient respectivement:

$$\nabla P_{1t} = u^P (\underline{P}_1 - P_{1t-1}) \quad \text{où} \quad \nabla P_{1t} = P_{1t} - P_{1t-1} \quad (6.128)$$

$$\nabla S_{1t} = u^S (\underline{S}_1 - S_{1t-1}) \quad \text{où} \quad \nabla S_{1t} = S_{1t} - S_{1t-1} \quad (6.129)$$

En éliminant l'opérateur de différence ∇ , on retrouve:

$$P_{1t} - P_{1t-1} = v^P (P_{1t} - P_{1t-1})$$

$$S_{1t} - S_{1t-1} = v^S (S_{1t} - S_{1t-1})$$

$$P_{1t} = v^P P_{1t} + (1-v^P) P_{1t-1} = v^P \sum_j \pi_{1j} E_j + (1-v^P) P_{1t-1} \quad (6.130)$$

$$S_{1t} = v^S S_{1t} + (1-v^S) S_{1t-1} = v^S \sum_j \sigma_{1j} P_j + (1-v^S) S_{1t-1} \quad (6.131)$$

On remarque facilement l'isomorphie des deux dernières équations avec les systèmes spatio-temporels décrit dans le sous-chapitre 5.3. En effet, leurs derniers termes sont autorégressifs dans le temps car, par définition, $0 < v^P, v^S < 1$, donc $0 < (1-v^P), (1-v^S) < 1$. De plus, leurs premiers termes contiennent des matrices de structures spatiales ($\Pi = \{\pi_{1j}\}$ et $\Sigma = \{\sigma_{1j}\}$) déjà standardisées qui capturent les interactions provenant de l'autre distribution. Partant de cette analogie fondamentale, on peut faire la généralisation suivante dans laquelle sont incorporées des autorégressions spatiales, des structures de décalage alternatives et des valeurs paramétriques appropriées:

$$P_{1t} = \Omega_I(z) P_{1t} + \Omega_W(z) \sum_j W_{1j} P_{jt} + A_I(z) \{S_{1t} + E_I^P\} + A_{\Pi} \sum_j \pi_{1j} \{S_{jt} + E_j^S\} \quad (6.132a)$$

$$p_t = [\Omega_I(z) I + \Omega_W(z) W] p_t + [A_I(z) I + A_{\Pi}(z) \Pi] \{s_t + e_t^P\} \quad (6.132b)$$

de même que:

$$S_{1t} = \Gamma_I(z) S_{1t} + \Gamma_Q(z) \sum_j Q_{1j} S_{jt} + B_I(z) P_{1t} + B_{\Sigma}(z) \sum_j \sigma_{1j} P_{jt} \quad (6.133a)$$

$$s_t = [\Gamma_I(z) I + \Gamma_Q(z) Q] s_t + [B_I(z) I + B_{\Sigma}(z) \Sigma] p_t \quad (6.133b)$$

Les matrices W et Q sont standardisées et peuvent représenter soit la compétition des zones voisines, soit les forces agglomératives (dépendant des signes des polynômes $\Omega_W(z)$ et $\Gamma_Q(z)$). Les termes régressifs non-spatiaux comme $A_I(z) \{s_t + e_t^P\}$ et $B_I(z) p_t$ traduisent quant à eux les relations internes entre les activités (compétition pour le sol, effets d'entraînement locaux). Enfin, les termes typiquement spatiaux comme $A_{\Pi}(z) \Pi \{s_t + e_t^P\}$ et $B_{\Sigma}(z) \Sigma p_t$ évoquent le véritable processus de localisation émanant du principe d'optimalité des bénéfices. Contrairement aux formulations statiques, ce processus n'est plus seul mais couplé avec des phénomènes de

décalage, d'autocorrélation spatiale, de permanence des distributions antérieures, de relations fonctionnelles intra-zonales. Cette modélisation s'avère donc plus complète.

En dépit des processus autoregressifs spatiaux et temporelles qui se greffent désormais aux strictes processus de localisation, il n'existe qu'un état d'équilibre atteignable dans des conditions d'état soutenu du vecteur exogène e_t . L'inertie causée par l'autocorrélation temporelle et les décalages distribués ainsi que l'autocorrélation spatiale ne font que ralentir l'évolution des distributions vers l'état de bénéfices de localisation maximales. Sans aucun doute, l'état d'équilibre reste inatteignable, du moins longtemps, car il suppose un environnement en état soutenu; distributions constantes de l'emploi basique, du facteur d'attraction résidentielle (compris dans le terme π_{1j}) et du facteur d'attraction commerciale (compris dans le terme σ_{1j}) de même qu'un réseau de transport inchangé (répercussion sur c_{1j} inclus dans π_{1j} et σ_{1j}). Dans l'appendice 6.A, on définit les conditions d'équilibre statique du modèle (6.132-6.133).

Dans la section 6.4.5, il a été question d'un problème de localisation dans laquelle le secteur commercial était conditionné par la distribution de l'emploi. Selon cette conceptualisation, les individus, à leurs lieux de travail, se révèlent des consommateurs potentiels tout comme ceux à leurs lieux de résidence. Une mobilité différente y était présumée. Cette considération, faite par Wilson [1974], s'inspire du fait que les centres-villes nord-américains accaparent une bonne part du commerce de détail métropolitain bien que certains soient quasi-dépourvus de résident. Leur marché en est un de travailleurs. Selon le formalisme du bénéfice de localisation, un terme ϕ_{1j} a été défini comme un modèle d'interaction (voir 6.116). Le sous-modèle de localisation commerciale (6.133) devient alors:

$$S_{1t} = \Gamma_1(z)S_{1t} + \Gamma_0(z)\sum_j \phi_{1j} S_{jt} + B_1(z)P_{1t} + B_2(z)\sum_j \sigma_{1j} P_{jt} + F_1(z)\{S_{1t} + E_{1t}^s\} + F_2(z)\sum_j \phi_{1j}\{S_{jt} + E_{jt}^s\} \quad (6.134a)$$

$$s_t = \left[\Gamma_1(z)I + \Gamma_0(z)Q \right] s_t + \left[B_1(z)I + B_2(z)\Sigma \right] p_t + \left[F_1(z)I + F_2(z)\tilde{\Sigma} \right] \{s_t + e_t\} \quad (6.134b)$$

Nous avons inclus un terme régressif non-spatial en $F_1(z)e_t$ par souci

de complémentarité. Deux processus autorégressifs spatiaux s'exercent dorénavant au sein du vecteur s_t . L'un d'eux incarne l'autocorrelation géographique via la matrice de structure Q . L'autre désigne l'interaction liée à la consommation des employés commerciaux. Les deux termes se distinguent du fait que $Q \neq \tilde{Q}$, alors $Qs_t \neq \tilde{Q}s_t$. Le calcul de la distribution d'équilibre statique se fera dans l'appendice 6.A.

Le sous-chapitre suivant traitera des propriétés dynamiques que possèdent les modèles spatio-dynamiques de localisation. On découvrera que certaines conditions font que la convergence vers l'équilibre est importunée par des discontinuités. Plutôt que sources d'embarras, celles-ci ont un rôle particulièrement constructif, suscitant l'émergence de nouvelle organisation spatiale.

6.5 ANALYSE DYNAMIQUE ET PROPRIETES DE BIFURCATION

6.5.1 Quelques notions

L'analyse dynamique consiste à examiner les propriétés qualitatives d'un système doté d'une dimension temporelle. Ces propriétés concernent spécifiquement des caractéristiques évolutives des états du système. On peut nommer cinq éléments fondamentaux de descriptions:

- (1) le système d'équations dynamiques;
- (2) les états d'équilibre;
- (3) la stabilité structurelle des états d'équilibre;
- (4) la forme des trajectoires d'état;
- (5) les propriétés bifurcatoires des trajectoires d'état [Zeeman, 1977]

Le système d'équations dynamiques est un système mathématique décrivant l'évolution d'un ensemble de variables d'état. Une variable d'état est une mesure d'un quelconque attribut du système définie pour toute unité de temps de la période. Pour un temps continu, les variables sont décrites par des équations différentielles; pour des unités de temps discrètes, elles sont reproduites par équations de différences ou par des systèmes autorégressifs, de décalages distribués et/ou de moyennes mobiles.

Un état d'équilibre équivaut à un état d'immobilité soudain, stable ou instable, du système dynamique. Si la variable d'état est contenu dans un vecteur-fonction $x(t)$ dont l'évolution est décrite par le vecteur différentiel $dx/dt = f(x,u)$, le vecteur d'équilibre x^* correspond à l'égalité $dx/dt = f(x^*,u) = 0$. Pour identifier ces points critiques, il suffit de résoudre le système d'équations $f(x,u) = 0$.

L'intérêt de la recherche des états d'équilibre réside dans la reconnaissance de leur multiplicité et de leur stabilité. Pour comprendre ces concepts, il suffit d'analyser des équations polynômiales univariées comme fit le mathématicien René Thom [1972] dans sa proposition d'une théorie des modèles «morphogénétiques» (ou "genèse de la forme"). Incidemment, Thom introduisit le concept de «catastrophe», point critique délimitant deux ou plusieurs comportements évolutifs du système.

La notion de stabilité structurelle telle qu'employée dans la théorie des catastrophes, concerne la stabilité de la forme d'une fonction vis-à-vis les paramètres qui la composent. Considérons une fonction $f(x,u)$ où x est la variable d'état du système et u , un paramètre. Les points singuliers réguliers de f deviennent les maxima et minima de cette fonction i.e. $\partial f / \partial x = 0$. On remarquera que les valeurs de x respectant cette égalité sont données par une fonction de u seulement. Les singularités structurelles (dégénérées) de f coïncident par contre aux points se rapportant à $\partial^2 f / \partial x^2 = 0$ au voisinage de $\partial f / \partial x = 0$. Ces points identifient les valeurs de x et u où apparaissent ou disparaissent un maximum ou un minimum (une valeur d'équilibre). Pour un système à multiples localisations (x_i ; $i=1,2,\dots,n$), l'instabilité structurelle se vérifie lorsque la matrice Hessienne de dérivées secondes ($H = \{\partial^2 f / \partial x_i \partial x_j\}$) est singulière ($|H| = 0$) [Wilson 1981].

Prenons par exemple un polynôme cubique de x doté de paramètres a , b , c et d : $f(x) = ax^3 + bx^2 + cx + d$

Les points singuliers réguliers sont donnés par la première dérivée annulée: $\partial f / \partial x = 3ax^2 + 2bx + c = 0$

Selon la formule de Viète, les racines de ce polynôme quadratique sont:

$$x^* = \frac{-2b \pm \sqrt{4b^2 - 12ac}}{6a}$$

Présumons que le terme $(b^2 - 4ac)$ soit positif pour s'assurer que les racines x^*_1 et x^*_2 soient réelles. De par la fonction considérée, ces deux points critiques représentent un minimum et un maximum de la fonction. La dérivée seconde calculée sur ces points donnera l'identité exacte. En effet:

$$\left. \frac{\partial^2 f}{\partial x^2} \right|_{x=x^*} \begin{cases} > 0 & x^* \text{ est un minimum} \\ < 0 & x^* \text{ est un maximum} \end{cases}$$

Qu'arrive-t-il lorsque $\partial^2 f / \partial x^2 = 0$? Le point x^* n'est alors plus un minimum ni plus un maximum. Il ne peut y avoir non plus deux de ces points. En fait, au voisinage de l'annulation de la dérivée seconde, les deux extrêma se confondent pour ne former qu'un seul point d'inflexion. Un tel point s'appellent une «singularité structurelle» car à son passage, la structure de la fonction change dramatiquement de forme. La figure 6.2 montre le passage qualitatif brutale du graphique de la fonction.

Ce point singulier est aussi affublé du nom évocateur de «catastrophe» car il correspond à une situation (statique) de transition brusque entre un système à deux points d'équilibre à un système qui en est dépourvu. Le changement peut se faire aussi dans l'autre direction. Dans un contexte dynamique, les points de catastrophe marquent des discontinuités dans la trajectoire d'évolution. Dans le sous-chapitre précédent, il a été dit que les systèmes gradients tendent inexorablement vers un état d'équilibre stable. Si, par contre, des variations paramétriques dénotant des changements de l'environnement se produisent, il se peut qu'apparaissent ou disparaissent d'autres états d'équilibre stables. Le système peut alors entreprendre une évolution qualitativement différente même si la trajectoire temporelle se montrera graduelle. La discontinuité n'apparaît pas comme une coupure de la trajectoire mais comme une réorientation tel un navire qui change de cap après avoir reçu ordre, au milieu de sa traversée, d'accoster dans un port autre que celui prévu.

Le concept de catastrophe est directement relié à celui de la multiplicité des états d'équilibre; il évoque les endroits où leurs nombres augmentent ou diminuent. La multiplicité engendre aussi une caractérisation des états d'équilibre selon leur stabilité. En fait, il ne peut y avoir deux états stables voisins sans que s'interpose un état instable. Les points instables ont autant d'intérêt que leurs rivaux car ils sont à l'origine de d'autres discontinuités.

Regardons encore la figure 6.2a. Considérons un système tendant à la minimisation de $f(x)$ tel une bille glissant sur les contours de la fonction. Le point x^*_2 devient alors stable car une bille s'y localisant ne peut guère s'échapper. Aussi dans l'intervalle $[x^*_1, \infty]$, le système ne peut que converger vers x^*_2 . Si initialement le système (ou la bille) se situe en deçà de x^*_1 , il poursuivra sa course vers l'infini négative ou une valeur limite inférieure de x . Dans le langage de la théorie des systèmes dynamiques, ces intervalles de convergence se nomment «bassins attracteurs» car ils délimitent un espace où tous les points (ou états) ne peuvent que converger vers un seul point stable. On comparera aisément cette définition à celle d'un "bassin versant" dont les eaux qui s'y recueillent sont irrémédiablement canalisées vers un seul point, l'embouchure.

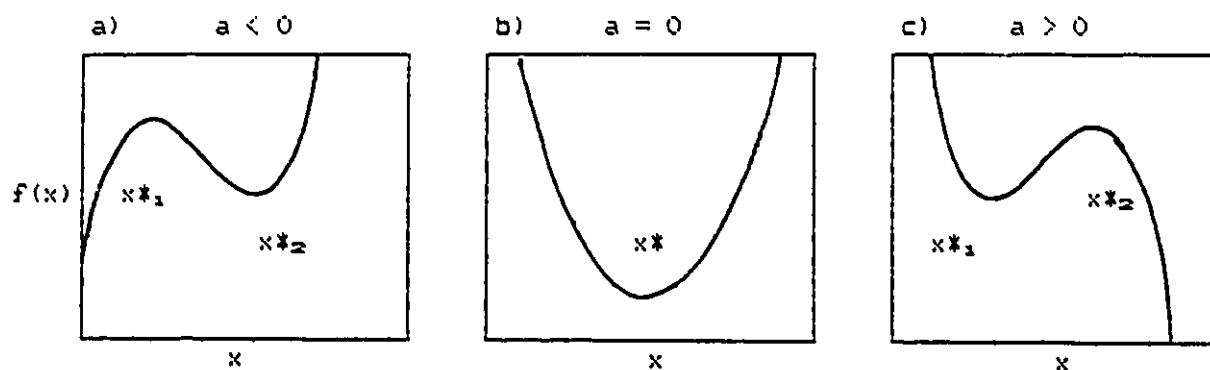


Figure 6.2: Formes génériques d'un polynôme cubique selon différentes valeurs de a : $f(x) = ax^3 + bx^2 + cx + d$ (x^*_n : valeurs critiques)

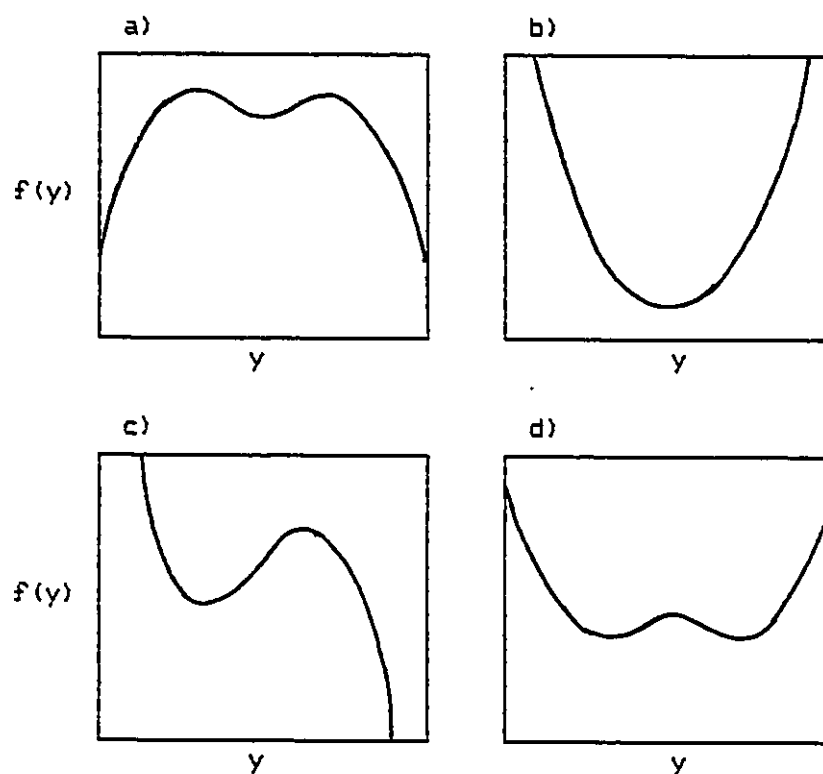


Figure 6.3: Formes génériques d'un polynôme quartique selon différentes valeurs paramétriques de $f(y) = ay^4 + by^3 + cy^2 + dy + e$

On poursuivra l'analogie en assimilant les points d'équilibre instables à la crête des montagnes délimitant les bassins versants. Dans la figure 6.2a, le point d'équilibre x^*_1 est naturellement instable car un état x d'un système minimisant peut alternativement se diriger vers x^*_2 ou le point infini $-\infty$ de même qu'une bille en équilibre ne pourra que chuter dans l'un ou autre receptacle. La figure 6.3 montre le graphique d'un polynôme quartique pourvu de trois points critiques. Si le système y est minimisant, il y a deux états d'équilibre stables et un seul instable. Ici, le point y^*_2 agit comme point de «bifurcation» entre deux états stables définis. Les bassins attrapeurs sont également bien définis.

Les bifurcations et les catastrophes (qui sont en fait un type de bifurcation [Wilson, 1981]) ont beaucoup d'implications dans la morphogénèse. Dans un système quelconque, les états d'équilibre stables évoquent les états les plus courants. Supposons que le système décrit par la variable d'état y (figure 6.3a) se situe au voisinage du point stable y^*_3 . Deux types de comportements discontinus s'avèrent possibles. D'abord, la fonction décrivant la caractéristique à minimiser $f(x)$ est munie de paramètres. Ceux-ci décrivent les modalités de l'environnement de y . Au passage de certains seuils paramétriques, la courbe quartique se transforme et perd de ses propriétés qui la caractérisent. Tantôt son lobe (point instable) disparaît si bien que les deux points stables se fondent ensemble. Dans d'autres circonstances, dépendant des paramètres affectés, la forme initiale généralement convexe devient momentanément cubique pour reprendre une forme quartique mais cette fois généralement concave (voir 6.2d). Ces transformations sollicitent toutes des modifications importantes des bassins attrapeurs si bien qu'un point y , en présence de perturbations paramétriques, peut échapper brusquement à son bassin initial et se retrouver dans un autre. Le système subit alors une «catastrophe» car il y a disparition ou émergence d'une forme stable du système.

L'autre type de discontinuité n'implique pas une intervention de la structure paramétrique. Elle relève directement de la variable d'état, de ses fluctuations normales et de sa proximité d'un état instable. Si ces fluctuations s'exercent lorsque le système évolue au voisinage de l'équilibre, les forces de convergence ramèneront plus ou moins rapidement la variable d'état vers la stabilité. Par contre, aux alentours de points

instables, une perturbation ayant une amplitude suffisante fera franchir les limites du bassin initial. Une bifurcation en résultera. Parfois, les systèmes se localisent dans des sections hautement instables i.e. où de nombreux bassins de petite taille se côtoient si bien que les moindres fluctuations sont déterminantes.

Dans la modélisation dynamique de nombreux phénomènes, l'analyse de bifurcation ouvre de nouvelles avenues de recherche car elle admet des comportements discontinus (voir Hazewinkel, Jurkovich et Paelinck [1985] pour une revue synthétique des applications). Thom [1972] fit ressortir le rôle déterminant des discontinuités dans l'appréhension de la naissance, la croissance et le déclin d'une forme qui correspondent, somme toute, à la morphogenèse. Le mathématicien mit surtout en relief les effets de catastrophes qui se présentent comme des caractéristiques statiques ayant des implications dynamiques. Les bifurcations d'origine fluctuante possèdent aussi des propriétés morphogénétiques.

A la fin des années soixante-dix, de nouvelles perspectives de modélisation furent mises de l'avant par des chercheurs de toute discipline inspirés des travaux du physicien Ilya Prigogine [Nicolis, Prigogine, 1977; Prigogine, Stengers, 1979; Prigogine, Allen, Herman, 1977; Prigogine, Allen, 1982]. On y réfère à de nouveaux concepts d'évolution qui admettent le fait que les systèmes, naturels et sociaux, tendent plutôt à se complexifier et se diversifier et, par conséquent, échappent à la fatalité de l'équilibre stable. Il s'agit spécifiquement d'une tendance à l'auto-organisation, connue sous le nom de «structure dissipative», résultant de l'action combinée de mécanismes d'interaction internes et de fluctuations externes. La présence de non-linéarités dans les mécanismes d'interaction fait que les perturbations sont amplifiées ou atténuées [Allen, 1982]. Ces non-linéarités créent une alternance de points critiques stables et instables d'un système. Au cours de l'évolution de celui-ci, certaines fluctuations, soit très prononcées, soit à proximité d'un point instable, font basculer le système vers des états alternatifs plus stables. Ces bifurcations de trajectoire amorcent l'émergence d'une nouvelle structure.

Ce paradigme de l'évolution fut appliqué à la modélisation du développement inter-urbain [Allen, Sanglier, 1979] et intra-urbain [Allen, Boon, Sanglier, 1980]. Dans le dernier cas, les auteurs proposèrent une formulation dynamique fondée sur un mécanisme de localisation à la Lowry. Les non-linéarités comme les économies d'échelle et les coûts de transport sont à l'origine de comportements bifurcants. Le système dénote des propriétés auto-organisantes du fait qu'au lieu de converger vers un état d'équilibre de la distribution spatiale des activités, on assiste à l'émergence ou à la disparition de certaines structures causées par des perturbations plus ou moins aléatoires. Cette dynamique nous apparaît irréversible car la trajectoire joignant l'état initial et l'état final a été ponctuée de bifurcations. Dans un système intra-urbain, le degré d'organisation se mesure à travers la spécialisation spatiale des activités. C'est en effet une distribution très éloignée d'une distribution uniforme laquelle semble peu organisée du point de vue de l'entropie informative. En absence de non-linéarité, le système converge inexorablement vers l'état maximisant les bénéfices de localisation, état dont le degré d'organisation dépend de celui donné par la distribution spatiale des variables exogènes.

Un segment de la théorie des systèmes dynamiques s'intéresse spécifiquement à la forme de la trajectoire et les bifurcations amenant un changement de cette forme. Plutôt que de scruter la stabilité structurelle (des états du système) de l'évolution, l'analyse porte désormais sur la stabilité dynamique (configuration des trajectoires). Cette théorie étudie comment la nature des solutions d'équations différentielles dynamiques change lorsque les paramètres varient [Hirsch, Smale, 1974]. Dans l'analyse d'un système dynamique, l'identification des valeurs paramétriques critiques semble le volet suscitant le plus d'intérêt.

L'analyse rigoureuse de la stabilité dynamique reste relativement difficile dès que le nombre de variable d'état dépasse trois. Pourtant, les changements qualitatifs de trajectoires les plus pertinents sont observés lorsque le contexte co-évolutif est riche. Les applications géographiques se révèlent plutôt limitées du fait qu'un système multi-zonal et multivarié se décrit par un nombre de variable d'état correspondant au nombre de zones multiplié par le nombre de variable. Néanmoins, les concepts et certaines méthodes se prêtent à l'observation et l'explication des comporte-

ments évolutifs complexes. La description y perd au chapitre du détail mais offre au chercheur un éclairage approprié à l'étude des phénomènes dynamiques. Les applications spatiales paraissent récentes et encore peu nombreuses [voir Dendrinou, Mullaly, 1985].

6.5.2 Stabilité structurelle des modèles de localisation

Dans le programme d'optimisation des bénéfices de localisation G (6.92-6.97) où alors G' (6.112-6.113, 6.93-6.97), les variables optimisées (p_{ij} et s_{ij}) sont des variables d'interaction entre les lieux i et j . Les variables de localisation se présentent tout simplement comme une somme de ces interactions. En aucun temps, elles symbolisent une quelconque mesure de bénéfice. Les problèmes G et G' demeurent donc intrinsèquement linéaires.

Pourtant, il paraît vraisemblable que la variable de localisation puisse aussi interpréter la variable d'attraction. Par exemple, la taille démographique peut devenir un élément d'attraction résidentielle de la zone. L'emploi commercial peut aussi être un indicateur lié aux bénéfices de consommation dans la zone en question. Cela se traduit algébriquement par:

$$Q_i = f(P_i) \quad \text{et} \quad W_j = f(S_j)$$

Afin de présenter une analyse plus claire, nous considérerons l'allocation d'une activité à la fois (l'autre étant tenue constante).

Supposons un problème d'optimisation G^* pour une allocation spatiale des résidences selon une distribution fixe de l'emploi de service $\{S_j\}$. Il correspond à la première moitié de la fonction (6.92) sujet à la contrainte (6.93). La solution de G^* est donnée par l'équation (6.98) et établit les conditions d'équilibre. (6.98) est linéaire en respect de P_i et est complètement déterminée par les paramètres et les variables exogènes. Pour chaque distribution $\{Q_i, E_j, S_j, c_{ij}\}$ et chaque combinaison paramétrique $\{\alpha, \psi^*\}$, il existe donc une distribution positive et unique $\{P_i\}$. Parce que (6.98) est linéaire, la matrice Hessienne de dérivées secondes est nulle et, par conséquent, le système est structurellement stable (aucune singularité). Par exemple, une variation du paramètre ψ^* modifie la répartition d'équilibre de l'activité résidentielle de façon continue.

Si un problème G^* pour l'activité commerciale, analogue à celui présenté précédemment, est défini, on en viendra aux mêmes conclusions. Cependant, il est bien connu que l'attrait d'une zone commerciale pour un consommateur dépend beaucoup de la taille des facilités. Par ailleurs, il semble concevable d'envisager que cette taille est proportionnelle au personnel employé dans les établissements commerciaux. Il sera présumé que, pour le consommateur, des économies d'échelle peuvent être réalisées. Le facteur d'attraction devient alors une fonction curvilinéaire de S_j :

$$W_j = f(S_j) = S_j^\alpha$$

Ecrivons le programme G^* correspondant:

$$\max_{\{S, s\}} G^* = \frac{-1}{\psi^\alpha} \sum_j \sum_i s_{ij} \left[\ln s_{ij} - 1 - \alpha \ln S_j \right] - \sum_{ij} s_{ij} c_{ij} \quad (6.135)$$

$$\text{sujet à } \sum_j s_{ij} = \phi P_i \quad \sum_j S_j = S = \phi P \quad (6.136-6.137)$$

Les distributions sont optimales si les dérivées partielles sont nulles:

$$\frac{\partial G^*}{\partial s_{ij}} = \frac{-1}{\psi^\alpha} \ln s_{ij} + \frac{\alpha}{\psi^\alpha} \ln S_j - c_{ij} + \eta_i = 0 \quad (6.138)$$

$$\frac{\partial G^*}{\partial S_j} = \frac{\alpha \sum_i s_{ij}}{S_j} \quad (6.139)$$

Selon les conditions précédentes, on en déduit les formulations suivantes:

$$s_{ij} = S_j^\alpha f_{ij} A_i \phi P_i \quad (6.140)$$

$$\text{où } f_{ij} = \exp(-\psi^\alpha c_{ij}) \quad A_i = [\sum_k S_k^\alpha f_{ik}]^{-1}$$

$$S_j = \sum_i s_{ij} \equiv R_j \quad (6.141)$$

Cette formulation s'apparente avec le modèle d'achat de Lakshmanan et Hansen [1965] qui a été présentée dans la section 6.2.2 et dont les propriétés de catastrophe ont été maintes fois étudiées notamment par Harris et Wilson [1978], Clarke [1981], Wilson [1981], Harris et al. [1982], Kaashoek et Vorst [1984]. L'interdépendance des équations (6.140-6.141) suppose une ou des situations d'équilibre par combinaison paramétrique $\{\bar{\alpha}, \alpha, \psi^\alpha\}$ et distribution $\{P_i, c_{ij}\}$. L'équation d'équilibre coïncide avec une singularité régulière de la fonction d'optimisation, soit la première dérivée mise à zéro:

$$\partial G^* / \partial S_j \approx R_j - S_j \quad (6.142)$$

La forme générale de la surface d'équilibre est donnée par l'ensemble des valeurs de S_j respectant l'égalité (6.142). Il semble y avoir trois cas particuliers;

- 1- $S_j^* = 0$, donc $R_j = 0$
- 2- $0 < S_j^* = R_j < S$ (S = valeur maximale)
- 3- $S_j^* = S \mid S_k = 0; \forall k \neq j$.

Bien que ces valeurs soient réalisables, elles ne sont pas toutes pour autant stables. Une valeur d'équilibre est dite stable si une faible perturbation suscite un retour vers celle-là. Autrement dit, la condition de stabilité équivaut à la condition des maxima pour G^* ; elle s'établit comme suit:

$$\frac{\partial^2 G^*}{\partial (S_j)^2} = \frac{\partial (R_j - S_j)}{\partial S_j} < 0 \quad \text{pour } S_j^* = R_j \quad (6.143)$$

Dans un système dynamique gradient, les points d'équilibre stables jouent les «attracteurs» car, à leur proximité, la trajectoire converge obligatoirement. Les points qualifiés d'instables sont dispersifs puisqu'ils repoussent la variable vers des états stables. Ils apparaissent donc des facteurs de bifurcation lorsque des perturbations accompagnent la dynamique du système. Le croisement de l'un de ces seuils change radicalement le régime évolutif et réoriente la variable vers un autre état stable; il y a alors émergence d'une nouvelle structure ou forme. Ce processus morphogénétique animé par les fluctuations s'apparente à ceux rencontrés chez les phénomènes physico-chimiques et biologiques [Nicolis, Prigogine, 1977]. Cette perspective de modélisation s'appuie sur la capacité des mécanismes d'interaction non-linéaires de doter le système des propriétés d'«auto-organisation» lui permettant de modifier sa structure au gré de son évolution.

Harris et Wilson [1978] examinèrent la dérivée de l'équation d'équilibre du modèle dynamique de structure commerciale. Cependant, les conclusions qu'ils tirèrent ne tiennent pas ici car la variable endogène étudiée, la taille des facilités commerciales, n'était pas assujettie à une contrainte linéaire de total comme (6.137). Cette contrainte modifie passablement le comportement variationnel de S_j puisqu'elle introduit une dépendance entre chacun des effectifs zonaux. La seconde dérivée du problème (6.135-6.137) en respect de S_j s'écrit:

$$\frac{\partial^2 G^*}{\partial (S_j)^2} = \sum_i \alpha S_j^{\alpha-1} f_{1j} A_i \gamma P_i - \sum_i \alpha S_j^\alpha f_{1j} A_i \gamma P_i \cdot \left[S_j^{\alpha-1} f_{1j} - \sum_{k \neq j} S_k^{\alpha-1} f_{1k} / (N-1) \right] - 1 \quad (6.144)$$

Lorsque la contrainte (6.137) est inactive, le second terme entre crochet disparaît tout simplement puisque les variations de S_j n'impliquent plus une correction chez les autres effectifs. Avec (6.144), la forme graphique générale de R_j - S_j peut être tracée à partir de la pente calculée aux points d'équilibre. A cause de la non-linéarité engendrée par l'exposant α , l'analyse doit tenir compte des valeurs prises par ce paramètre. Examinons d'abord le cas le plus simple, soit $S_j^* = 0$:

$$\frac{\partial^2 G^*}{\partial (S_j)^2} = \begin{cases} \rightarrow \infty & (\alpha < 1) \\ = \sum_i A_i f_{1j} \gamma P_i - 1 & (\alpha = 1) \\ = -1 & (\alpha > 1) \end{cases}$$

Lorsque $\alpha < 1$, l'état équivalent à l'absence d'activité commerciale n'apparaît jamais une possibilité stable. Quand les économies d'échelle semblent directement proportionnelles à l'importance de l'activité ($\alpha = 1$), la convergence ou divergence vers cette situation d'équilibre dépend de la localisation relative de la zone i . Pour une zone favorablement localisée par rapport à un marché, la dérivée étant positive, l'activité tend plutôt à émerger. Néanmoins, l'équilibre nul demeure un point attracteur en présence d'économies d'échelles progressives ($\alpha > 1$).

Pour poursuivre notre analyse lorsque l'équilibre est positif non-nul, il convient de reformuler (6.144) en des termes plus révélateurs. Il suffit d'employer le terme d'interaction s_{1j} :

$$\begin{aligned} \frac{\partial^2 G^*}{\partial (S_j)^2} &= \alpha \cdot \sum_i \frac{s_{1j}}{S_j} - \alpha \cdot \sum_i \frac{(s_{1j})^2}{S_j \gamma P_i} + \alpha \cdot \sum_i \frac{s_{1j}}{\gamma P_i} \cdot \left[\sum_{k \neq j} \frac{s_{1k}}{S_k (N-1)} \right] - 1 \\ &\equiv \alpha \cdot [J_j - K_j + L_j] - 1 \end{aligned} \quad (6.145)$$

Cette formulation possède des caractéristiques aidant à la découverte de la stabilité. En effet, d'après la contrainte (6.136) et la condition d'équilibre (6.141) :

$$\sum_j s_{1j} = \bar{o} P_1 \quad \sum_i s_{1j} = S_j^*$$

Par conséquent, les trois termes de (6.145) concernent exclusivement des sommes de probabilités. Ici, des relations constantes sont vérifiées si tous les S_j sont non-nuls :

$$J_3 = 1, \quad 0 < K_3 \leq 1, \quad 0 < L_3 \leq 1 \quad \text{et} \quad K_3 \geq L_3$$

Ceci permet de conclure pour un équilibre de type $0 < S_3^* = R_3 < S$:

$$-1 < \alpha \cdot [1 - K_3 + L_3] - 1 \leq \alpha - 1$$

Forcément:

$$\frac{\partial^2 G^*}{\partial (S_3)^2} = \begin{cases} < 0 & (\alpha < 1) \\ \leq 0 & (\alpha = 1) \\ > 0 & (\alpha > 1) \end{cases}$$

Pour $\alpha \leq 1$, il ne peut qu'y avoir qu'une seule valeur d'équilibre stable car la fonction $R_3 - S_3$ ne croise que négativement l'axe $R_3 - S_3 = 0$. Le cas $\alpha > 1$ soulève de l'intérêt car il contient parfois une valeur stable. Intuitivement, si une telle valeur existe, alors existe aussi une valeur instable plus petite étant donné que $S_3^* = 0$ reste toujours stable. En effet, la fonction $R_3 - S_3$ doit obligatoirement croiser l'axe $R_3 - S_3 = 0$ positivement entre deux croisements à pente négative.

Le cas particulier $S_3^* = S$ ($S_k = 0$; $\forall k \neq i$) où toute l'activité commerciale de la région urbaine se situe dans une même zone paraît une valeur d'équilibre plausible. Puisque:

$$S_3^* = R_3 > 0 \quad \text{et} \quad S_3^* = S = \partial P, \quad \text{alors}$$

$$J_3 = 1, \quad K_3 = 1 \quad \text{et} \quad \partial^2 G^* / \partial (S_3)^2 = \alpha \cdot L_3 - 1$$

Par ailleurs, cette possibilité de distribution impliquant des valeurs nulles chez les zones voisines, le terme L_3 ne présente plus la limite unitaire. En se référant à (6.145), on retrouve les valeurs limites suivantes:

$$\frac{\partial^2 G^*}{\partial (S_3)^2} = \begin{cases} > \infty & (\alpha < 1) \\ = \sum_k \frac{\Sigma_{k \neq j} f_{1k}}{(N-1)f_{1j}} \cdot \frac{P_1}{P} - 1 & (\alpha = 1) \\ = -1 & (\alpha > 1) \end{cases}$$

Dans le cas où $\alpha < 1$, puisque l'absence d'activité est instable, la concentration commerciale dans une même zone, impliquant une disparition chez les autres, est également instable. Pour que le monopole soit soutenable avec $\alpha = 1$, il faut réunir des conditions spécifiques plutôt rares. En effet, cela exige que la somme de tous les rapports entre la moyenne hors-zone d'accessibilité et l'accessibilité de la zone se solde par une fraction de l'unité. Dans un système muni d'économies d'échelle progressives ($\alpha > 1$),

la situation de monopole demeure toujours réalisable. Cela suppose aussi que s'il existe une valeur d'équilibre stable non-nulle, il y a une valeur répulsive entre le maximum S et l'autre valeur.

L'identification des valeurs d'équilibre et leurs statuts en matière de stabilité permet de tracer la forme générale la fonction R_j-S_j . La figure 6.4 montre la forme de cette fonction pour quatre valeurs de α . Avec la croissance du paramètre, la valeur stable principale de S_j devient de plus en plus grande jusqu'à monopolisation. Il s'agit ici du comportement de la zone la mieux localisée dans un système de deux zones. Il faut noter la pente infinie pour les valeurs extrêmes $S_j=0$ et $S_j=S$ lorsque α est inférieur à un. On remarque aussi l'apparition de croisement positif dès que α dépasse l'unité. L'observation de ce graphique permet de dessiner la courbe d'équilibre avec l'exposant en abscisse (fig. 6.4a-b). Les lignes continues réfèrent aux équilibres stables (régimes attractifs) tandis que les pointillés identifient les équilibres instables (régimes dispersifs). Après la valeur pivot de 1, les états extrêmes deviennent stables. Il y a alors apparition de deux limites répulsives. L'une d'elles s'avère un rebroussement du régime principal stable. Selon Thom [1972, chap.5], il s'agit d'une catastrophe de type «pli». A la pointe du rebroussement ($\alpha = \alpha'$), un système doit définitivement sauter à l'un des états extrêmes. Dans un environnement dynamique, le système fait face à une bifurcation causée par un changement paramétrique, en l'occurrence une variation du facteur d'économie d'échelle. L'espace située entre deux équilibres instables désigne un bassin d'attracteur à l'intérieur duquel le système est convergent. En présence de plusieurs bassins d'attracteur, un autre type de bifurcation peut arriver. En effet, si de fortes perturbations font quitter une variable d'état de son bassin initial, celle-ci cheminera dorénavant vers un autre état d'équilibre. Au voisinage de la valeur critique $\alpha = \alpha'$, le système demeure instable et très sensible à la moindre fluctuation; ici, les bifurcations semblent plus propices. Un tel comportement bifurcatoire réitère le rôle crucial des perturbations dans l'évolution d'une structure. Dans un système où les mécanismes d'interaction ne présente pas de non-linéarité (problème G''), les déviations exogènes sont incessamment corrigées et ne peuvent modifier qualitativement la structure. Par contre, les systèmes dotés de non-linéarités génèrent des moments d'instabilité et se servent des perturba-

tions pour aiguiller leurs trajectoires [Allen, Sanglier 1981]. Par conséquent, ils auto-organisent leurs structures.

Dans un système urbain, les perturbations interprètent les initiatives d'aménagement, privées ou centralisées, qui dérogent aux mécanismes spatio-économiques dominants. L'amélioration des réseaux de transport, l'implantation d'une grande firme industrielle ou d'un équipement collectif majeur, l'expansion d'un centre commercial dans le but d'augmenter sa part du marché sont autant de perturbations affectant la structure spatiale de la ville. Celle-ci peut aussi bien «absorber» la perturbation que se réorganiser en fonction du nouvel environnement.

La surface d'équilibre peut être visionnée à travers un feuilletage différent comme celui donné par le paramètre de transport ψ^* . Selon l'équation (6.144), la fonction du coût de transport $f_{1,j}$ n'affecte pas la stabilité des points d'équilibre. En principe, les variations de ψ^* (et de $c_{1,j}$) modifient plutôt les seuils de rebroussement mais n'altèrent aucunement le genre de catastrophe. Il s'agit d'une propriété fondamentale des modèles morphogénétiques:

«On peut identifier le type de catastrophe et son origine dynamique sans qu'on ait à connaître explicitement tous les paramètres internes dont dépend le système» [Thom 1972 p.62].

Les figures 6.6a-c illustrent les courbes d'équilibre avec le paramètre ψ^* en abscisse selon trois valeurs de α . La catastrophe-pli se manifeste encore sur le graphique où $\alpha > 1$. On reconnaît alors que la fonction de bénéfice G^* est structurellement singulière au point $\alpha = 1$. Le régime principal correspondant au maximum global de G^* est également interrompu à la valeur $\psi^{*'}$. Ce point singulier varie toutefois selon α , la matrice de transport C , le vecteur de population p et la zone i considérée. Quand $\alpha = 1$, il existe une singularité fixe à $\psi^* = \psi^{*'}$. A $\psi^* = 0$, la friction du déplacement n'existe plus et toutes les zones ont accès équitablement au marché régional; la valeur d'équilibre $S_j^* = S/N$ est accordée à tous. Pour $\alpha < 1$, ce point apparaît stable; lorsque $\alpha = 1$, toutes les valeurs $0 \leq S_j \leq S$ sont stables. Dès que α dépasse l'unité, seules les états extrêmes restent tandis que $S_j^* = S/N$ joue un rôle séparateur. Quand l'impédance du transport est faible mais non-nulle, le modèle tend à privilégier les zones

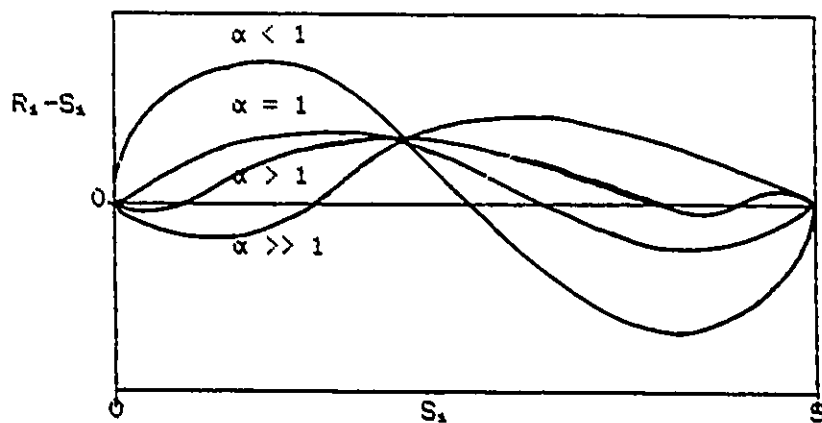
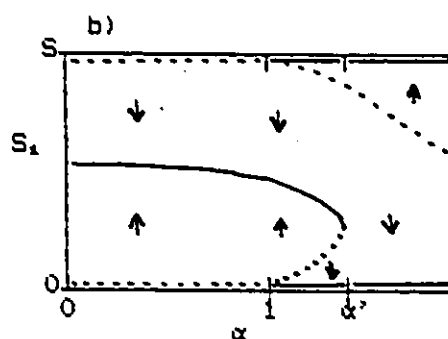
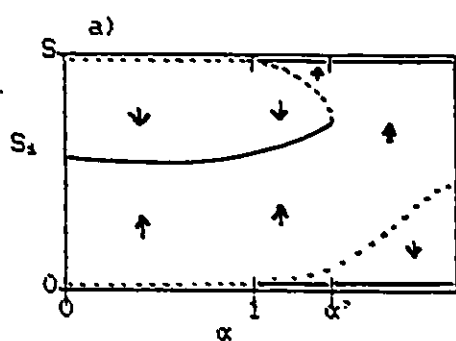


Figure 6.4:
Courbes $R_1 - S_1$
versus S_1



Régime
— stable
.. instable

Figure 6.5: Courbes d'équilibre pour " α ": a) zone avantageusement localisée
b) zone marginale

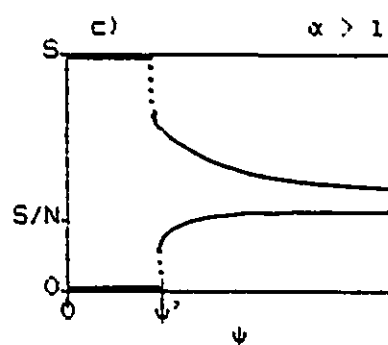
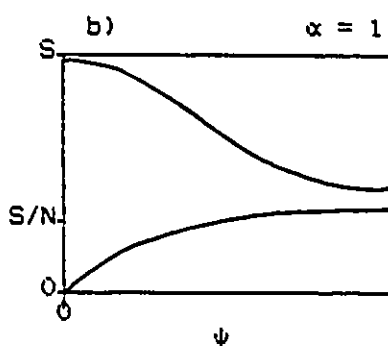
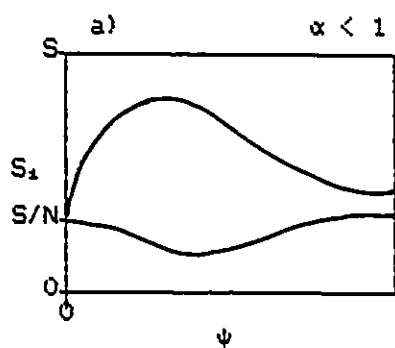


Figure 6.6: Courbes d'équilibre pour " ψ ": a) $\alpha < 1$; b) $\alpha = 1$; c) $\alpha > 1$

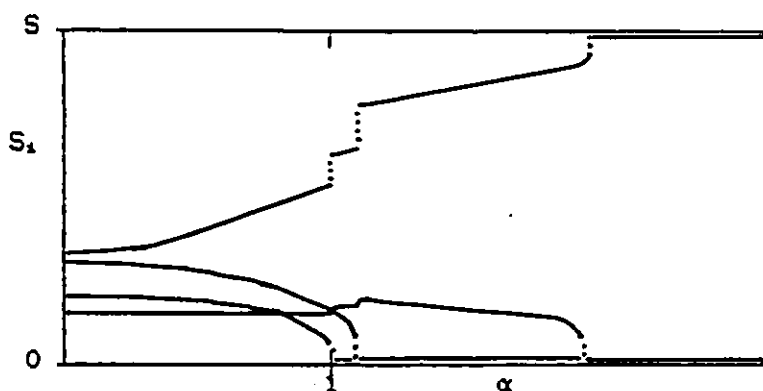


Figure 6.7:
Courbes d'équilibre
pour " α " pour un
système de 4 zones
(régimes principaux
seulement)

favorablement localisées au dépens des zones moins pourvues. Dans certaines conditions paramétriques, ce privilège signifie la monopolisation de l'activité (voir fig. 6.6c). A mesure que ψ^* croît, l'activité commerciale d'une zone tend à être proportionnel au marché local ($S_j \propto \phi_j$).

En fait, les représentations graphiques des courbes d'équilibre simplifient grandement la réalité que confrontent habituellement les systèmes urbains. Dans un contexte multi-zonal, les seuils paramétriques de catastrophe (point de rebroussement) sont propres à chaque zone mais affectent le comportement d'équilibre des autres. Ainsi, la disparition de l'activité chez une zone correspond à une amplification subite du régime principal de ses voisins, les quantités retirées étant réallouées ailleurs. Ce processus de «récupération» opère évidemment plus intensément avec les espaces à proximité.

Pour témoigner de ce phénomène, les valeurs d'équilibre ont été évaluées pour un système de quatre zones. Le graphique de la figure 6.7 expose les changements brusques des régimes principaux liés aux catastrophes-plis manifestées chez les zones voisines et générées par des variations du paramètre α . On observe la formation d'états paliers à chaque fois qu'une autre zone liquide ses activités commerciales. Chaque palier correspond à une organisation commerciale distincte de la ville. Un système multi-spatial poursuivant une dynamique où α est croissant, fait face à des catastrophes successives. La présence de nombreux paliers d'un régime principal signifie aussi de nombreux secteurs d'instabilité où le rôle des perturbations permet à chaque fois des changements discrets de la structure d'équilibre.

La section présente a montré que les modèles de structure urbaine dotés de non-linéarités occasionnent une multiplicité des états d'équilibre. Dans une analyse dynamique, le statut de stabilité des ces états distingue les points de convergence de ceux de divergence. Les propriétés de bifurcation qui ont été dévoilées relève, d'une part, des variations chez les paramètres qui, à leur valeur singulière, provoque une cessation de stabilité d'un état et, conséquemment, impose subitement un nouveau point de convergence. D'autre part, dans un environnement paramétrique constant, la co-existence d'états stables et instables fait qu'une fluctuation de la

variable d'état fait basculer celle-ci dans un nouveau bassin d'attracteur. La section suivante ne portera plus sur le comportement de l'état d'équilibre (stabilité structurelle) comme facteur de bifurcation mais sur les conditions affectant l'apparence de la trajectoire dans le temps.

6.5.3 Stabilité dynamique d'un processus co-évolutif de localisation

La stabilité dynamique réfère au comportement de la variable d'état, usuellement à proximité de l'équilibre. Il y a bifurcation dynamique si la trajectoire d'état change qualitativement à cause d'une variation de paramètre. Cette trajectoire est habituellement interprétée par la solution des équations différentielles commandant l'évolution du système étudié.

Les équations (6.98) et (6.99) représentaient chacune la solution des fonctions dynamiques (6.125) et (6.126) prises séparément i.e. en tenant constante l'autre distribution. Dans ce contexte univarié, le point d'équilibre «attracteur» contrôle le comportement évolutif de l'activité dans l'espace urbain. La section précédente s'est particulièrement intéressée à la stabilité (structurelle) de ces points (correspondant ici aux maxima de la fonction de bénéfice).

Ici, l'investigation ne porte plus sur la nature de l'équilibre mais sur la dynamique conjointe des activités en rapport avec l'équilibre. Pour examiner les propriétés qualitatives des trajectoires, il est préférable d'analyser les solutions d'un système linéarisé autour de l'équilibre. La linéarisation permet des solutions plus simples et toutes autant révélatrices sur la forme des trajectoires [Dendrinos, Mullaïy 1985]. Voici un système d'équations différentielles représentant l'évolution conjointe de variables x_i ; $i=1,2,...,N$:

$$\frac{dx}{dt} = F(x) \quad (6.146)$$

$$dx/dt = \begin{bmatrix} dx_1/dt \\ dx_2/dt \\ \vdots \\ dx_N/dt \end{bmatrix} \quad F(x) = \begin{bmatrix} F_1(x_1, x_2, \dots, x_N) \\ F_2(x_1, x_2, \dots, x_N) \\ \vdots \\ F_N(x_1, x_2, \dots, x_N) \end{bmatrix}$$

$F(x)$ est un vecteur de N fonctions multivariées. Les valeurs d'équilibre x^* sont obtenues dans la condition d'immobilité suivante:

$$\frac{dx^*}{dt} = 0 \quad (6.147)$$

Par contre, de par la non-linéarité de $F(x)$, la solution de (6.146) reste difficile voire parfois introuvable, même pour les formes fonctionnelles les plus simples. La linéarisation de $F(x)$ au voisinage du vecteur d'équilibre x^* facilite la recherche d'une solution. L'expansion de Taylor de $x^* + \zeta$ où ζ est un vecteur de petits incréments, donne:

$$F_1(x^* + \zeta) = F_1(x^*) + \sum_j \frac{\partial F_1}{\partial x_j} \cdot \zeta_j + \sum_j \frac{\partial^2 F_1}{\partial (x_j)^2} \cdot \frac{(\zeta_j)^2}{2!} + \dots \quad (6.148)$$

Le premier terme apparaît nul de par la condition d'équilibre (6.147). En éliminant les termes supérieurs au premier degré (ils sont insignifiants car $\zeta_j \rightarrow 0$; $\forall j$), on obtient l'approximation linéaire:

$$\frac{dx}{dt} = Ax \quad \text{où } A_{N \times N} = \{a_{ij}\}, \quad a_{ij} = \left. \frac{\partial F_1}{\partial x_j} \right|_{x^*} \quad (6.149)$$

Les éléments a_{ij} de la matrice Jacobienne A précisent le comportement du système près de l'état x^* . La solution du système approximant (6.149) a la forme générale suivante (voir appendice 6.B):

$$x(t) = G \cdot \exp(\lambda t) \quad (6.150)$$

La matrice G contient des constantes relatives aux conditions initiales tandis que λ est un vecteur des N racines propres de la matrice A . Chacune des racines résout le polynôme du $N^{\text{ème}}$ degré suivant:

$$|A - \lambda I| = 0 \quad (6.151)$$

Les racines λ_i sont des nombres complexes du type $\lambda = v + \omega i$ où $i = \sqrt{-1}$. La solution (6.150) apparaît donc une somme de fonctions exponentielles complexes. Par ailleurs, on sait pertinemment que:

$$\exp[(v + \omega i)t] = \exp(vt) \cdot [\cos \omega t + i \sin \omega t]$$

La partie réelle $\text{Re}(\lambda) = v$ indique une croissance ou un déclin alors que la partie imaginaire $\text{Im}(\lambda) = \omega$ introduit un mouvement oscillatoire. L'analyse de bifurcation dynamique peut porter alors sur les valeurs que prennent ces racines en relation avec les entrées de la matrice dynamique A .

L'évaluation des racines caractéristiques demeure une tâche très ardue car on doit développer le déterminant (6.151) et recourir à des méthodes de solution polynômiale. Cependant, il existe certaines propriétés des racines permettant d'identifier leurs natures à partir de certains attributs de la matrice. La somme des parties réelles des racines $\sum_i \text{Re}(\lambda_i)$ est toujours égale à la somme des éléments diagonaux de A i.e. le trace $\text{Tr}(A)$; si celui-ci est positif, il y a au moins une racine positive et le système devient instable (explosif). Si $\text{Tr}(A) < 0$, il existe trois possibilités:

- ($a_{ii} \leq 0; \forall i$): toutes les racines sont nécessairement négatives et, en conséquence, le système est convergent;
- Certaines entrées sont positives: si les toutes les racines sont complexes ($\text{Im}(\lambda_i) \neq 0$), $\text{Re}(\lambda_i) = \text{Tr}(A)/N$; elles sont donc toutes négatives;
- Certaines entrées sont positives: si certaines racines sont purement réelles, il y en a qui sont positives; le système devient globalement instable car au moins une trajectoire est divergente.

Les racines propres sont complexes lorsque [voir Hirsch, Smale 1974]:

$$\text{Det}(A) > (\text{Tr}(A)/N)^2$$

La déterminant est une somme du produit des éléments diagonaux avec les cofacteurs d'ordre $N-1$. Le déterminant est toujours inférieur au terme de droite lorsque les entrées symétriques sont du même signe. Sinon, le nombre de racines complexes dépend de la magnitudes relatives entre entrées diagonales et triangulaires.

Reformulons le modèle dynamique de localisation (6.125-6.126) afin qu'il puisse s'assimiler à un système comme (6.149):

$$dP_i/dt = \varphi \cdot (X_i - P_i + \sum_j \pi_{ij} S_j) \quad (6.152)$$

$$dS_i/dt = s \cdot (\sum_j \sigma_{ij} P_j - S_i) \quad (6.153)$$

où $X_i = \sum_j \pi_{ij} E_j$ (constant); π_{ij} et σ_{ij} définis par (6.100) et (6.101).

Pour ce système dynamique à deux activités, la matrice A reste facile à déterminer car les coefficients sont linéaires en respect de P_i et S_j :

$$A = \begin{bmatrix} -I & \begin{matrix} T \\ \hline \end{matrix} \begin{matrix} \{\pi_{ij}\} \\ \hline \end{matrix} \\ \begin{matrix} \{\sigma_{ij}\} \\ \hline \end{matrix} & -I \end{bmatrix} \quad (6.154)$$

Un tel système occasionne des trajectoires stables car $\text{Tr}(A) = -2N$ et $a_{kk} = -1$; $k=1,2,\dots,2N$. Il ne peut y avoir non plus d'oscillations car les entrées triangulaires sont toujours non-négatives. Les volumes d'activité convergent donc vers la distribution d'équilibre correspondant au maximum de bénéfice (programme G, éqs 6.92-6.97). Maintenant, considérons une dynamique alternative en supposant que la croissance d'une activité dans une zone inhibe celle de l'autre en diminuant la quantité de sol disponible. Cette relation compétitive pour le sol peut être symbolisée par une relation linéaire négative avec paramètre g et h :

$$dP_i/dt = \varphi \cdot (X_i - P_i + \sum_j \pi_{ij} S_j - g S_i) \quad (6.155)$$

$$dS_i/dt = s \cdot (Y_i - \sum_j \sigma_{ij} P_j - h P_i - S_i) \quad (6.156)$$

où $X_i = \sum_j \pi_{ij} E_j + \text{constante}$; $Y_i = \text{constante}$

Les équations d'équilibre apparaissent comme suit:

$$P_i = X_i + \sum_j \pi_{ij} S_j - g S_i \quad S_i = Y_i + \sum_j \sigma_{ij} P_j - h P_i \quad (6.157-6.158)$$

Les coefficients g et h équivalent aux taux de consommation de sol des activités commerciales et résidentielles respectivement. Ici, les distributions d'équilibre ne sont plus seulement déterminées par l'accessibilité mais tiennent compte aussi de l'occupation du sol. Un tout nouveau patron se dessine car les zones à forte concentration d'une activité tendent à chasser l'autre même si la proximité de celle-ci lui est généralement favorable. Cette relation ambivalente ressort davantage quand g et/ou h sont élevés. La matrice Jacobienne prend la forme suivante:

$$A = \begin{bmatrix} -I & \text{}^T \{\pi_{ij}\} - gI \\ \text{}^T \{\sigma_{ij}\} - hI & -I \\ \text{}^T \text{} & \text{}^T \end{bmatrix} \quad (6.159)$$

Ici, les éléments triangulaires de A sont en tous temps positifs à l'exception des entrées $\{\pi_{ii} - g\}$ et $\{\sigma_{ii} - h\}$ correspondant aux relations intra-zonales. Le premier terme désigne l'interaction positive liée aux échanges inter-activités tandis que le deuxième définit l'effet négatif occasionné par l'occupation conjointe du sol. Lorsque, pour une zone i , l'effet inhibiteur dépasse l'effet bénéfique, la relation inter-activité intra-zonale devient globalement négative. Si g et h sont tels que $\pi_{ii} - g$,

$\sigma_{11}-h < 0$; $\forall i$, la système converge vers une structure urbaine spécialisée zonalement. L'absence d'oscillation est démontrée du fait que les éléments symétriques de A conservent le même signe.

Maintenant, supposons que l'occupation du sol commerciale limite passablement le développement résidentiel et g et h sont tels que $\pi_{11}-g < 0$, $\sigma_{11}-h > 0$; $\forall i$. Ici, l'activité domiciliaire tire parti de la proximité des commerces (sources d'emploi) seulement si ceux-ci se localisent dans les zones voisines. Le développement commercial, quant à lui, profite généralement de l'accessibilité de la population, ceux-ci occupant une faible portion du sol zonal ($h > 0$). Il y a alors possibilité d'oscillation de par l'inversion des signes chez les entrées symétriques; cette possibilité est augmentée lorsque l'écart entre $\pi_{11}-g$ et $\sigma_{11}-h$ et entre π_{1j} et σ_{1j} ($i \neq j$) est grand. Une corrélation négative entre $\{Q_i\}$ et $\{W_i\}$ accentue évidemment cet écart. Un ψ^P faible (signifiant une grande mobilité pour les déplacements de travail) couplé à un ψ^S élevé (indiquant une mobilité restreinte pour les déplacements pour consommation) favorise également l'écart numérique entre les entrées hors-diagonales de A. Cet environnement paramétrique particulier est fréquent dans les métropoles nord-américaines actuelles. La variation combinée de g , h , ψ^P et ψ^S permettant la transition entre une solution stable monotone et une solution stable oscillatoire provoque une bifurcation essentiellement dynamique. La structure urbaine vers laquelle chemine ce système deviendra hautement spatialement spécialisée, les établissements de service occupant des secteurs à dominance commerciale, recherchant l'accessibilité des domiciles bien que les refoulant hors de leur propre territoire. La localisation des résidences reste plus indépendante des lieux d'emplois et coïncide avec les zones zones dotées d'attrait résidentiel. L'antagonisme des deux activités donnera lieu à des fluctuations de leur croissance dans une même zone et mais favorisera la dispersion spatiales des lieux résidentiels.

La dynamique de la structure urbaine décrite précédemment s'apparente à celle de la relation «prédateur-proie» en écologie. Cette dynamique fut maintes fois étudiée en recourant au formalisme de Lotka-Volterra (voir May [1981] pour une synthèse de l'écologie mathématique). Le système de Lotka-Volterra est un système d'équations différentielles représentant la dynamique de population de deux communautés écologiques. La solution d'un

modèle prédateur-proie conduit à une évolution cyclique en présence de perturbations [May, Oster, 1976]. Ici, l'activité commerciale paraît prédatrice car sa croissance dans une zone fait décroître la population bien qu'elle profite de celle-ci en tant que marché. Cependant, les relations inter-zonales restent "symbiotiques" car les deux activités bénéficient mutuellement de leurs croissances.

L'augmentation du paramètre h tel que $\sigma_{11}-h < 0$ remplace la relation prédatrice par une géographie compétitive (pour le sol). Il y a transition d'une solution oscillatoire vers une solution exponentielle parfois dotée de lobes. La bifurcation dynamique suscitée s'apparente à une bifurcation de nature écologique. La structure urbaine d'équilibre restera spécialisée mais la dynamique de convergence devient plus constante. Les graphiques des figures 6.8a-c montrent des comportements évolutifs qualitativement différents selon 3 combinaisons de valeurs paramétriques. La figure 6.8a montre une trajectoire véritablement asymptotique qui s'apparente à un système comme (6.152-6.153); la figure 6.8b décrit une dynamique "prédatrice" sollicitant des variations cycliques déphasées; dans 6.7c, la compétition intra-zonale engendre une solution régulière. On remarquera la faible amplitude cyclique dans 6.7b due aux effets stabilisants des interactions inter-zonales toujours positives. Aussi, à l'approche de l'état d'équilibre, les oscillations s'estompent. Cette dernière propriété est partagée par tous les systèmes dynamiques convergents mais oscillants [Hirsch, Smale 1974; Bennett, Chorley 1978]. Dans tous les cas, ont été octroyées des valeurs initiales fortement distantes des états d'équilibre de façon à amplifier les propriétés convergentes des trajectoires. Un événement perturbateur apparaissant dans un temps intermédiaire susciterait aussi une réponse du système ayant les mêmes caractéristiques dynamiques de retour à l'équilibre [voir Cansela da Fonseca, Drach 1985].

Les trajectoires dynamiques exhibées par les modèles urbains inspirés du mécanisme de Lowry [1964] n'ont pas tous, comme on pourrait le croire, la même forme. Cette diversité doit par conséquent avoir des implications significatives sur la genèse de la structure. La présence d'une co-évolution cyclique suppose que les activités urbaines se développent par saccade de manière à s'ajuster à l'autre; le mécanisme d'ajustement reste

$$a) \quad g = 0 \quad h = 0$$

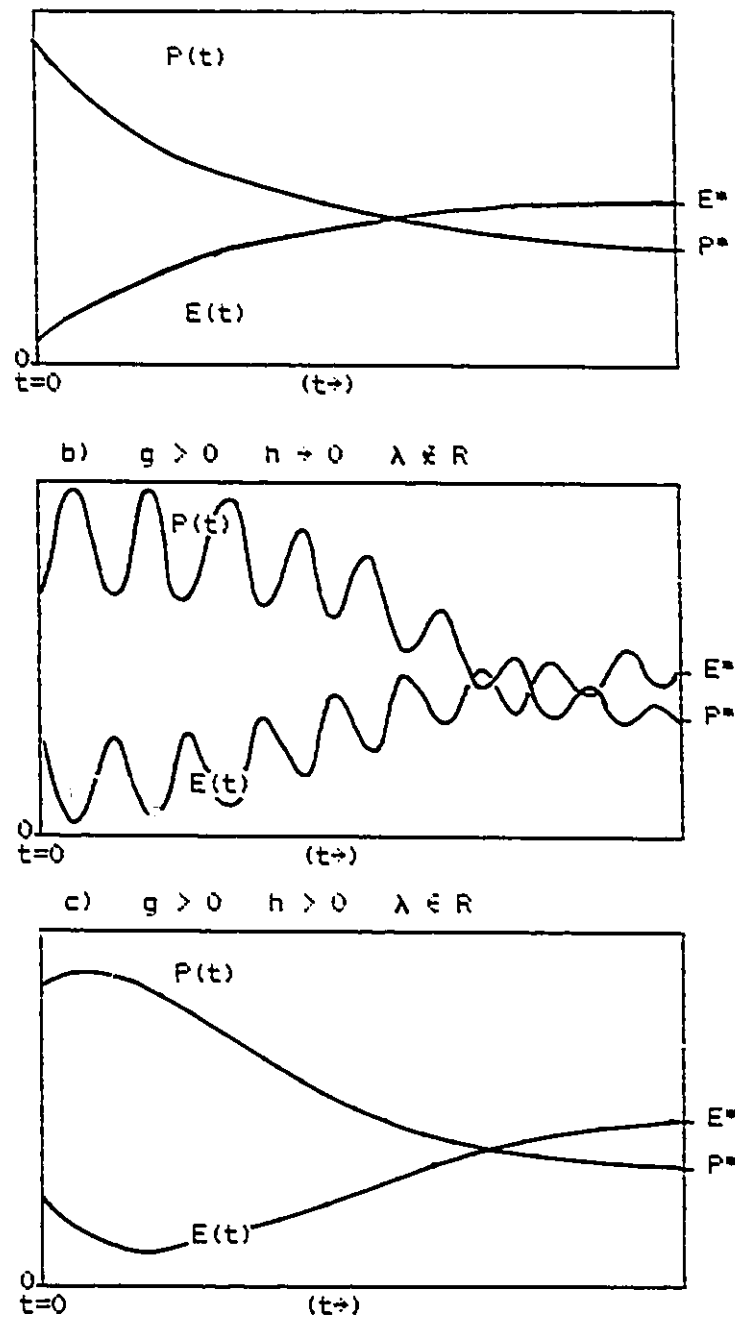


Figure 6.8: Trajectoire dynamique des variables d'emploi de service (S_i) et de population (P_i) tirées d'une même zone.

- a) sans compétition pour le sol
- b) activité commerciale «predatrice»
- c) compétition pour le sol

normal car la distribution spatiale d'une activité est intimement liée à celle des autres.

6.5.4 Notes concernant un modèle opérationnel

Les analyses dynamiques qui ont été effectuées dans les sections 6.5.2 et 6.5.3 portaient sur des versions simplifiées de modèles spatio-temporels de localisation d'activité. Cependant, les propriétés de catastrophe et de bifurcation décelées sont conservées dans une formulation plus élaborée comme la paire d'équations (6.132-6.133) ou (6.132, 6.134) dans laquelle des relations autoregressives temporelles et spatiales, des interactions spatiales et des relations fonctionnelles sont inscrites. Il en résulte toutefois des manifestations beaucoup moins claires et plus enchevêtrées.

Par ailleurs, certaines précautions doivent être prises dans l'attribution à des modèles dotés d'unités de temps discrètes de propriétés dynamiques détectées dans des systèmes d'équations différentielles. Dans le dernier cas, les progressions sont infinitésimales, de sorte que lorsque le système approche d'un point stable, il tend à rester immobile en absence de perturbation. Dans un système à temps discrets, il y a certes convergence vers un point d'équilibre stable mais l'ajustement se fait de façon moins précise. En effet, les réponses du système ne sont pas fondées sur l'état courant mais sur celui d'une période antécédente. Par conséquent, si le processus réagit avec beaucoup d'intensité, il risque d'outrepasser le point attracteur. A la période suivante, une correction s'ensuit mais, encore une fois, l'objectif est encore mal visé. Un comportement chaotique est alors observé. Au-delà un certain seuil de réponse, le système ne converge même plus car les oscillations discrètes délogent le système du bassin attracteur dans lequel il était. Un tel comportement se nomme «chaos de May-Chin-Li» et est propre aux systèmes dynamiques discrets [Dendrinos, Mullaly, 1985 chap.1].

Si l'intensité des réponses systémiques reste modérée, les propriétés dynamiques conventionnelles peuvent se manifester. Dans le modèle constitué des équations spatio-temporelles (6.132) et (6.134), le fait d'émettre une relation fonctionnelle d'économie d'échelle telle que l'attraction $W_j = f(S_j)$ suscitera une liaison non-linéaire chez la variable d'activité $\{S_j\}$.

Il est aussi vraisemblable que l'attraction résidentielle Q_i devienne une fonction quelconque de la population P_i . Dans ce cas, les possibilités de catastrophes sont moindres puisqu'il semble peu réaliste que l'attraction soit plus que proportionnelle à la population de par les limites de densité résidentielle qui tendent à restreindre l'accueil des résidents à mesure que la population zonale croît.

Puisque le modèle propose décrit un processus co-evolutif de deux activités, des bifurcations dynamiques sont envisageables. Dans la paire d'équations (6.132, 6.134), la distribution de l'activité exogène (vecteur e_z) détermine la distribution d'équilibre (voir 6.118, 6.120) si toutefois les matrices d'interaction Π et Σ ne comportent pas de relations non-linéaires entre p_t et entre s_t . La forme des $2N$ trajectoires dépend uniquement des relations inter-activités soit:

pour $s_t \rightarrow p_t$, la matrice composée $\left[A_z(z)I + A_n(z)\Pi \right]$ et

pour $p_t \rightarrow s_t$, la matrice composée $\left[B_z(z)I + B_n(z)\Sigma \right]$

Selon toute vraisemblance, les termes de $A_n(z)\Pi$ et de $B_n(z)\Sigma$ seront positifs car ils dénotent les interactions spatiales favorables à la localisation de l'autre activité. Analogiquement aux scalaires g et h de la section 6.5.3, les matrices diagonales $A_z(z)I$ et $B_z(z)I$ traitent exclusivement de relations inter-activités intra-zonales. Les signes des polynômes $A_z(z)$ $B_z(z)$ sont sans restriction et définissent le type de relations co-évolutives entre l'activité résidentielle et commerciale dans une zone.

Les nombreuses relations autorégressives temporelles (spatiales et non-spatiales) décident, quant à elles, de la vitesse de convergence. Une forte mémorisation indique une inertie du système, freinant alors l'évolution vers l'équilibre. L'intensité et le signe des autorégressions spatiales conduisent soit à une uniformisation ou une différenciation géographique du développement. Dans les deux cas, elles nuisent à la progression systémique vers l'état d'équilibre i.e. l'état des bénéfices maximaux de localisation.

6.6 CONCILIATION DES MODELES DE TRANSPORT ET DE LOCALISATION

6.6.1 Cadre conceptuel

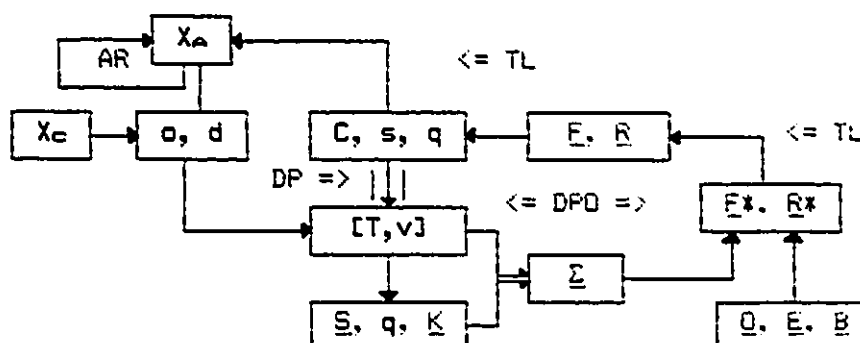
Tout au long des chapitres 3, 4 et 6, on a tenté de démontrer la compatibilité de certains principes d'optimisation. Les modèles d'offre et de demande de transport peuvent être dérivés et intégrés en maximisant des coûts généralisés de transport. La maximisation des bénéfices de localisation permet de formuler des modèles consistants de localisation d'activités urbaines; dans ce cas-ci, un cadre dynamique apparaît applicable.

Jusqu'ici, deux intégrations de modèles ont été effectuées. Toutefois, comme il a été dit précédemment, une intégration complète s'avère difficile puisque les processus à concilier agissent sur des périodes de magnitudes différentes. Le comportement des voyageurs est généralement guidé par des conditions actuelles: distribution spatiale des activités, géométrie des réseaux, trafic, tarification, etc... Celui des fournisseurs de l'offre de transport résulte en grande partie d'une appréciation d'un système existant. Néanmoins, la période d'intervention peut être plus ou moins longue, dépendant du type de modification apportée. Le processus de localisation d'activités s'exerce, comme on sait, à moyen et long terme. La permanence des établissements assure une apparente immuabilité des activités.

En raison de ces divergences d'échelle temporelle, il semble préférable et plus pratique, dans un contexte de planification, de "mettre en interface" plutôt qu'intégrer les cadres de modélisation des systèmes de transport et de localisation d'activités.

Dans le sous-chapitre 4.1, un schéma conceptuel de l'offre de transport a été élaboré à partir de ceux suggérés par Florian et Gaudry [1980]. L'intérêt de ce schéma est qu'il assimile les éléments de demande en tant que mesure du système. En surcroît, il s'accommode de deux niveaux d'équilibre, l'un «instantané» à travers l'équilibre du trafic sur le réseau (principe de Wardrop), l'autre à moyen ou à long terme entre la demande générale et des modifications des systèmes de transport.

Dans ce schéma, le processus de localisation précède la génération spatiale des déplacements (case $[X, o, d]$). Désormais, cette case se scinde en deux parties, la matrice X contenant les facteurs économiques zonaux se séparant des vecteurs d'origines et de destinations o et d . Procédons maintenant à une partition des facteurs de la matrice X de manière à isoler les caractéristiques socio-économiques de la population des variables d'activités. Le premier groupe de facteurs est dorénavant considéré comme exogène (nommons cette sous-matrice X_C). La sous-matrice restante, disons X_A , contient des variables spatialisées de fonctions urbaines (population, emplois, étudiants, etc...). Dans le schéma, la case qui s'y identifie est authentiquement endogène puisque ces variables sont tributaires des distributions spatiales antérieures, de certaines activités exogènes et des conditions antérieures de transport en état soutenu. Il s'agit d'une nouvelle boucle d'équilibre, l'équilibre «transport - localisation». Le schéma suivant illustre le cadre conceptuel suggéré ici.



DP: équilibre demande - performance

DPD: équilibre demande - performance - offre

TL: équilibre transport - localisation

AR: effets autorégressifs spatio-temporels des activités

Etant essentiellement une conceptualisation générale, on s'empressera ici de définir un cadre dit "opérationnel" qui traduira les cases schématiques en des phases d'application de modèles. Dans chacune de ces phases, on retrouve un modèle complètement intégré. Entre chacune d'elles, les interdépendances se manifestent par des variables communes, produites par l'une et requises dans l'autre, et des principes d'optimisation consistants.

6.6.2 Cadre opérationnel

Le schéma illustré dans la figure 6.2 peut incidemment être reformulé par un «cadre opérationnel». Celui-ci consiste en quatre phases distinctes qui traduisent respectivement une relation linéaire et la réalisation de trois équilibres.

Prenons comme conditions initiales une distribution spatiale d'activités X_A et certains attributs de la population X_C au temps t . Le cycle débute par une génération zone par zone des déplacements produits et attirés. La phase de génération spatiale est interprétée par la paire de fonctions (3.1-3.2) ou (3.7-3.8). Elle est mise à l'écart des autres modèles de désagrégation de la demande parce qu'elle ne partage pas la même rationalité; les origines et les destinations sont essentiellement conditionnées par les activités et les caractéristiques socio-économiques mais très peu par les éléments relatifs au transport proprement dit. De cette phase sont issus les vecteurs o et d .

La phase suivante consiste à distribuer simultanément cette demande selon les trajets origines-destinations, les modes disponibles et leurs itinéraires admissibles. Il s'agit de réaliser l'équilibre «demande - performance» en solutionnant le programme de minimisation (3.144-3.151). A cette étape, le réseau $L(1)$ des routes est identifié par un vecteur q de capacité et un vecteur s de fonctions de congestion. Par hypothèse, le réseau $L(2)$ de lignes de transport en commun possède des capacités illimitées et n'est conséquemment défini que par un vecteur s de coûts constants. L'équilibre «demande - performance» permet d'évaluer les matrices spatiales (O-D) de déplacement T et de coût de transport associé C ; ces matrices peuvent être désagrégées par mode. A l'instar, le vecteurs de flux v et de coûts sur réseau $s(v)$ sont ajustés pour le réseau routier $L(1)$.

La demande étant mesurée pour une période " t ", la phase suivante consiste à sélectionner les modifications de réseau permettant d'améliorer le plus les conditions de déplacement. L'équilibre «demande - performance - offre» est représenté par le problème de conception de réseau avec demande variable (4.29-4.41). En fait, ce problème ne diffère du précédent que par sa fonction objective traitant uniquement des liens de réseau; la modéli-

sation de la demande devient alors un sous-problème.

L'équilibre DPQ ne survient qu'à moyen et long terme. La modification des réseaux dictée par le modèle se réalise donc avec un décalage qui se mesure en année. Ainsi, l'optimum DPQ n'est pas nécessairement obtenu puisque ce décalage suffit pour que la distribution migratoire soit substantiellement changée.

Les conditions de transport renouvelées sont interprétées par une matrice de coûts composées. Celle-ci alimente un système d'équations comme (6.132, 6.134) qui relocalise une part des activités en tenant compte des distributions spatiales antérieures. Cet équilibre «transport - localisation» n'est pratiquement jamais atteint car ses temps des réaction paraissent lents.

La boucle se ferme quant la nouvelle distribution d'activités se met à générer des déplacements: le système reste néanmoins en déséquilibre puisqu'il surviendra des changements répétés de réseaux et de structure urbaine.

Les problèmes d'optimisation utilisés dans la représentation des équilibres sont "génériques". Ils peuvent être reformulés différemment selon la nature de l'application et la disponibilité des données. Par exemple, le problème de demande de transport peut s'accommoder d'une affectation routière "tout-ou-rien" pour l'un des modes. La définition des coûts modaux peut tenir compte de caractéristiques "hors-routes". Il n'y a aussi aucune obligation quant à la variabilité de la demande et la capacité limitée des liens dans un problème de conception de réseau; l'intégration conceptuelle y perdrait mais non l'applicabilité. Enfin, on laisse à la discrétion de l'utilisateur, la catégorisation des activités et le choix des relations spatio-temporelles à retenir.

6.6.3 Les implications d'une conciliation

En principe, le modèle général permettrait d'évaluer la demande de transport urbain à partir d'une organisation urbaine donnée et des conditions de transport courantes. On obtient alors l'équilibre «demande -

performance». Il dicterait également les améliorations de réseaux de transport les plus appropriées, considérant le patron migratoire observé et le budget de développement accordé. Cet ajustement progressif de l'offre s'assimile à l'équilibre «demande - performance - offre» bien qu'il n'est pas sûr qu'un véritable équilibre puisse s'établir. On assisterait plutôt à une succession de correctifs permettant de conformer les réseaux à la demande sans parvenir à un état stable et satisfaisant.

Plusieurs facteurs empêchent l'obtention de l'équilibre DFO. D'abord, il y a le décalage entre l'appréciation du système de transport et les interventions. La limite budgétaire peut aussi ne pas permettre de réaliser les améliorations idéales tout comme le sous-ensemble des liens projetés peut omettre les projets les plus prometteurs. Par ailleurs, du fait que la plupart des modifications de réseaux sont discrètes (ajout d'une route ou d'une ligne d'autobus, élargissement de la chaussée, correction géométrique d'un carrefour), on ne peut que rester en deça ou dépasser les réels besoins. Il faut savoir que tout changement local d'un réseau affecte l'entièreté du système de transport; réorganisation des itinéraires, transfert modal, redistribution origine-destination. Dans le cas de l'automobile, l'existence de la congestion amplifie les rajustements de la demande puisque que les conditions de déplacement apparaissent "flot-tantes".

Pourtant, tous ces facteurs sont appréhendés par le problème de conception de réseau car il évalue d'avance la demande conséquente aux changements envisagés. En fait, les conditions d'optimalité du problème garantissent des améliorations optimales compte tenu du budget alloué et de la définition des liens projetés. Si on ne doute pas de l'optimalité d'une solution, on ne peut quand même pas affirmer de facto que celle-ci reflète un état d'équilibre stable d'un point de vue dynamique. En effet, rien ne nous assure que les améliorations apportées ne deviennent pas caduques quelques temps après.

Pourquoi en serait-il ainsi? Parce que l'agglomération urbaine croît globalement et que sa structure spatiale interne évolue différentiellement. Si bien que les facteurs générant la demande spatiale change continuellement. La croissance économique et démographique de la région dépend essen-

tiellement de conditions exogènes. Par contre, la distribution interne de cette croissance est déterminée par le potentiel (amenités, possibilité de développement) de chaque zone de la ville et par leur accessibilité relative. Or, l'expansion des réseaux de transport et les variations spatiales de la congestion (causée par celles de la demande) modifient continuellement le patron d'accessibilité. Mais la structure urbaine n'en est pas chambardée pour autant. Caractérisée avant tout par sa permanence, elle répond aux changements d'accessibilité qu'avec retardement. De surcroît, ses réponses sont discrètes plutôt que continues et restent localisées.

La conciliation d'un modèle de transport avec celui de localisation vise à reproduire ce phénomène d'interaction entre le système de transport et la structure urbaine. Dans notre modélisation, tous les éléments évoqués précédemment semblent pris en compte. La croissance démographique est exprimée à travers la variable d'activité exogène tandis que le potentiel de chaque zone est inscrit dans un terme d'attraction (distinct pour les services et les résidences). L'accessibilité s'évalue à partir de la matrice de coût composé de transport. La structure autorégressive du modèle de localisation autorise à la fois la permanence des établissements et les délais de réaction. Seuls les changements discrets ne peuvent être appréhendés mais l'emploi d'un système zonal agrégé et d'unités de temps suffisamment longues permettent d'assimiler ces changements à une évolution continue.

L'intérêt d'un modèle spatio-dynamique dans un contexte de planification est de permettre une meilleure évaluation des impacts du développement du transport urbain. Les impacts se situent à deux niveaux. D'une part, il y a les conséquences sur la structure urbaine comme l'étalement résidentiel, la spécialisation spatiale, la concentration ponctuelle des services et des commerces, l'émergence de nouvelles zones et le déclin des anciennes. Ces phénomènes sont déjà observables dans nombre de métropoles et s'expliquent souvent, en partie comme en tout, par des transformations antérieures du système urbain de transport.

D'autre part, il y a les impacts sur l'offre et la demande de transport. Ceux qui surviennent immédiatement semblent facilement mesurables au moyen de modèles statiques de transport. Les répercussions à plus long

terme paraissent évidemment plus difficiles à prévoir car beaucoup de facteurs échappent à l'analyste. Les changements de comportement des voyageurs (exprimés par des paramètres) demeurent peu prévisibles de même que l'évolution de certains coûts (comme celui du carburant). Par contre, le modèle que nous présentons devrait apprendre les changements intra-urbains résultants des interventions en matière de transport. Ces changements ont des répercussions si importantes sur l'offre et la demande de déplacement qu'ils compromettent même les bénéfices attendus des améliorations du transport.

Deux mesures de transport paraissent particulièrement sensibles à l'organisation urbaine. Citons d'abord les parts modales. Par exemple, l'équation "bonification du réseau => gain de clientèle" n'est pas aussi simple. Généralement, l'étalement urbain et l'éclatement des origines et des destinations défavorisent les services de transport collectif, lesquels sont plus adaptés à des tissus urbains compacts et des déplacements concentriques. Mentionnons ensuite les coûts moyens de transport. Dans des conditions d'état soutenu, tout aménagement de réseau de transport devrait faire diminuer les coûts de transport sinon au moins les maintenir. Mais une agglomération en expansion spatiale imposera un allongement des parcours. A la fois, elle privilégiera l'usage de l'auto et, par le fait même, entraînera une hausse de la congestion routière. Des parcours allongés sollicitent davantage d'éléments routiers et impliquent alors une surcharge additionnelle. Par ailleurs, l'expansion démographique et la croissance économique (activant la mobilité des biens et des personnes) de la ville font sans cesse le flot de voyageurs sur les réseaux de transport et infligent des conditions de déplacement de plus en plus pénibles.

Les deux chapitres suivants, lesquels constituent la troisième partie, seront consacrés à l'application du cadre de modélisation présenté précédemment et développé tout au long des chapitres antérieurs. Le premier des deux chapitres est consacré à l'opérationnalisation i.e. à l'ensemble des étapes permettant l'utilisation du modèle à des fins empiriques. Quant au chapitre huit, il présentera les analyses et commentaires découlant d'une application à la région de Montréal.

6.A APPENDICE A: Equilibre statique des modèles de localisation spatio-temporels

Le premier des deux modèles est forme de la paire d'équations (6.132-6.133). Il possède deux composantes distinctes, l'une autogressive, l'autre d'interactions spatiales, soit:

$$p_t = [\Omega_x(z)I + \Omega_w(z)W]p_t + [A_x(z)I + A_n(z)\Pi](s_t + e_t) \quad (6A.1)$$

$$s_t = [\Gamma_x(z)I + \Gamma_o(z)Q]s_t + [B_x(z)I + B_x(z)\Sigma]p_t \quad (6A.2)$$

Les matrices W et Q évoquent les structures spatiales à travers lesquelles des effets d'autocorrélation sont transmis. Les matrices Π et Σ contiennent des éléments d'interaction spatiale entre les activités résidentielles et commerciales. La matrice I est celle d'identité de même taille que les précédentes (N·N).

Dans des conditions d'état soutenu où les vecteurs exogènes e_t sont constants pour tout $t=0,1,2,\dots$, l'opérateur de décalage z devient sans pertinence et les vecteurs endogènes p_t et s_t sont constants pour tout t. Alors, les polynômes de décalage se resument à des scalaires résultant de la somme des coefficients:

$$\begin{aligned} \text{si } z \equiv 1, \quad \underline{\Omega}_x &\equiv \Omega_x(z), \quad \underline{\Omega}_w \equiv \Omega_w(z), \quad \underline{A}_x \equiv A_x(z), \quad \underline{A}_n \equiv A_n(z), \\ \underline{\Gamma}_x &\equiv \Gamma_x(z), \quad \underline{\Gamma}_o \equiv \Gamma_o(z), \quad \underline{B}_x \equiv B_x(z), \quad \underline{B}_x \equiv B_x(z) \end{aligned}$$

On peut alors simplifier la représentation matricielle en créant de nouvelles matrices:

$$\Omega = \underline{\Omega}_x I + \underline{\Omega}_w W \quad A = \underline{A}_x I + \underline{A}_n \Pi \quad (6A.3-6A.4)$$

$$\Gamma = \underline{\Gamma}_x I + \underline{\Gamma}_o Q \quad B = \underline{B}_x I + \underline{B}_x \Sigma \quad (6A.5-6A.6)$$

Conséquemment, (6A.1-6A.2) deviennent;

$$p = \Omega p + A s + A e^s \quad (6A.7)$$

$$s = \Gamma s + B p \quad (6A.8)$$

Cette forme s'apparente beaucoup à celle présentée dans la section 6.3.5 (eqs. 6.102-6.103), à l'exception des matrices d'autocorrélation Ω et Γ . Ces composantes peuvent cependant être isolées en les convertissant en multiplicateurs spatiaux:

$$\begin{aligned} p - \Omega p &= As + Ae^0 \\ p &= [I - \Omega]^{-1} As + [I - \Omega]^{-1} Ae^0 \end{aligned} \quad (6A.9)$$

$$\begin{aligned} s - rs &= Bp \\ s &= [I - r]^{-1} Bp \end{aligned} \quad (6A.10)$$

Cette reformulation effectue la convolution issue de l'autocorrelation spatiale. Une fois ces effets isolés, (6A.7) et (6A.8) peuvent se noter comme suit:

$$p = \bar{A}s + \bar{A}e^0 \quad \text{où } \bar{A} = [I - \Omega]^{-1}A \quad (6A.11)$$

$$s = \bar{B}p \quad \text{où } \bar{B} = [I - r]^{-1}B \quad (6A.12)$$

Dès lors, la démonstration effectuée dans la section 6.3.5 peut être refaite de la même manière puisque (6A.11) et (6A.12) ont exactement la même forme que (6.102) et (6.103):

$$\begin{aligned} s &= \bar{B}p = \bar{B}\bar{A}s + \bar{B}\bar{A}e^0 \\ s - \bar{B}\bar{A}s &= \bar{B}\bar{A}e^0 \\ s &= [I - \bar{B}\bar{A}]^{-1} \bar{B}\bar{A}e^0 \end{aligned} \quad (6A.13)$$

Le multiplicateur pour p est dérivé à partir de (6A.13):

$$\begin{aligned} p &= \bar{A}s + \bar{A}e^0 = [I - \bar{B}\bar{A}]^{-1} \bar{B}\bar{A}e^0 + \bar{A}e^0 \\ p &= \bar{A}[I - \bar{B}\bar{A}]^{-1} e^0 \end{aligned} \quad (6A.14)$$

(6A.13) et (6A.14) sont les équations d'équilibre pour le modèle (6A.1-6A.2); dans leurs formes compactes, ils sont isomorphes à celle décrivant les conditions d'optimalités (6.107-6.108) du programme G (6.92-6.97). Maintenant, revenons aux composantes initiales:

$$\begin{aligned} p &= [I - \Omega]^{-1}A \left[I - [I - r]^{-1}B[I - \Omega]^{-1}A \right]^{-1} e^0 \\ s &= [I - [I - r]^{-1}B[I - \Omega]^{-1}A]^{-1} [I - r]^{-1}B[I - \Omega]^{-1}A e^0 \end{aligned}$$

où Ω , A , r et B définies par (6A.3-6A.6).

La détermination de l'état d'équilibre du modèle composé par les équations (6.132) et (6.134) ne présente pas de difficulté supplémentaire même un nouvel effet d'interaction s'exerce au sein du vecteur commercial. Réécrivons d'abord l'équation (6.134):

$$s_t = \left[r_t(z)I + r_a(z)Q \right] s_t + \left[B_t(z)I + B_x(z)\Sigma \right] p_t + \left[F_t(z)I + F_x(z)\tilde{Q} \right] (s_t + e_t^p) \quad (6A.15)$$

Le troisième terme désigne les interactions de l'emploi total vers l'emploi commercial. Puisque dans des conditions d'état soutenu, $e_t^p = e^p$, pour tout t , les polynômes de décalage $F_t(z)$ et $F_x(z)$ se réduisent aux scalaires $\underline{F}_t$ et $\underline{F}_x$ de même que tous les autres polynômes de (6A.15) et (6A.1). Créons une matrice F tel que:

$$F = \underline{F}_t I + \underline{F}_x \tilde{Q} \quad (6A.16)$$

Alors (6A.15) devient:

$$\begin{aligned} s &= r s + B p + F s + F e^p \\ s - r s &= B p + F s + F e^p \\ s &= [I - r]^{-1} B p + [I - r]^{-1} F s + [I - r]^{-1} F e^p \end{aligned} \quad (6A.17)$$

Faisons l'équivalence:

$$F = [I - r]^{-1} F$$

Alors, selon les définitions précédentes de p (6A.11) et de $\bar{B}$ (6A.12):

$$\begin{aligned} s &= B p + F s + F e^p = \bar{B} \bar{A} s + \bar{B} e^p + F s + F e^p \\ s - \bar{B} \bar{A} s - F s &= \bar{B} \bar{A} e^p + F e^p \\ s &= [I - (\bar{B} \bar{A} + F)]^{-1} [\bar{B} \bar{A} + F] e^p \end{aligned} \quad (6A.18)$$

Le lecteur a noté un déroulement similaire à celui conduisant à (6.118). De même, le vecteur d'équilibre p est dérivé à partir de (6A.18) et son équation est isomorphe à (6.120):

$$p = \bar{A} [I - (\bar{B} \bar{A} + F)]^{-1} e^p \quad (6A.19)$$

Avec les matrices initiales (6A.3-6A.6) et (6A.16), on a alors:

$$p = \left[[I - Q]^{-1} A \right] \left[I - [I - r]^{-1} B [I - Q]^{-1} A - [I - r]^{-1} F \right]^{-1} e^p \quad (6A.20)$$

$$s = \left[I - [I - r]^{-1} B [I - Q]^{-1} A - [I - r]^{-1} F \right]^{-1} \left[[I - r]^{-1} B [I - Q]^{-1} A + [I - r]^{-1} F \right] e^p \quad (6A.21)$$

Il faut se rappeler que les équations d'équilibre statiques valent pour un vecteur e^p constant, des polynômes de décalage stationnaires et des matrices spatiales W , Q , Π et Σ constantes et exogènes.

6.B APPENDICE B: Solution d'un système dynamique approximant

La solution d'un système dynamique approximant doté d'une matrice Jacobienne A:

$$\frac{dx}{dt} = Ax \quad \text{où } A = \{a_{ij}\}_{N \times N}, \quad a_{ij} = \left. \frac{\partial F_i}{\partial x_j} \right|_{x^*} \quad (6B.1)$$

est trouvée en diagonalisant A à l'aide de la transformation linéaire $V^{-1}AV$ où:

$$V_{N \times N} = \{v_{ij}\} = [x_1, x_2 \dots x_N]$$

dans lequel x_i sont les N vecteurs invariants associés à chacune des racines λ_i du polynôme caractéristique:

$$|A - \lambda I| = 0$$

Les vecteurs invariants (ou vecteurs propres) sont évalués en solutionnant:

$$[A - \lambda I] \cdot x = 0$$

Une telle transformation linéaire permet d'obtenir une matrice diagonale formée des N racines propres λ_i :

$$V^{-1}AV = \text{diag}(\lambda_1, \lambda_2 \dots \lambda_N) \quad (6B.2)$$

Maintenant, supposons le système transformé:

$$\frac{dz}{dt} = Bz \quad \text{où } \frac{dz}{dt} = V^{-1} \frac{dx}{dt} \quad \text{et } z = V^{-1}x \quad (6B.3)$$

La pré-multiplication par V donne:

$$\frac{dx}{dt} = VBz = VB V^{-1}x \quad (6B.4)$$

Alors,

$$VBV^{-1} = A$$

En pré-multipliant V^{-1} , on retrouve le système transformé (6B.3) défini par la matrice de transformation linéaire (6B.2):

$$\frac{dz}{dt} = V^{-1}AVz \quad (6B.5)$$

D'après (6B.2), le système présente la forme canonique suivante:

$$\frac{dz_i}{dt} = \lambda_i z_i \quad (6B.6)$$

Par intégration.

$$z_1 = K_1 \exp(\lambda_1 t) \quad (6B.7)$$

où K_1 est la constante d'intégration relative aux conditions initiales.

Revenons maintenant à x . Selon les définitions en (6B.3) et (6B.7):

$$x_1 = \sum_j v_{1j} z_j = \sum_j v_{1j} K_j \exp(\lambda_j t) \quad (6B.8)$$

La solution d'un système dynamique multivarie ou multispacial prend donc la forme générale suivante:

$$x = G \cdot \exp(\lambda t) \quad (6B.9)$$

$$G_{N \times N} = \{v_{1j} \cdot K_j\} \quad \exp(\lambda t)_{N \times 1} = \{\exp(\lambda_1 t)\}$$

Troisième partie

UNE APPLICATION DU CADRE DE MODELISATION

Chapitre 7

OPERATIONNALISATION DU CADRE DE MODELISATION

7.1 INTRODUCTION A L'APPLICATION

7.1.1 Objectifs

Le cadre de modelisation qui a été presente dans cette these apparait certes un cadre ideal. D'abord, le (sous-) modele de demande considere un trafic d'equilibre pour les automobiles et un partage modal avec le transport en commun. Ensuite, selon les conditions de transport et les budgets de developpement, les reseaux peuvent prendre de l'expansion. En surcroit, la distribution spatiale des activites se modifie, quoique lentement, selon les changements de l'accessibilite relative aux autres activites.

Ce cadre emanant avant tout d'une reflexion theorique, il convient de l'appliquer a une region urbaine afin d'en apprecier la portee dans un contexte de planification. En fait, le principal objectif de l'application consiste a definir le role d'une telle modelisation comme instrument d'analyse. Dans le cadre de la planification des transports urbains, il s'agit de simuler des expansions alternatives de reseaux en conjonction avec le developpement urbain consequent.

Mais l'application a une region comme celle de Montreal semble plus qu'un bon pretexte et vise plutot a eprouver notre methodologie. Recourir a une region hypothetique aurait ete beaucoup plus simple et aurait surement permis de deceler certaines faiblesses de la modelisation. Malheureusement, un cas hypothetique engendre par l'auteur du modele risque d'etre trop adapte a la methodologie si bien que celle-ci afficherait artificiellement une trop bonne performance.

Une vraie region d'etude avec toute sa complexite et ses contradictions exige d'un modele une grande souplesse. Le decoupage geographique y est heterogene, les comportements y apparaissent inconstants et les donnees s'avèrent imparfaites sinon incompletes. Pourtant, ces conditions sont normales si bien qu'un modele, aussi elegant qu'il peut etre, doit d'abord venir à bout des "imperfections" des problemes d'aménagement.

Pour évaluer le modèle dit spatio-temporel, on a choisi d'étudier ses possibilités de simulation conjointe de la structure urbaine et des réseaux de transport. Le terme simulation se distingue de prévision au sens qu'il signifie une reproduction d'une évolution possible à partir de différentes conditions prédéfinies. La simulation génère certes des prévisions mais elle sert plutôt à évaluer les conséquences de ces conditions formulées par le chercheur. Quant à analyse strictement prévisionnelle, elle vise précisément à obtenir des informations sur l'évolution d'un phénomène dont on connaît le comportement passé.

La complexité du problème étudié ne permet pas de prétendre à des prévisions. D'abord, l'évolution des réseaux de transports est entièrement soumise aux politiques de transport. Celles-ci obéissant à toutes sortes d'imperatifs souvent contradictoires et inconstants, il serait périlleux de prédire le développement probable des réseaux. Cependant, il reste permis d'analyser certaines politiques-types afin d'établir la fourchette des développements conséquents. Dans ce sens, il s'agit d'une simulation.

La validation du modèle par une analyse rétrospective a été par ailleurs écartée. L'indisponibilité d'une base cohérente de données ne permet même pas de retracer adéquatement l'évolution antérieure. La section 7.2.2 donnera davantage de motifs d'abandon en rapport avec les données observées.

Les développements simulés seront auscultés sous divers éclairages. A l'égard de la demande de transport, il faut se demander si les nombreux paramètres décrivant la perception de la distance ne deviennent pas caduques lorsque les réseaux de transport et la distribution des déplacements sont modifiés. En effet, il est possible que la sensibilité des paramètres conviennent à une organisation spatiale donnée et se transfère mal à une autre. A propos de l'amélioration des réseaux, on s'interrogera sur la pertinence de minimiser seulement les coûts de transport. Aussi, est-il possible que la confrontation de deux modes mettent en évidence l'incomparabilité des coûts perçus de transport et des budgets d'améliorations? Quant au processus de localisation des activités, est-il plus sensible aux disparités spatiales des facteurs attractifs qu'aux variations de l'accessibilité. Si tel est le cas, on met en doute le rôle prédictif

des modèles de localisation tout en diminuant le rôle déterminant des politiques de transport urbain.

Les résultats contre-intuitifs deviennent un autre élément d'analyse de la performance du modèle. Une calibration inadéquate peut expliquer une sensibilité excessive ou insuffisante des sous-modèles. Mais la plus grande source de résultats contre-intuitifs reste sans doute le traitement simultané de plusieurs phénomènes. Même si les modèles semblent bien intégrés du point de vue théorique, plusieurs relations s'enchevêtrent si bien que certaines imprecisions risquent de se multiplier.

Dans les sous-chapitres 7.2 et 7.3, il sera question de rendre opérationnel le modèle intégré. Dans un premier temps, il s'agit de déterminer le territoire d'étude, son découpage, l'horizon de simulation, la périodicité. Ensuite, il convient d'identifier les données observables qui interpréteront les variables génériques du modèle. Enfin, on procédera à l'estimation des paramètres nécessaires à la simulation. Cette phase d'opérationnalisation est une phase critique du modèle. Elle permet d'apprécier son applicabilité non seulement dans le contexte spécifique de la région montréalaise mais aussi dans le contexte plus général des problèmes associant transport et développement urbain.

7.1.2 La région métropolitaine de Montréal comme terrain d'essai

Pour faire ressortir les propriétés exclusives du cadre de modélisation proposé, il convient de l'appliquer à une agglomération urbaine dont le développement du transport promet plusieurs alternatives. La région métropolitaine de Montréal a été choisie parce que, spécifiquement, on y trouve de nombreuses possibilités d'expansion des réseaux de transport urbain. Le système autoroutier, le plus déployé dans une ville canadienne, n'est toutefois pas encore complètement connecté. Son système de métro souterrain connaît une extension considérable depuis son ouverture en 1966 et il paraît fort possible que cela puisse continuer. Dans le cadre de cette application empirique, il importe qu'un large éventail d'options de transport soit présenté. Sinon, le (sous-) modèle de conception de réseau n'y trouve plus son utilité, l'ensemble des liens projetés (L^p) devenant si petit que les liens optimaux peuvent être sélectionnés à l'aide d'une

enumeration complete.

Le cas de Montreal apparait d'autant plus interessant du fait qu'aucune alternative de developpement du transport semble se detacher du peloton. Aujourd'hui encore, les multiples niveaux de gouvernement et organismes concernes ne s'entendent pas sur les priorites du developpement du transport urbain. Les enjeux sont de taille car il implique des coûts énormes d'immobilisation, des options d'aménagement du territoire, des choix modaux et une conciliation des besoins des transporteurs de marchandises.

La métropole quebecoise soulève aussi de l'intérêt par sa dynamique intra-urbaine disparate. La ville centrale a perdu de 1971 à 1986 plus de vingt pourcent de ses residents. Les couronnes suburbaines ont beaucoup beneficie de l'etalement de l'habitat bien que les banlieues moins lointaines semblent subir un sort davantage semblable à leur grande voisine. Cette dispersion de l'activite résidentielle n'est pas qu'attribuable à l'expansion considerable du réseau autoroutier. Elle est concomitante au decloisement géographique de l'emploi, particulièrement l'emploi manufacturier. Certes, le centre-ville montréalais reste le pôle économique principal mais les unités de production et les services à la consommation tendent de plus en plus à occuper le territoire métropolitain.

L'attention particulière dont bénéficie la région de Montréal dans les dossiers politiques récents apportent coïncidemment une motivation supplémentaire. La proposition de l'«Option préférable d'aménagement» [O.P.D.Q., 1987], les positions de la Communauté Urbaine de Montréal sur les transports [C.U.M., 1988], le «Plan d'action» du Ministère des transports [M.T.Q., 1988] témoignent de l'actualité du thème "transport -développement urbain". De plus, le dossier brûlant du sort de l'autoroute Métropolitaine, l'expansion incertaine du métro dans l'Est, les problèmes de transbordement du transport en commun ravivent la curiosité de tous envers le transport urbain.

7.2 DEFINITION ET SPECIFICATION DE LA BASE DE DONNEES

7.2.1 Système spatial

Dans une étude de transport urbain, il n'y a pas de système spatial parfait mais seulement de compromis. Avec un système finement désagrégé, on obtient des mesures de coût de transport, des affectations sur les réseaux et des répartitions modales plus précises. Par contre, la taille des matrices de coût, des matrices de déplacement et des listes d'itinéraires croît géométriquement avec le nombre de zones. Une aggrégation spatiale excessive occulte cependant de nombreuses caractéristiques de transport puisqu'un nombre trop important de voyages s'effectue à l'intérieur des zones (voir Aveni [1979] pour une évaluation de l'aggrégation spatiale à travers la proportion des déplacements intra-zonaux). Par surcroît, la distribution des activités urbaines devient artificiellement homogène dans un système de grandes zones.

Nous avons construit un système de 60 zones qui couvre la région métropolitaine de Montréal selon une définition de la STCUM [1983]. La région comprend, outre les îles de Montréal et Jésus, une Rive Sud qui s'étend de Lery à Vercheres allant jusqu'à la rivière Richelieu et une Rive Nord partant d'Oka jusqu'à St-Sulpice. Elle contient également la MRC de Vaudreuil-Soulanges. Le découpage territorial proposé par la STCUM reste à peu près inchangé de façon à pouvoir utiliser les données de transport fournies par cette organisme. Certaines zones ont été fusionnées car jugées trop petites tandis que d'autres ont été scindées en raison de leurs territoires disloqués (nuisant à la représentativité de leurs centroïdes). Les fusions ne posent aucun problème quant aux données issues de la STCUM puisqu'elles n'impliquent qu'une aggrégation des valeurs. Dans le cas des zones divisées, les données de transport n'ont été que réparties équitablement.

La carte 7.1, présentée à la page 298, montre le territoire d'étude et son découpage. Les zones de même que leurs parties constituantes sont identifiées dans le tableau de l'annexe 7.A (pp 299-300). La carte principale présente 48 zones de la Communauté Urbaine de Montréal, de la ville de Laval et de la proche Rive Sud. Dans de nombreux cas, chaque zone coïncide

avec une ou deux entités municipales. Toutefois, les villes de Montréal et de Laval, trop grandes, sont respectivement subdivisées en 15 et 8 zones. Pour la ville de Montréal, la définition zonale s'harmonise généralement avec celles des «arrondissements de planification» sans pour autant correspondre totalement. Fait intéressant, le centre-ville s'étend sur deux zones, l'une restreinte à son cœur (de St-Mathieu à St-Hubert et de Sherbrooke au Vieux-Montréal), l'autre couvrant sa périphérie immédiate (d'Atwater à Papineau, de la montagne au fleuve). Quant à la ville de Laval, les zones s'accordent avec le découpage usuel des grands quartiers.

Le carton présenté au bas de la carte 7.1 illustre les douze zones les plus périphériques des couronnes Nord et Sud. Ces divisions correspondent généralement à des «municipalités régionales de comté» bien que nous avons procédé à certains regroupements logiques de municipalités lorsque les territoires définis par la STCUM nous semblaient trop étendus. Les zones périphériques se distinguent nettement des autres par leur faible densité d'occupation du sol. Ceci explique leur niveau d'agrégation spatiale plus élevé. De par l'importance du territoire, la périphérie de l'agglomération montréalaise apparaît néanmoins non-négligeable.

7.2.2 Horizon temporel et périodes

La difficulté d'obtention des données antérieures empêche toute analyse rétrospective. Seules des données concernant les caractéristiques de la population sont disponibles selon une base spatiale suffisamment désagrégée et une base périodique régulière (i.e. quinquennale). Quant aux données relatives aux activités non-résidentielles, elles se révèlent plutôt ponctuelles tant géographiquement que temporellement. Les seules variables convenables, des variables d'emploi, n'étaient compilées que pour 1971 et 1981. D'une part, deux observations temporelles ne suffisent pas pour une analyse rétrospective. D'autre part, ces données sont issues d'un système spatial plus agrégé que le nôtre, obligeant ainsi à une répartition zonale. Or, la répartition envisagée requiert des informations fournies par les enquêtes origines-destinations qui ont été menées selon une périodicité particulière (1970-74-78-82) et un système géographique variable.

Devant toutes ces contraintes, il a fallu se résoudre à écarter toutes analyses fondées sur des informations antérieures. L'application spatio-temporelle se restreindra alors à simuler l'évolution de la structure urbaine et des réseaux de transport. L'année de référence devient 1981, année quasi-commune à tous les types de données. La périodicité reste de préférence de cinq ans pour coïncider avec les années censitaires.

L'horizon de simulation est l'an 2001, non pas pour son symbolisme mais parce qu'il correspond d'abord aux horizons de prévision de plusieurs analyses prospectives dont certaines données sont essentielles à la bonne marche de l'étude. Par surcroît, il s'agit d'une durée suffisante pour qu'apparaissent certaines modifications de la structure urbaine mais ni trop longue pour que les incertitudes deviennent prépondérantes.

7.2.3 Définition des réseaux de transport

Dans notre application, deux modes principaux sont considérés, soit l'automobile et le transport collectif. A chacun de ces deux modes, est reconnu un réseau complètement indépendant de l'autre. Cette présomption n'apparaît pas complètement réaliste sachant que la quasi totalité des parcours des autobus ne sont pas en site propre. La présomption d'indépendance facilite certes la modélisation de l'affectation des voyageurs. Une prise en compte de réseaux interdépendants, signifiant des coûts variables pour les autobus, reste toutefois possible et fut proposée par Florian [1977].

Afin de maintenir la banque de données à une taille manipulable et de simplifier les analyses, on n'a retenu que les éléments principaux des réseaux. Par exemple, seules les autoroutes et les artères dites rapides sont considérées dans le réseau routier. Pourtant, maints parcours n'utilisent pas en partie ou en totalité ces liens majeurs. Notre modélisation (voir appendice 3.B) permet cependant de contourner ce problème d'agrégation en admettant pour chaque trajet O-D, un lien «auxiliaire» synthétisant les itinéraires alternatifs. Ce type de lien possède une capacité illimitée mais aussi un coût de transport fixe plus élevé. L'hypothèse de la fixité est appuyée par le fait que les conditions de transport semblent généralement plus stables sur les liens secondaires.

Cette hypothèse est également maintenue pour les parties des itinéraires situées entre le centroïde de zone et un noeud d'accès au réseau.

Quant au transport collectif, une agregation similaire du réseau est supposée. Le réseau principal se compose du métro souterrain ainsi que des lignes ferroviaires de banlieue. Les trajets d'autobus sont traités comme des liens auxiliaires. Dans ce cas-ci, la division du réseau en deux exige peu de présomption à l'égard de leurs caractéristiques car, dans un cas comme dans l'autre, les coûts sont invariables.

Les réseaux sont constitués de noeuds et de liens unidirectionnels. Les noeuds routiers retenus coïncident évidemment aux intersections des liens principaux et à certains points d'accès stratégiques. Quant aux noeuds du réseau public, ils désignent outre les stations de correspondance et les terminus, certaines gares intermédiaires. Tous les noeuds sont geo-codes dans un système de coordonnées identiques à celui des zones et réfèrent aux liens qui y aboutissent. Les liens contiennent une référence des noeuds de début et de fin. Les liens identiques mais de sens opposés ont un même numéro se différenciant par un signe; de cette façon, les données de liens sont enregistrées une seule fois.

Les cartes 7.2 (p. 301) et 7.3 (p. 303) illustrent les routes et lignes retenues dans la composition des réseaux L(1) et L(2). On trouve dans les annexes 7.B et 7.C l'énumération de ces liens avec le numéro correspondant sur les cartes respectives. La numérotation est celle qui a été utilisée dans la codification des itinéraires. Chacune des routes ou des lignes est une séquence exclusive de liens adjacents. Par exemple, l'autoroute 40 (Transcanadienne, Métropolitaine, Rive-Nord) est considérée comme une «route» et une ligne de métro peut correspondre exactement à une «ligne».

Sur les liens routiers, les coûts minimaux (s_i) résultent du produit de la longueur du lien avec la vitesse maximale permise. Les autoroutes typiquement urbaines (Métropolitaine, Decarie, Ville-Marie) et certains ponts autoroutiers (Champlain, Mercier) limitent les automobilistes à une vitesse de 70 km/h. La vélocité autorisée des autoroutes suburbaines et périurbaines atteint généralement les 100 km/h. La plupart des artères retenues permettent une conduite excédant les 50 km/h mais la présence de

feux de circulation diminue du tiers la vitesse praticable; des vitesses de 40 à 45 km/h ont été, selon le cas, assignées.

Les capacités ont été établies en référence des normes conventionnelles i.e. 1800 véhicules par heure sur chaque voie autoroutière. Cependant, pour être réaliste, nous avons octroyé sur certaine voie rapide des valeurs moindres (1400 à 1600) en raison de caractéristiques conceptuelles moins généreuses (par exemple, voie surélevée de la Métropolitaine). Sur les artères connaissant des interruptions de la circulation, les capacités sont limitées à environ 500 veh/hre/voie.

La vitesse commerciale moyenne du métro approche les 36 km/h. Ce nombre tient compte des temps d'arrêt aux stations. La vitesse moyenne sur les trains de banlieues varie beaucoup non seulement d'une ligne à l'autre mais aussi au sein d'une même ligne, dépendant de l'espacement des gares. Néanmoins, nous avons fixé une vitesse normative de 43 km/h. Pour les correspondances «métro à métro», une pénalité de 3 minutes est imposée tandis qu'un supplément de 6 minutes est rajouté aux parcours sollicitant un transfert «train à train» ou «metro à train».

Les temps de déplacement sur les liens auxiliaires des deux réseaux font l'objet d'une approximation particulière. En effet, il s'agit d'évaluer la distance entre deux points (zone à zone ou zone à noeud) selon une grille rectilinéaire imaginaire. La méthode du calcul à vol d'oiseau, plus simple, a été rejetée car elle fait fi de la géométrie quadrilatérale du réseau des routes secondaires. La mesure a été effectuée en tenant compte de l'orientation générale des rues.

Il faut noter que les déplacements franchissant le fleuve St-Laurent et la rivière des Prairies ne se prévalent pas d'un lien auxiliaire interzonal puisque la traversée implique l'usage des réseaux principaux.

Une démarche consistante a été préconisée à l'égard du temps de parcours des liens auxiliaires. La distance est d'abord évaluée et est convertie par la suite par une vitesse constante.

Les distances intra-zonales ont été estimées selon une formule connue en analyse spatiale:

$$s_{11} := (2/3) \cdot (S_1/\pi)^{1/2} \quad (7.1)$$

où s_{11} : distance moyenne à l'intérieur de la zone i

S_1 : superficie urbanisée de la zone i

Il s'agit tout simplement de calculer le rayon median d'un cercle ayant une surface équivalente à la zone. Ainsi plus la zone est vaste, plus la distance moyenne augmente. Dans le cas d'un lien auxiliaire reliant une zone et un noeud (échangeur ou gare), il convient d'introduire la distance intra-zonale à celle du centroïde au noeud puisque les voyageurs ne transigent généralement pas au point central de la zone. Cependant, la contribution de la distance interne n'apparaît pas simplement additive puisqu'elle est importante si noeud et zone sont proches mais peu significative dans le cas contraire. La règle de contribution est formulée de la façon suivante:

$$\bar{s}_{1n} := \{(s_{11})^2 + (s_{1n})^2\}^{1/2} \quad (7.2)$$

où $\bar{s}_{1n}$: distance moyenne entre la zone i et le noeud n

s_{1n} : distance quadrilatérale entre le centroïde de i et le noeud n

Cette formulation est parfaitement consistante avec celle de la distance intra-zonale car si le noeud se localise à même le centroïde ($s_{1n} = 0$), la distance nodale-zonale équivaut à la distance intra-zonale. De même que si la zone a une surface nulle, la distance moyenne devient simplement inter-nodale ($\bar{s}_{1n} = s_{1n}$).

La distance des liens auxiliaires inter-zonaux est conséquemment évaluée en mettant à contribution les distances internes de chaque zone:

$$\bar{s}_{ij} = \{(s_{11})^2 + (s_{1j})^2 + (s_{jj})^2\}^{1/2} \quad (7.3)$$

où $\bar{s}_{ij}$: distance moyenne entre les zone i et j

s_{ij} : distance quadrilatérale entre les centroïdes de i et j

Une vitesse moyenne de 30 km/h assignée aux liens auxiliaires routiers permet d'établir leurs temps de transport. Dans le cas du transport collectif, où les liens synthétiques correspondent à des parcours d'autobus, on a octroyé une vitesse uniforme de 15 km/h. Ce nombre convient surtout

aux lignes de la STCUM desservant les districts les plus denses. Mais il devrait aussi convenir aux trajets plus périphériques où les autobus, généralement plus rapides, infligent en contrepartie des temps d'attente plus longs.

7.2.4 Énumération des itinéraires

Les modèles d'affectation, de type logit (eq. 3.122) ou par trafic d'équilibre (3.125-3.127, 3.129), sollicitent une énumération des itinéraires admissibles pour chaque paire O-D. L'identification de ces chemins devient rapidement un travail fastidieux. En effet, dans notre application, il y a 3600 (60^2) paires O-D. En excluant les 60 paires intra-zonales qui ne nécessitent aucune spécification des itinéraires et en réduisant la recherche des chemins que pour une direction (la liste des itinéraires est symétrique pour i-j et j-i mise à part l'inversion du signe des liens), on obtient 1770 paires où une énumération est requise. Ainsi, chacune de ces paires se dote d'environ un à huit séquences de liens admissibles; la moyenne est de 4.57 pour le réseau routier et de 3.08 pour le réseau collectif.

La sélection des chemins a été sans aucun doute la tâche la plus ardue de la préparation de la banque de donnée. Elle aurait pu se faire à l'aide d'une carte routière de l'agglomération montréalaise. Cependant, avant d'admettre un itinéraire, il faut calculer son temps de transport afin de le comparer au plus court. Avec des temps fixes, la tâche semble facilitée mais lorsque les temps sont variables, il faut évaluer diverses possibilités de congestion. Évidemment, il convient de confier la sélection des chemins à un programme informatisé. Celui-ci procède de la façon suivante:

- À chaque zone est affectée exogènement un certain nombre de noeuds d'accès: il y a de un à dix noeuds possibles;
- On trouve le plus court chemin entre tous les noeuds du système et on calcule leurs temps de transport;
- À chaque paire O-D, on génère autant d'itinéraires qu'il y a de croisements de noeuds d'accès (de un à cent);
- On évalue leur temps totaux en ajoutant les temps «zone à noeud» et «noeud à zone» (voir section précédente);

- On retient l'itinéraire le plus court et ainsi que ceux se trouvant à l'intérieur d'une marge prédéfinie;

Pour les itinéraires en mode collectif, le programme n'est appliquée qu'une seule fois. Par contre, dans le cas du réseau routier, on l'applique selon diverses conditions extrêmes de circulation afin de ne pas exclure des itinéraires qui ne sont fréquentées que lorsque les axes principaux sont congestionnés. Les deux listes incluent aussi des parcours empruntant des liens projetés afin que conviennent aux applications simulatoires.

Une codification spéciale des chemins a été développée de manière à réduire la taille de la banque. Ainsi un parcours n'est identifié que par les numéros des zones d'origine et de destination, les numéros des deux noeuds d'accès, les distances entre zones et noeuds d'accès (calculées selon la méthode 7.2) et les numéros des «routes» ou «lignes» utilisées (voir annexe 7.8 pour leur identification). En connaissant les points d'entrée et de sortie, on connaît alors les liens empruntés. Par ailleurs, si seulement une route est sollicitée par l'itinéraire, ces points correspondent aux noeuds d'accès. S'il y a plus d'une route, les noeuds d'intersection deviennent des entrées et/ou des sorties de la route. Selon l'ordre des numéros d'entrée et de sortie, on reconnaît la direction prise sur la route en question; le signe (+ ou -) des liens empruntés en dépend.

7.2.5 Données de demande de transport

L'essentiel des données de demande de transport provient de l'enquête origine-destination menée par la STCUM en 1982 [STCUM, 1983]. Au moment de la rédaction de cette thèse, les données issues de l'enquête de l'automne 1987 restaient indisponibles. Les données publiées des enquêtes antérieures (1970, 1974, 1978) n'ont pas été utilisées parce qu'elles furent compilées selon un système spatial inconstant et incompatible avec celui choisi en 1982.

L'enquête O-D de 1982 s'est déroulée durant le trimestre d'automne, période durant laquelle l'achalandage journalier moyen s'approche le plus de l'achalandage journalier moyen annuel. Le sondage a été mené auprès de

75,000 menages dont l'un des représentants devait énumérer et décrire les déplacements quotidiens de tous les membres du ménage.

Les matrices de déplacement décrivent le comportement des voyageurs durant une journée typique de semaine. Elles excluent les «retours au domicile» puisqu'ils sont implicites dans la plupart des cas. Dans notre cas, cela facilite la caractérisation de l'origine et de la destination selon la composition des activités d'une zone. Les matrices qui furent enregistrées sont:

- tout motif, tout mode, 24 heures
- tout motif, mode automobile (conducteur + passager), 24 heures
- tout motif, mode STCUM, 24 heures
- tout motif, autres modes, 24 heures
- selon cinq motifs (travail, étude, achat, loisirs, autres),
tout mode, 24 heures

Les matrices n'ont pas été utilisées telles quelles car, d'une part, notre système zonal diffère quelque peu de celui de la STCUM et, d'autre part, les données d'activités (population, emploi) ne sont disponibles que pour les années de recensement. Dans les deux cas, la procédure de correction reste la même.

Par ailleurs, les populations zonales telles qu'inscrites dans le rapport de l'enquête se révèle inexacte puisqu'un facteur d'expansion tenant compte du nombre de répondants y fut appliqué. Cependant, le rapport entre déplacements et population, c'est-à-dire le taux de mobilité, est juste si bien que nous avons réestimé les origines selon la population 1981 recensée par Statistique Canada. Selon cette méthode, le traitement des zones scindées et la conversion 1982-1981 s'effectue simultanément et de façon consistante. La distribution des destinations n'a pas été touchée: mise à part la correction du total global.

A partir des totaux marginaux corrigés, les matrices ont été reconstituées selon la méthode de balancement multiproportionnel de matrice:

$$t_{i,j} := a_i O_i b_j D_j y_{i,j} \quad (7.4)$$

$$a_i := 1/\sum_j b_j D_j y_{i,j} \quad b_j := 1/\sum_i a_i O_i y_{i,j}$$

où $t_{i,j}$: nombre corrigé de déplacement de i vers j ; $y_{i,j}$: nombre initial de déplacement de i vers j ; a_i, b_j : facteurs de balancement; O_i, D_j : origine et destination corrigée.

Le lecteur notera l'isomorphisme de cette formulation avec du modèle de distribution spatiale doublement contraint (eq. 3.20). La différence réside dans la substitution de la fonction de transport par le nombre observé de déplacement. Etant donnée que la distribution des totaux marginaux observés et corrigés diffère très peu, les entrées de la matrice corrigée deviennent quasi-identiques. Ils permettent toutefois de respecter les nouveaux totaux marginaux.

Un autre problème qu'il a fallu surmonter est celui de l'indisponibilité d'une véritable matrice compilant les déplacements selon tout mode collectif. Le rapport présente certes une matrice «STCUM» et une autre dite «autres modes» mais leur addition ne résulte pas en une matrice comprenant seulement les modes collectifs. En effet, selon leur définition, les «autres modes» incluent outre les voyages effectués par les autres sociétés de transport mais aussi les voyages non-motorisés (piétons et cyclistes). Heureusement, les totaux marginaux de ceux-ci sont donnés et on sait pertinemment que la grande majorité de ces déplacements s'effectue au sein d'une même zone (la preuve est que le nombre d'origines est quasi-identique à celui des destinations de la zone). Toutefois, à cause des effets de frontière et de la mobilité plus grande des cyclistes, il semble inapproprié de destiner tous les voyages non-motorisés à leurs zones de départ.

La méthode d'approximation des déplacements non-motorisés s'appuie sur le fait que dans le centre de la Communauté Urbaine, les «autres modes» sont en majorité des «non-motorisés» (les déplacements en transport en commun utilisant au moins un trajet STCUM sont automatiquement répertoriés STCUM). En respectant les totaux marginaux des non-motorisés du centre de la CUM, il a donc été possible d'estimer un paramètre de transport permettant de distribuer spatialement les piétons et cyclistes. La matrice «non-

motorisée» pour l'ensemble de la région d'étude a pu donc être évaluée selon un modèle de gravité muni d'un effet frictionnel très sévère. La matrice «tout mode collectif» résulte de la soustraction de la matrice «non-motorisée» avec la matrice «tout mode».

7.2.6 Données d'activités urbaines

Nous avons restreint les données d'activité aux données de population et d'emplois selon certaines catégories. La population reste un indicateur général de l'activité résidentielle. Elle demeure une donnée facile à obtenir selon divers découpages géographiques. Cependant, elle occulte des caractéristiques démographiques permettant de nuancer l'évolution et la distribution de l'activité résidentielle. Par exemple, le nombre de logements peut s'accroître en simultanéité d'un déclin de la population si la taille moyenne des ménages diminue. De plus, la composition de cette population (âge, sexe, revenu) est déterminante dans la mobilité quotidienne [Bussière et al., 1988]. Néanmoins, les caractéristiques de la population ont été ignorées de manière à simplifier l'application.

Pour les zones correspondant à une ou plusieurs entités municipales, la population a été calculée à partir des données par subdivision de recensement. Pour les zones inscrites dans un découpage municipal plus grand (Montréal et Laval), il a fallu s'en remettre aux données ventilées par secteurs de recensement. Seules les données de 1981 et 1986 ont été retenues pour la simulation.

La formation d'une banque de données d'emploi constitue un casse-tête infiniment plus compliqué. En fait, il n'existe pas de données d'emploi fiables convenant au niveau d'agrégation spatial voulu. Pourtant, ces informations apparaissent essentielles à la mise en opération de notre modèle. Beaucoup d'études combinant transport et localisation se sont butées à l'indisponibilité de telles données (voir Miller, Cubukgil [1981] pour la région de Toronto).

L'application désirée ne vise à pas à dresser des prévisions strictes (d'ailleurs comment prévoir des choses qui sont méconnues pour le temps présent?). Elle tente plutôt de démontrer l'utilité d'une méthodologie

dans un contexte de planification. Dans ce cas-ci, les données d'emploi spatialisées n'ont pas à être d'une précision absolue mais plutôt indicatives.

En admettant cet assouplissement, il devient permis de construire une banque de 60 zones en répartissant adéquatement des données d'emploi fiables provenant d'une base spatiale plus agrégée. L'INRS-Urbanisation fit une commande à Statistique Canada pour l'obtention de deux banques d'emploi, pour les années 1971 et 1981, pour la région métropolitaine de Montréal. Bien que pourvues d'imprécisions, ces banques sont à la fois finement désagrégées par secteur et spatialisées en dix sous-régions. Dans le cadre de notre étude, les données présentées dans le rapport de Lamonde, Ledent et Polèse [1984] suffisent pour établir notre petite banque.

Dans le rapport, on n'y retient que les chiffres de 1981 du secteur manufacturier et des grandes catégories du secteur tertiaire. Les sous-régions utilisées sont:

- | | |
|-----------------------------|-------------------|
| 1- Centre-Ville de Montréal | 6- Est de la CUM |
| 2- Centre de la CUM | 7- Ville de Laval |
| 3- Ouest de la CUM | 8- Couronne Nord |
| 4- Sud-Ouest de la CUM | 9- Longueuil |
| 5- Centre-Ouest de la CUM | 10- Couronne Sud |

Les sous-régions 2 à 5 s'accordent avec les limites couramment utilisées au Service d'aménagement du territoire de la C.U.M.. Les couronnes Nord et Sud adoptent le contour de la région métropolitaine de recensement de 1971 et, du fait, ne correspondent pas exactement à celles constituées par nos zones d'étude. Aussi, notre zone Centre-Ville est légèrement plus étendue que la leur. Toutefois, des ajustements peuvent facilement être faits bien qu'en général, nos zones s'inscrivent dans ces sous-régions.

La méthode de répartition de l'emploi se base sur certaines relations fonctionnelles entre les activités. Dans un premier temps, l'emploi total de chaque sous-région est répartie selon la part relative des destinations à motif de travail. Cette dernière donnée provient de l'enquête O-D de la STCUM et, par conséquent, s'accorde avec notre système spatial. Cette première répartition semble facile à réaliser et peu risquée puisqu'il y a une forte corrélation entre l'emploi et la donnée de transport.

Les services à la consommation comme le commerce de détail et les services personnels sont repartis selon la part relative des destinations à motif d'achat. Cette variable semble très indicative de la fonction commerciale d'une zone. Les emplois liés à l'enseignement sont alloués d'après les destinations à motif d'étude tandis que ceux associés aux divertissements et loisirs ainsi qu'à l'hébergement et la restauration s'en remettent aux destinations à motif de loisirs.

Certaines catégories d'emploi ne peuvent pourtant être réparties à l'aide des données de destinations. Il importe de faire certaines présomptions de leurs comportements de localisation. Les services médicaux et sociaux ont par exemple été distribués en deux étapes. D'abord, le répertoire de l'Association Canadienne des Hôpitaux [A.C.H., 1981] fournit des données sur le personnel des établissements hospitaliers. Ces nombres peuvent être directement affectés aux zones en localisant les hôpitaux. Ensuite, la partie restante est répartie en proportion des populations résidentes.

Les services aux entreprises et les services divers sont partagées au prorata des emplois totaux. Les catégories «finances, assurances, affaires immobilières» et «administration publiques» sont distribuées selon une combinaison linéaire de l'emploi total, l'emploi commercial et la population; la pondération de cette combinaison a été estimée d'après la distribution observée au sein des dix sous-régions.

Enfin, les emplois manufacturiers, du commerce de gros et des services publics (transport, communication, énergie) sont la partie restante de l'emploi total soustraite par les autres catégories d'emploi.

Les marges d'erreurs d'une telle répartition zonale semblent, à notre avis, acceptables. D'une part, les emplois observées sont déjà spatialisées; les erreurs de répartition se restreignent donc à l'intérieur des sous-régions. D'autre part, la zone dont la composition apparaît la plus originale, le centre-ville, dispose de données propres à elle.

Les erreurs de repartition sont par surcroît atténuées par un regroupement des catégories d'emplois selon un comportement présumé de localisation. Cette reclassification est nécessaire par l'application du modèle spatio-temporel de localisation d'activités.

L'emploi exogène (noté E^e) se compose du secteur manufacturier, du commerce de gros et des services publics. L'association de ces deux catégories tertiaires avec le manufacturier se justifie par des comportements de localisation apparentes. En effet, les entreprises de commerce de gros tendent à se grouper dans les zones industrielles et semblent relativement indépendantes des autres activités tertiaires. Les services publics se localisent en fonction d'impératifs spécifiques et affichent une certaine inertie géographique (terminaux de transport, services d'entretien des réseaux de communication, de transport d'énergie, etc...).

Le second regroupement contient les emplois de service surtout orientés vers les marchés résidentiels. Le commerce de détail ainsi que les services personnels et éducatifs appartiennent à cette classe. Toutefois, la proximité des travailleurs peut devenir un élément incitatif de localisation. On appelle ce regroupement «emplois tertiaires de type 1» (noté S^1).

Le troisième groupe concerne les catégories tertiaires qui tendent à se concentrer dans les grands pôles d'emplois comme les finances, l'administration publique et les services médicaux. Par ailleurs, ils sont également liés à l'évolution de la population. Il s'agit des emplois tertiaires de type 2 (noté S^2).

Le dernier groupe (noté S^3) se rapporte aux services aux entreprises et aux services divers qui tendent à accompagner les concentrations d'emplois.

Deux autres types d'information sur les activités sont nécessaires à la mise en opération du modèle de localisation. D'abord, la distribution spatiale de l'emploi exogène doit être connue pour les périodes futures. Pour cela, les prévisions sous-régionales 1986-2001 du rapport de Lamonde et al. [1989; chap.2] ont été réparties selon la distribution zonale de 1981.

Ensuite, pour les années à venir, le total régional de chacune des activités à localiser (P , S^1 , S^2 , S^3) doit être fixé préalablement. Pour obtenir des prévisions de population consistantes, on recourt au scénario démographique de référence de ce même rapport. Seuls quelques correctifs ont été apportés pour compenser notre définition territoriale légèrement différente. Les totaux métropolitains des trois catégories tertiaires sont issus des prévisions sectorielles d'emploi de Lamonde et al. [1989].

7.3 PRESENTATION DU MODELE OPERATIONNALISE

7.3.1 Introduction

Les chapitres 3 à 6 ont présenté plusieurs développements possibles à l'égard des sous-modèles appartenant au modèle général. Les versions proposées paraissent applicables que lorsque les données pertinentes sont disponibles et que les paramètres sont mesurables. Le modèle opérationnel qui sera exposé ci-après suit certes le schéma défini dans la section 6.6.1. Mais les sous-modèles qui le composent tiennent compte des données recueillies et de la spécificité du cas montréalais.

Les sections qui suivent décriront les différentes phases du cycle opérationnel en citant la forme fonctionnelle des modèles utilisés, la méthode de calibration et les résultats de cette calibration.

7.3.2 Génération spatiale des déplacements

La génération spatiale des déplacements au temps "t" est une étape relativement simple qui consiste à évaluer le nombre de voyageurs produits et attirés par zone à partir de variables d'activités et d'attributs socio-économiques de la population. Une paire de modèles de régression comme (3.1-3.2) permet d'estimer ces nombres. En fait, les régresseurs apparaissent des indicateurs d'activités identifiés géographiquement, et tout particulièrement, des variables d'effectifs se rapportant à ces activités (population, emplois totaux, emplois commerciaux, etc...). À l'égard du transport, ces catégories d'effectifs correspondent implicitement à des grands motifs de déplacement.

La population zonale, indicateur de l'activité résidentielle, peut être cependant modulée par un taux de mobilité lui-même modélisé par certaines caractéristiques pertinentes de la population (voir équation 3.9). Selon les applications régionales choisies, il n'est pas certain que ce taux soit clairement déterminé par des attributs bien identifiés. De plus, les relations souvent non-linéaires entre ce taux et ses facteurs de variation doivent être spécifiées (la transformation Box-Tukey peut être fort utile ici). Par conséquent, la sous-modélisation du taux reste facultative et

n'a pas été expérimentée ici de manière à limiter l'application.

Les coefficients des deux équations de régression (3.1-3.2) peuvent être estimés à partir de données issues de l'enquête origine-destination. Dans notre cadre, on présume une constance des coefficients à travers les périodes; toutefois, cette restriction peut être relâchée si on dispose d'une connaissance suffisante de la dynamique des coefficients. La formulation des modèles régressifs est reprise ici:

$$D_i = \alpha_0 + \sum \alpha_r X_{ri} \quad (7.5)$$

$$D_j = \beta_0 + \sum \beta_r Y_{rj} \quad (7.6)$$

où α_r, β_r : coefficients de régression; X_{ri}, Y_{rj} : variables d'activité; D_i, D_j : origine et destination.

La méthode des moindres carrés (OLS) semble à prime abord adéquate mais elle doit être appliquée en tenant compte de la contrainte d'égalité des sommes d'origines et de destinations i.e.:

$$\sum_i D_i = \sum_j D_j$$

Les méthodes linéaires d'estimation ne sont habituellement pas sujettes à de telles contraintes. Cette égalité étant normalement respectée chez les observations, elle le restera pour les estimés de la même période; en effet les estimés OLS conservent toujours une moyenne identique à des observations. Là où le risque d'inégalité entre les deux totaux surviendra, c'est pour les périodes futures. En pratique, cette divergence pourrait s'accroître avec le temps mais sera somme toute très faible, considérant l'ampleur des nombres impliqués. L'application d'un facteur de correction ramenant artificiellement l'égalité permettra d'éliminer cette incohérence sans trop contrevenir à la rigueur de notre cadre théorique.

Les modèles de prédiction des origines et des destinations ont donc été spécifiés (identification des régresseurs) et estimés à partir des données de 1981. Les variables dépendantes sont les totaux marginaux de déplacements de 1981 tandis que les régresseurs se retrouvent parmi la population et les catégories d'emplois décrites dans la section 7.2.6.

Pour le modèles des origines, plusieurs combinaisons de régresseurs ont été essayées. Les critères de sélection des variables sont classiques i.e.

la contribution à la variance expliquée et la validité respective de l'estimateur. Voici un tableau dressant brièvement l'analyse de régression qui a obtenu les résultats les plus satisfaisants.

Tableau 7.1 Résultats de l'estimation du modèle d'origine

<u>Régresseurs</u>	<u>Coefficients</u>	<u>Erreur-type</u>
Constante	2330.7	(839.4) ***
Population 1981 (P)	0.7338	(0.0150) ***
Emploi de service 1981 ($S = S^1 + S^2 + S^3$)	0.2710	(0.0246) ***
<u>Variable dépendante</u>	<u>Nombre de cas</u>	<u>Coef. de détermination (R^2)</u>
Origine 1981	60	0.9786

*** valide à 99% selon une distribution t_{57}

Dans tous les essais, la variable «population» était présente de par la forte relation entre l'activité résidentielle et le nombre quotidien de départs. La variable «emploi de service» est constituée des trois catégories d'emploi non-exogènes. Sa contribution reste complémentaire mais essentielle dans l'explication des variations spatiales des origines (les zones concentrant ces types d'emploi génèrent beaucoup de déplacements à motif d'affaires). L'exclusion d'une variable «emploi total» tient au fait qu'on n'a observé aucune relation statistique significative entre la distribution de l'emploi exogène (dominé par le secteur manufacturier) et celle des origines. Intuitivement, cette indépendance n'étonne pas.

Concernant le nombre de destinataires, diverses combinaisons de régresseurs ont été testés. Le modèle de régression le mieux ajusté réunit cette fois, la population, l'emploi total et une variable groupant les emplois de consommation et bureaux (types 2 et 3). L'emploi total symbolise les déplacements de travail. Les emplois orientés vers la consommation et ceux de bureaux interprètent l'ensemble des déplacements motivés par les achats, les études, les affaires, les loisirs et toute autre raison. Le rôle de la variable «population» semble ici de nature mandataire i.e. représentant un ensemble de facteurs susceptibles d'attirer des gens dans une zone. Les résultats de la régression sont dressés sur le tableau suivant:

Tableau 7.2 Résultat de l'estimation du modèle de destination

<u>Regresseurs</u>	<u>Coefficients</u>	<u>Erreur-type</u>
Constante	1106.2	(1113.5) (1)
Population 1981 (P)	0.1814	(0.0200) ***
Emploi total 1981 (E)	0.6938	(0.0723) ***
Emploi cons/bur. 1981 ($S^1 + S^2$)	1.9079	(0.1478) ***

<u>Variable dependante</u>	<u>Nombre de cas</u>	<u>Coef. de détermination (R^2)</u>
Origine 1981	60	0.9908

(1) non significatif statistiquement

*** valide à 99% selon une distribution t_{57}

Le fait que la variable «emploi total» contient explicitement celle des emplois de 1^{er} et 2^e types pourrait présenter des risques de collinéarité. Dans notre application, cela ne s'est pas posé puisque les deux variables n'étaient pas suffisamment corrélées pour rendre invalide la contribution de l'une d'elles.

Le lecteur notera la force inhabituelle des coefficients de détermination. Ils démontrent la relation excessivement stable entre l'ampleur des activités dans une zone et les déplacements générés. De plus, la "loi des grands nombres" garantit une constance des relations lorsque les observations sont établies dans un système zonal passablement agrégé.

7.3.3 Distributions spatiale, modale et routière

Selon notre cadre de modélisation présenté dans le sous-chapitre 3.7, les déplacements générés sont distribués spatialement, répartis selon les modes et affectés sur des itinéraires de façon simultanée. Il suffit de solutionner le problème de minimisation (3.172-3.179).

Les variables sollicitées par cette étape sont:

- les nombres de déplacements produits et attirés par zone: $\{O_i\}$, $\{D_j\}$;
- le coût généralisé pour chaque itinéraire n de transport public d'un trajet $i-j$: $\{z_{ij}^n\}$;
- la capacité des liens l du réseau routier L^1 : $\{q_l\}$;

- la fonction de congestion des liens ℓ du réseau L^1 : $\{\ell_\ell(v)\}$;
- la matrice d'incidence lien-chemin: $\{\delta_{ij}^{\ell}\}$.

Dans le cadre de l'application à la région de Montreal, nous décidons de fixer les paramètres modal et spatial selon les distributions observées en 1981. Cette alternative se distingue de celle préconisant la satisfaisant des contraintes d'entropie appropriées à chaque usage du modèle intégré de demande de transport. Il suffira de calibrer les sous-modèles en question selon les conditions de transport observées en 1981.

Ces conditions sont définies par les deux matrices de coût modal de transport. Les coûts de transport collectifs se présentent comme des coûts composés issus d'un processus de répartition de type logit (voir eq. 3.183-3.184). Comme il a été mentionné dans la section 3.8.4, ce processus peut être traité à l'écart du problème intégré. L'affectation de la demande sur les chemins n de transport public peut se faire indépendamment de l'affectation routière des automobilistes car aucun effet de congestion n'est présumé pour le transport collectif.

Aucune donnée n'étant disponible concernant le choix d'itinéraire en transport en commun, il a fallu choisir intuitivement une transformation et un paramètre de partage. En fait, ce choix plutôt arbitraire a peu de conséquence ici car le coût composé, plus que l'affectation proprement dite, importe dans le modèle intégré. Pour cette raison, une distribution très sévère a été choisie de manière à ce que chaque coût composé tende vers le coût minimal par trajet.

Il a été présumé que la perception des usagers envers les différences de coûts de parcours est logarithmique. Autrement dit, les itinéraires sont comparés non pas en différences absolues mais plutôt selon un rapport. Le sous-modèle d'affectation des itinéraires devient comme suit:

$$P(n|ij) = \frac{\exp(-p \ln z_{ij}^2)}{\sum_n \exp(-p \ln z_{ij}^2)} \equiv \frac{(z_{ij}^2)^{-p}}{\sum_n (z_{ij}^2)^{-p}} \quad (7.7)$$

Le paramètre p a été fixé de telle sorte qu'un parcours 1.2 fois plus long est trois fois moins choisi; donc $p = \ln(3)/\ln(1.2) \approx 6.0257$. Selon cette formulation et l'expression (3.184), le coût composé c_{ij}^2 du trajet $i-j$

selon le mode 2 est défini de la façon suivante:

$$\ln(c_{ij}^2) = \frac{1}{-p} \ln \sum_n (z_{ij}^n)^{-p} + \frac{1}{p} \ln(N_{ij}) \quad (7.8)$$

Le terme N_{ij} désigne le nombre d'itinéraires admissibles dans le trajet $i-j$. Seuls les parcours présentant un coût inférieur à 1.2 fois celui du plus court étaient admis.

Pour les fins de la calibration, les coûts par automobile sont considérés exogènes mais restent des coûts d'équilibre. Pour obtenir cet équilibre, il suffit d'affecter sur le réseau routier de 1981 les déplacements automobiles 24 heures. Toutefois, la congestion se mesure habituellement par heure et varie énormément d'une heure à l'autre. Par contre, nous voulions obtenir un coût d'équilibre sur une période de 24 heures (et non seulement en heure de pointe) afin qu'il représente les conditions générales de transport.

Il ne suffit pourtant pas de multiplier les capacités horaires par 24 car la majorité des voyages s'effectue durant le jour et surtout en période d'affluence. Les effets moyens de congestion sont donc difficiles à cerner. Pour appréhender ce phénomène, nous choisissons d'étaler la capacité sur différentes tranches horaires et on procède chaque fois à l'affectation d'équilibre. Selon la fonction de congestion en vigueur, une période d'étalement de six heures est apparue convenable si on en juge la vraisemblance de la distribution sur le réseau; à moins de 5 heures, la congestion était excessive et, à plus de 7 heures, seuls les plus courts chemins étaient empruntés. Le chiffre de six peut aussi se justifier par les faits suivants:

- (1) la circulation se limite à 16-18 heures par jour;
- (2) les déplacements de retour sont exclus (coupant de moitié la période effective);
- (3) l'irrégularité horaire des déplacements fait que les 8-9 heures restantes équivalent à une congestion étalée sur à peu près 6 heures.

La fonction de congestion employée sera constante d'un lien à l'autre et d'une application à l'autre. Elle se présente ainsi:

$$s_c = s_0 \cdot \{1 + 0.3(v_c/q_c)^5\} \quad (7.9)$$

La capacité Q_r est étalée sur six heures tandis que le volume V_r est genéré sur une journée complète.

La calibration des paramètres modal et spatial se fait selon la méthodologie de «vraisemblance maximale» exposée dans la section 3.8.3. Toutefois, le sous-modèle de répartition modale reste inadéquatement spécifié. En effet, il ne semble pas assuré que les coûts modaux de transport restent les seuls facteurs discriminants. Pour cette spécification, on procède à une estimation préliminaire du partage modal en recourant aux proportions observées en 1981 parmi les déplacements motorisés.

Plusieurs essais ont été nécessaires pour spécifier le modèle logit de partage. D'abord, il est devenu évident que les différences de coût ne suffisent pas pour expliquer le partage observé sur certain trajet. Pour tenir compte de la difficulté de stationner, on a inscrit un coût terminal à chacune des destinations. Il s'agit d'une pénalité allant de 0 à 9 et s'adressant uniquement aux automobilistes. La proportion des ménages non-motorisés dans une zone apparaît également une variable pertinente. Elle concerne plutôt les origines et peut être considérée soit comme un bénéfice au transport en commun, soit comme un coût non-monétaire pour l'usage de l'auto. On a opté pour la dernière définition afin que toutes les mesures soient perçues comme des coûts.

Le sous-modèle était calibré selon une présomption d'un paramètre unique pour toute la région d'étude. Pour les coûts terminaux, cette restriction semblait convenable. Par contre, la magnitude des coûts de parcours diffère substantiellement d'un trajet O-D à l'autre. Les différences absolues entre les modes tendent alors à croître selon la longueur du trajet; elles se révèlent donc peu significatives. On ne saurait, cette fois, préconiser une transformation logarithmique des coûts de transport du fait que le sous-modèle dérive du problème de minimisation (3.172-3.179) qui ne peut s'accommoder d'une telle relation non-linéaire. Il peut toutefois s'accommoder de paramètres multiples, comme par exemple, un paramètre par paire O-D.

Un tel nombre de paramètres rend impossible toute calibration basée sur une seule observation par paire O-D. Le dilemme a été résolu de la façon suivante. Etant donné que l'hétérogénéité des longueurs de parcours a été reconnue comme l'obstacle à l'unicité paramétrique, nous avons, pour les besoins de la calibration, divisé tous les coûts modaux par leurs moyennes bi-modales. De cette manière, il suffit d'estimer un paramètre global qui, lorsque utilisé dans les prédictions, sera "régionalisé" par la moyenne bi-modale de chaque trajet O-D. Le sous-modèle de répartition se présente ainsi:

$$p_{Tj} = \frac{\exp(-\theta_{1j}c_{Tj} - \theta'g_{m1} - \theta''h_{mj})}{\sum_m \exp(-\theta_{1j}c_{Tj} - \theta'g_{m1} - \theta''h_{mj})} \quad (7.10)$$

où $\theta_{1j} = \theta/\bar{c}_{1j}$, $\bar{c}_{1j} = (c_{1j} + c_{2j})/2$

p_{Tj} : proportion modale du mode m dans le trajet $i-j$

g_{m1} : proportion des ménages sans auto ($g_{21} = 0$)

h_{mj} : indicateur des coûts de stationnement ($h_{2j} = 0$)

$\theta, \theta', \theta''$: paramètres (uniques pour le système)

Selon cette définition de la fonction de partage, le coût composé γ_{1j} devient:

$$\gamma_{1j} = \frac{1}{-\theta_{1j}} \ln \sum_m \exp(-\theta_{1j}c_{Tj} - \theta'g_{m1} - \theta''h_{mj}) \quad (7.11)$$

Le coût composé tient compte désormais non seulement des conditions de transport mais aussi de la disponibilité d'un véhicule-moteur et de la facilité du stationnement.

La procédure de Newton-Raphson présentée dans la section 3.8.3 a servi à estimer simultanément les paramètres modaux $\theta, \theta', \theta''$ et spatial β . On a préféré l'emploi des contraintes de coût à celles d'entropie parce que les coûts étaient fixes et que la convergence se faisait plus rapidement (les dérivés secondes des fonctions d'entropie sont imprécisément approximables de par leur non-linéarité élevée). Le tableau suivant expose les résultats de la calibration.

Tableau 7.3 Résultats de la calibration modale-spatiale

Nombre de déplacements	2,417,834	$\sum_{i,j} t_{i,j}$	(obs)
Voyageurs en automobile	1,504,607	$\sum_{i,j} t_{i,j} p_{i,j}$	(obs)
Usagers du transport public	913,227	$\sum_{i,j} t_{i,j} p_{i,j}^*$	(obs)
Entropies			
Observée (déplacements)	-2,724,737	$-\sum_{i,j} t_{i,j} \ln(t_{i,j}) - T \ln(T)$	
Estimée (déplacements)	-2,428,033		
Observée (répartitions modales)	2091.52	$-\sum_{i,j} t_{i,j} \sum_{\pi} p_{\pi,j} \ln(p_{\pi,j})$	
Estimée (répartitions modales)	2215.18		
Paramètres			
θ (coût de parcours)	2.57603	$(\theta_{i,j} = \theta/\bar{c}_{i,j})$	
θ' (possession d'auto)	0.66282		
θ'' (stationnement)	0.15666		
β (coût composé $\gamma_{i,j}$)	0.16954		

Les valeurs paramétriques estimées seront conservées tout au long des simulations prévisionnelles. Cette constance signifie des comportements stables au sein des voyageurs i.e. une perception inchangée à l'égard des coûts. C'est évidemment une hypothèse non-vérifiée mais une présomption contraire paraît risquée puisqu'on ne peut prévoir l'évolution des comportements.

7.3.4 Solution du problème de demande de transport

Dans la section 3.8.3, deux algorithmes singulièrement semblables ont été présentés. L'algorithme de Frank-Wolfe adapté par Florian, Nguyen et Ferland [1975] définit une direction de descente à partir de la solution d'un sous-problème de transport en programmation linéaire. L'algorithme d'Evans [1976] possède une structure semblable mais déroge à la rationalité du «type Frank-Wolfe» par une détermination différente de la direction de descente. Celle-ci émane de la solution d'un sous-problème non-linéaire, précisément celui permettant de dériver le modèle de distribution spatiale doté du même paramètre β du problème principal. A la lumière des résultats de LeBlanc et Farhangian [1981] et nos investigations préliminaires, l'algorithme d'Evans converge plus rapidement. Selon LeBlanc et Farhangian, le sous-problème linéaire n'attribue des directions de descente

qu'à certains trajets O-D et laisse intacte les autres. Ceci n'est qu'une apparence trompeuse car, à cause des contraintes marginales effectives sur les valeurs incrementées, les incréments doivent être compensés par des baisses ailleurs. Une progression oscillatoire en résulte. En utilisant le modèle de distribution spatiale, les entrées de la matrice O-D sont modifiées simultanément; les incréments et les décréments se font en accord.

L'avantage de l'algorithme d'Evans réside aussi dans la simplicité de la programmation de la sous-routine générant les directions de descente; une simple procédure de balancement des multiplicateurs de Lagrange est requise. De plus, la démarche algorithmique nécessite déjà cette procédure pour la distribution O-D initiale; la procédure est donc doublement employée. Dans l'algorithme alternatif, cette procédure reste nécessaire en plus de celle calculant les directions de descente. Cette dernière appartient à la famille du «Simplexe» et est très connue pour son extraordinaire efficacité en dépit de son apparente complexité.

Le programme solutionnant le problème d'optimisation de la demande est maintes fois sollicité. Il est à la fois utilisé pour évaluer la demande sur le réseau pour une période donnée et comme sous-problème du problème de conception de réseau. Dans toutes les applications, les liens auxiliaires (routiers) sont considérés comme des liens réguliers mais non assujétis à la congestion. Du trafic y est affecté que si leurs coûts fixes égalent ceux du coût minimal du trajet.

La considération des liens auxiliaires et l'inclusion du transport collectif semble faciliter la convergence vers un trafic d'équilibre sur le réseau routier. En effet, ils agissent en tant que "soupape" pour chaque trajet O-D; dès que la congestion fait augmenter les coûts d'un parcours, la désaffection des voyageurs bénéficie non seulement aux autres itinéraires (comme dans un problème à demande modale fixe) mais aussi au chemin auxiliaire (dépendant de son coût) et au transport collectif.

7.3.5 Phase de modification des réseaux de transport

Le modèle d'offre choisi ici se restreint à prescrire des additions de liens bi-directionnels de réseau pour une période donnée. Il s'agit de résoudre un «problème de conception de réseau» qui s'accommode de deux réseaux concurrents ($L(1)$ et $L(2)$, entre autre) et d'un effet de congestion chez l'un d'eux. Le problème consiste à sélectionner un sous-ensemble de liens parmi un ensemble de liens projetés qui puisse minimiser le total des coûts de transport tout en respectant une contrainte d'infériorité à un budget de construction préétabli. Le sous-ensemble solution ajoutée au réseau existant au temps " t " définissent le réseau à " $t+1$ ".

Le problème tel qu'il s'insère dans le cadre opérationnel est formulé par le programme (4.29-4.41). Ce programme possède deux niveaux distincts, l'un étant identifié comme le «problème principal», l'autre comme un «sous-problème de demande». Cette distinction s'explique par la diversité des variables décisionnelles; dans le problème principal, une variable dichotomique y_i (0 rejet / 1 ajout) reste l'enjeu tandis que le sous-problème sert uniquement à estimer le trafic sur lien (v_i) d'une configuration donnée de liens. Le calcul de cette demande est primordial pour le problème principal car il permet l'évaluation des coûts totaux.

Il faut réitérer la complexité du problème de conception de réseau à capacité limitée. D'une part, la fonction objective du problème principal (4.29) commande la résolution du sous-problème lequel établit un trafic d'équilibre; toutefois, les fonctions objectives de chacun d'eux, bien que similaires, diffèrent pour le type d'optimum, l'un étant pour le système, l'autre pour les usagers. D'autre part, la structure algorithmique «branche-et-borne» sollicite à chaque noeud de l'arbre d'énumération une évaluation d'une «borne inférieure». Cette borne est cependant établie par une solution optimale de système; en modifiant la fonction de congestion, la procédure ordinaire d'affectation de trafic d'équilibre peut être utilisée.

Dans l'application qui sera présentée ici, les paramètres spatial (β) et modaux ($\theta_1, \theta', \theta''$) déjà estimés pour 1981 sont repris pour solutionner le problème de congestion. Il aurait été possible de travailler plutôt

avec des contraintes d'entropie. La satisfaction de ces contraintes requiert une sous-procédure supplémentaire qui ralentit énormément la solution du programme principal car elle doit être appliquée à chaque itération. Par conséquent, on privilégie la résolution d'un problème dont la fonction objective contient des termes d'entropie déjà paramétrisés. La section 3.3.3 a démontré l'équivalence des deux formulations du point de vue des conditions d'optimalité.

7.3.6 Relocalisation des activités

Cette phase vise à déterminer les changements qui peuvent survenir à la structure urbaine en fonction de l'accessibilité. Celle-ci peut être modifiée par une variation de la congestion et par des améliorations du réseau. En même temps, les évolutions respectives de chacune des catégories d'activité s'influencent mutuellement.

Le système d'équation spatio-temporelles (6.132, 6.134) représente un processus dynamique de localisation. Il s'agit d'une version théorique difficilement applicable puisqu'elle est fortement paramétrisée. Elle commande également une identification et une spécification spatio-temporelle relativement élaborée. La base de données disponibles pour la région métropolitaine de Montréal apparaît nettement insuffisante.

Dans un premier temps, les termes d'autocorrélation et de corrélation spatiales sont éliminés car aucune structure géographique appropriée n'est connue. L'autocorrélation temporelle est conservée car elle exprime la permanence des activités dans une zone.

Dans un deuxième temps, seule l'activité résidentielle est représentée par une variable disponible par tranche de cinq ans. Les données d'emplois n'ont été réparties spatialement que pour 1981. Par conséquent, il ne peut y avoir d'identification des délais et d'estimation fondée sur une véritable série temporelle. Pour cette raison, dans notre application prévisionnelle, un décalage uniforme de cinq ans a été imposé. Non seulement ce délai concorde avec la durée inter-censitaire mais convient aussi aux processus décrits ici. En effet, ceux-ci évoluent de manière non-cyclique (processus de premier ordre mais non-stationnaires) si bien qu'ils

peuvent être décrits par le décalage disponible le plus court (voir section 5.2.2).

Tout en décrivant les procédures d'estimation, il convient de formuler les équations telles qu'elles seront appliquées. Le sous-modèle de localisation résidentielle s'écrit comme suit:

$$P_{i,t+1} = \Omega \cdot P_{i,t} + \sum_j \pi_{i,j,t} E_{j,t} \quad (7.12)$$

$$\text{où } \pi_{i,j,t} = \frac{\bar{a}_t \cdot Q_i \exp(-\psi^R \gamma_{i,j,t})}{\sum_i Q_i \exp(-\psi^R \gamma_{i,j,t})} \quad \bar{a}_t = (P_{t+1} - \Omega P_t) / E_t$$

$P_{i,t}$: population de la zone i au temps t ; Q_i : facteur d'attraction résidentielle de i ; $\gamma_{i,j,t}$: coût composé de transport au temps t ;
 $E_{j,t}$: emploi total de la zone j au temps t ;
 P_t, E_t : totaux de la région d'étude (exogènes).

Le paramètre de transport ψ^R a été préalablement estimé par vraisemblance maximale avec la matrice O-D de déplacements de travail 1981 et la matrice de coûts composés $\{\gamma_{i,j}\}$ de la même année. Le facteur d'attraction Q_i a été fixé selon le rapport $P_i / \sum_j \pi_{i,j} E_j$ de 1981. Le paramètre Ω a été estimé par régression à partir des 60 zones et de la période 1981-1986.

L'emploi tertiaire orienté vers la consommation, S^1 , est modelisé par la distribution résidentielle et celle des emplois de service. Ainsi:

$$S_{i,t+1}^1 = r \cdot S_{i,t}^1 + B \cdot \sum_j \sigma_{i,j,t} P_{j,t} + (1-B) \cdot \sum_j \varphi_{i,j,t} S_{j,t} \quad (7.13)$$

$$\text{où } \sigma_{i,j,t} = \frac{\bar{o}_t \cdot W_1 \exp(-\psi^S \gamma_{i,j,t})}{\sum_j W_1 \exp(-\psi^S \gamma_{i,j,t})} \quad \varphi_{i,j,t} = \frac{\bar{i}_t \cdot W_1 \exp(-\psi^R \gamma_{i,j,t})}{\sum_j W_1 \exp(-\psi^R \gamma_{i,j,t})}$$

$$\bar{o}_t = (S_{t+1}^1 - r S_t^1) / P_t \quad \bar{i}_t = (S_{t+1}^1 - r S_t^1) / S_t$$

$S_{i,t}^1$: emploi tertiaire de type 1; W_1 : facteur d'attraction de type 1;
 $S_{j,t}$: somme des emplois de type 1, 2, et 3; B : proportion de la clientèle d'origine résidentielle; S_t^1, P_t, S_t : totaux régionaux.

La valeur du paramètre de transport ψ^S provient d'une calibration avec une matrice de déplacements combinant les motifs d'achat, d'étude et de loisirs. Il interprète le déclin des interactions d'origines résidentielles à motif de consommation de service. Le paramètre ψ^R doit quant à lui symboliser des déplacements similaires mais originant du lieu d'emploi. Une matrice équivalente étant indisponible, la calibration s'est basée sur les voyages originant du centre-ville seulement. Dans cette zone, la

grande majorité des origines réfère à un lieu de travail plutôt qu'au domicile. Les facteurs d'attraction W_i ont par ailleurs été évalués de la même façon que dans le modèle précédent.

Il n'y a pas de données permettant d'évaluer le poids relatif (B et $1-B$) des deux activités localisantes et d'estimer le paramètre autoregressif r . En effet, au moins une autre période aurait été nécessaire. Afin de combler cette lacune, on a choisi de fixer B comme si un résident équivalait à un travailleur dans les services; ainsi $B = P_t / (P_t + S_t)$. Quant à r , différentes valeurs ont été essayées pour prévoir la distribution de l'emploi de type 1 pour 1986. On s'est arrêté à la valeur qui faisait concorder le plus nos prévisions avec celle de Lamonde et al. [1989] à l'échelle de leurs sous-régions. Il ne s'agit pas d'une procédure rigoureuse mais elle permet d'obtenir une valeur raisonnable pour r .

L'emploi tertiaire de type 2, regroupant les secteurs financiers, médicaux et d'administration publique est modélisé avec les mêmes variables. En effet:

$$S_{i,t+1}^2 = r \cdot S_{i,t}^2 + B \cdot \sum_j \sigma_{1,j,t} P_{j,t} + (1-B) \cdot \sum_j \phi_{1,j,t} S_{j,t} \quad (7.13)$$

où $\sigma_{1,j}$ et $\phi_{1,j}$ définis comme dans (7.12).

Toutefois, les valeurs paramétriques et les termes d'attractions diffèrent. Seul le paramètre ψ^S reste inchangé car il exprime encore le déclin des interactions liées à la consommation de service. Le paramètre ψ^F est modifié cette fois car la nature des interactions change. En effet, les interactions entre lieux de travail semblent gouvernées par des motifs d'affaires; la calibration de ψ^F a donc été faite sur des déplacements observés originant du centre-ville et à motifs classés «autres».

Encore une fois, la pondération (B , $1-B$) ne peut être véritablement estimée. Puisqu'il s'agit de la localisation d'une activité tendant à se concentrer dans les pôles d'emplois tertiaires existants, on fixe arbitrairement un poids relatif deux fois plus important aux emplois de service qu'à la population. Le paramètre autorégressif est estimé après coup selon la même procédure que précédemment.

Enfin, les services aux entreprises et les services divers sont considérés comme conditionnés par la localisation des autres emplois. Le modèle recourt donc à l'emploi total comme facteur localisant:

$$S\tilde{E}_{t+1} = r \cdot S\tilde{E}_t + \sum_j \phi_{1,jt} E_{jt} \quad (7.14)$$

$$\text{où } \phi_{1,jt} = \frac{i_t \cdot W_1 \exp(-\psi^r x_{ijt})}{\sum_j W_1 \exp(-\psi^r x_{ijt})} \quad i_t = (S\tilde{E}_{t+1} - r S\tilde{E}_t) / E_t$$

Le paramètre de transport ψ^r reste inchangé par rapport au modèle (7.13). Le reste des paramètres est évalué comme dans les autres sous-modèles de localisation d'emplois.

Le tableau 7.4 donne les valeurs paramétriques utilisées dans les simulations qui seront présentées au chapitre suivant. Ces valeurs ne doivent pas être considérées comme "descriptives" ni comme "universelles". D'une part, étant donné la quasi-absence d'information temporelle, il s'est avéré impossible d'appliquer une méthode rigoureuse d'estimation. On s'est plutôt attardé à calculer des valeurs raisonnables permettant de rendre opérationnel le processus de localisation. D'autre part, les valeurs paramétriques sont propres à la région montréalaise et conditionnées par le découpage zonal utilisé, la catégorisation des activités, par la période d'étude ainsi que par la qualité des données observées.

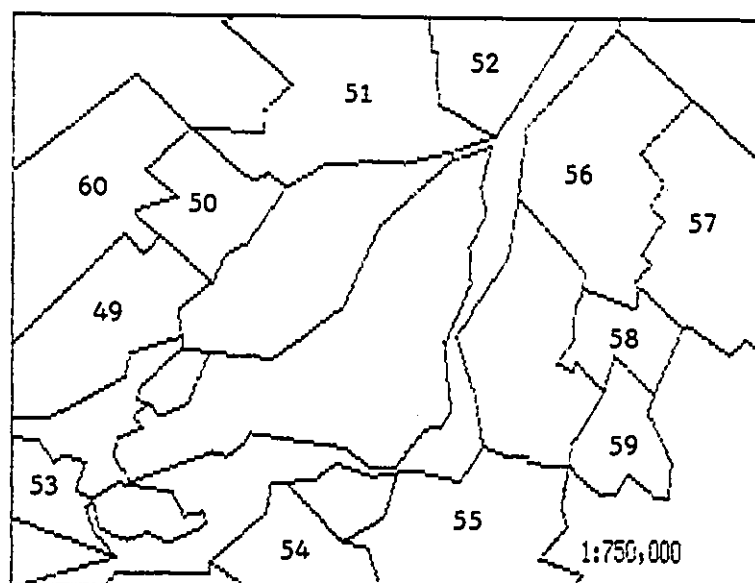
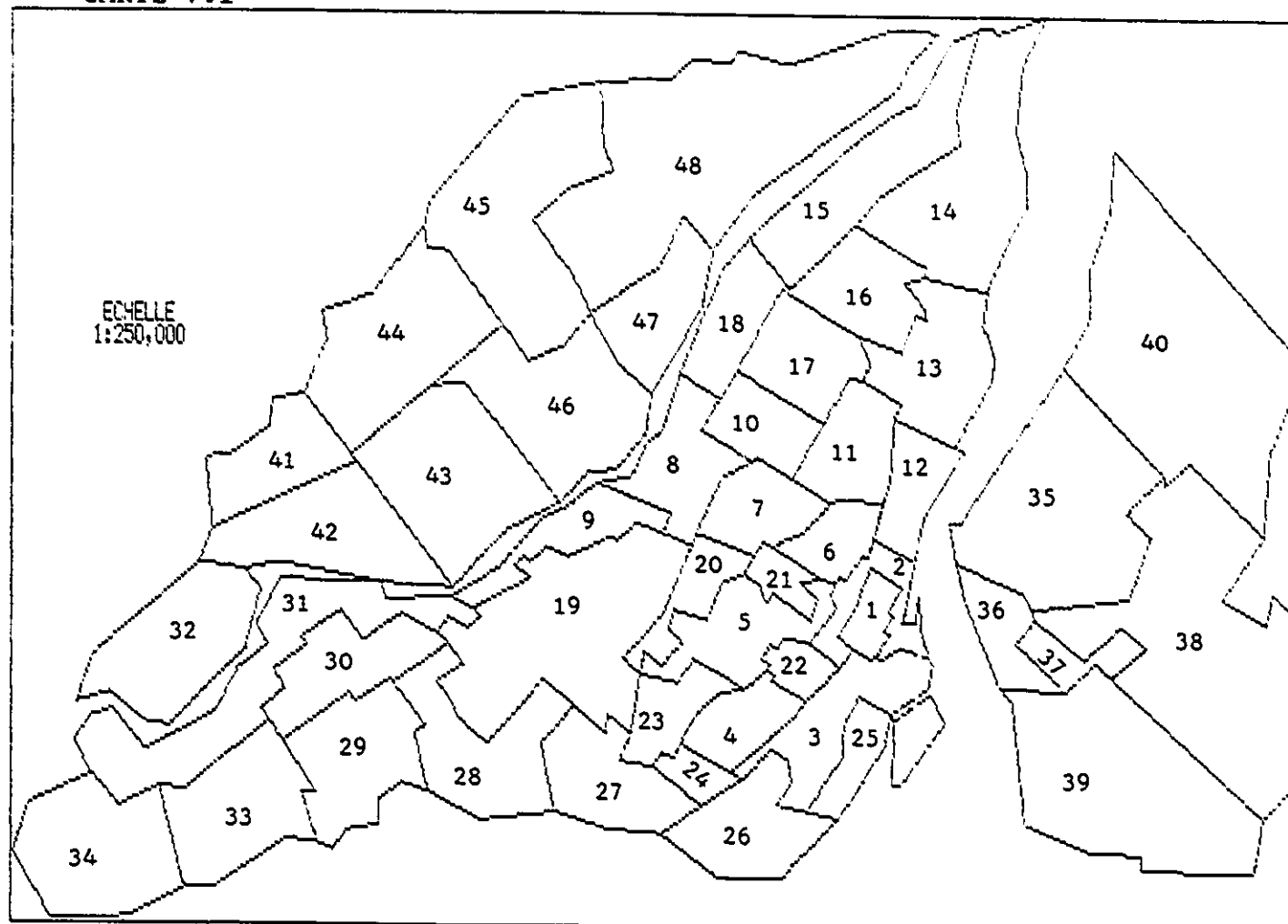
Tableau 7.4 Paramètres des modèles spatio-temporels de localisation

Variables localisées	Paramètres autorégressifs	Variables localisantes	Poids	Paramètres de transport
Population (P)	0.489 (r)	Emploi total (E)	1.0000	0.1451
Emploi type 1 (S ¹)	0.520 (r)	Population (P)	0.8084	0.2454
		Emploi (S) types 1,2,3	0.1916	0.9675
Emploi type 2 (S ²)	0.710 (r)	Population (P)	0.3333	0.2454
		Emploi (S) types 1,2,3	0.6667	0.4917
Emploi type 2 (S ³)	0.710 (r)	Emploi total (E)	1.0000	0.4917

Selon la définition des paramètres autorégressifs temporels (section 5.2.2), ils expriment le degré d'inertie de la variable dépendante. En examinant leurs valeurs sur le tableau 7.4, on pourrait penser que de 30% à

50% des activités sont déplacées à tous les cinq ans; un tel roulement semble irréaliste. En fait, la distribution spatiale des activités localisantes est telle (elle fortement corrélée avec la variable localisée) qu'elle favorise le maintien de la part dite "relocalisable". Les véritables déplacements se produiront que s'il y a éloignement de l'équilibre géographique.

CARTE 7.1



MONTREAL METROPOLITAIN
 NUMERO DES ZONES DE
 LA REGION D'ETUDE
 (VOIR ANNEXE 7.4)

Carte faite par GRAFON

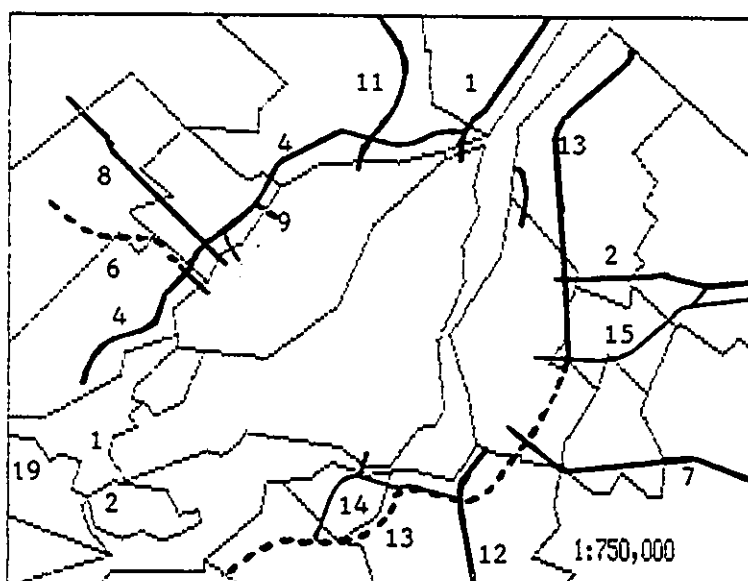
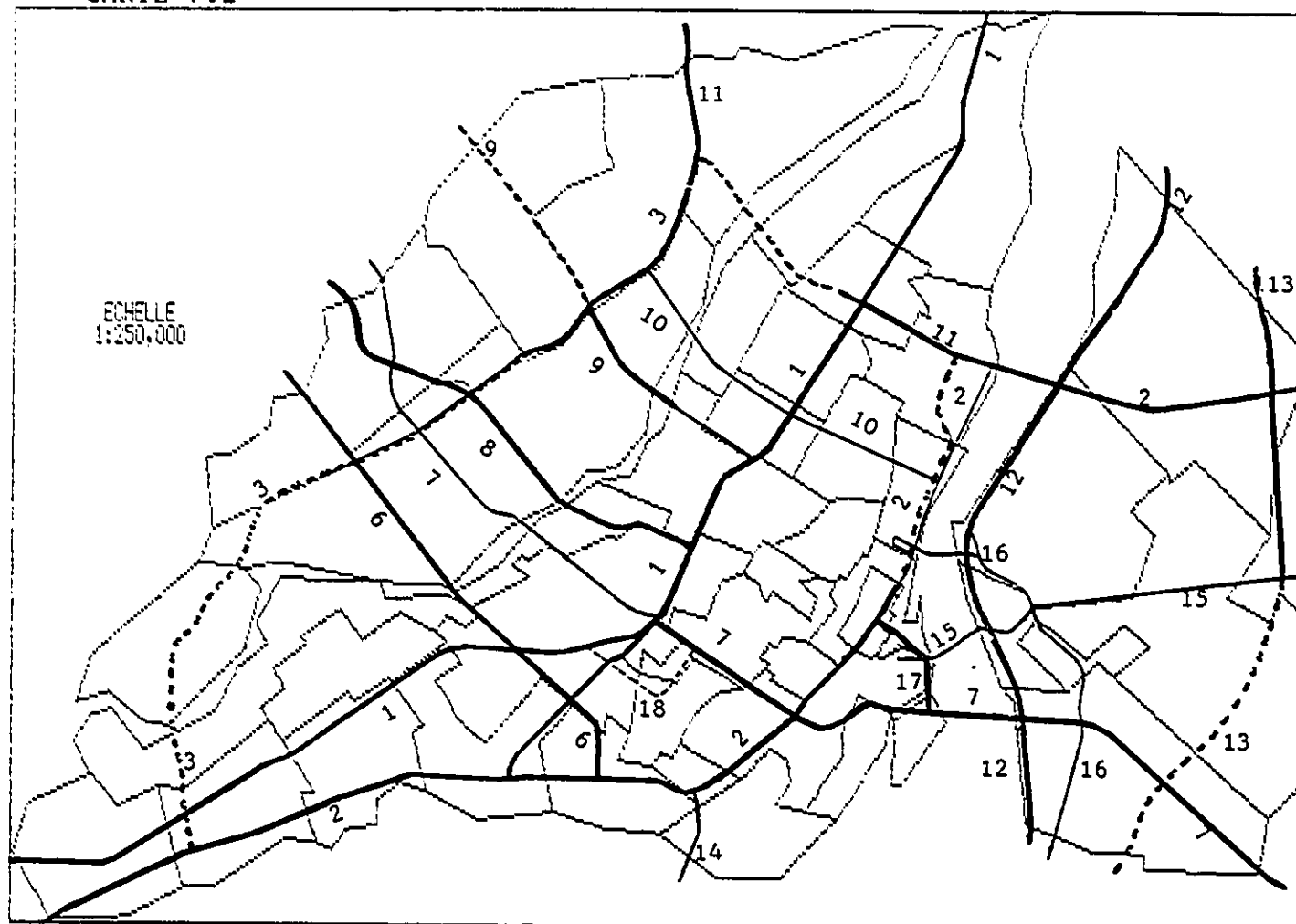
7.A ANNEXE: Liste des zones de la région étudiée (suite)

COURONNES PERIPHERIQUES

- 49 MRC de Deux-Montagnes [Oka, St-Joseph-du-Lac, Pointe-Calumet, Ste-Marthe-sur-le-Lac, St-Eustache, Deux-Montagnes]
- 50 MRC de Sainte-Thérèse * [Ste-Thérèse, Boisbriand, Blainville, Rosemère, Lorraine, Bois-des-Filions]
- 51 MRC des Moulins [Terrebonne, Mascouche, St-Louis-de-Terrebonne, Lachenaie, La Plaine]
- 52 MRC de L'Assomption [Repentigny, Le Gardeur, Charlemaque, Saint-Sulpice, L'Assomption, L'Epiphanie, St-Gérard-Magella]
- 53 MRC de Vaudreuil-Soulanges * [Ile-Perrot, Pointe-du-Moulin, Notre-Dame-de-l'Ile-Perrot, Pincourt, Terrasse-Vaudreuil, Ile-Cadieux, Dorion, Vaudreuil, Vaudreuil-sur-le-Lac, Hudson, St-Lazare, Pointe-des-Cascades]
- 54 MRC de Roussillon (Ouest) [Châteauquay, Lery, Mercier, St-Isidore]
- 55 MRC de Roussillon (Est) [Laorairie, Candiac, Delson, St-Constant, Ste-Catherine, St-Mathieu, St-Philippel]
- 56 MRC de Lajemmerais * [Ste-Julie, Varennes, Verchères, St-Amable, Calixa-Lavallée]
- 57 MRC de la Vallée-du-Richelieu (Nord) [Beloeil, Mont-St-Hilaire, McMasterville, Otterburn Park, St-Mathieu-de-Beloeil, St-Marc-sur-Richelieu, St-Charles]
- 58 MRC de la Vallée-du-Richelieu (Centre) [St-Bruno-de-Montarville, St-Basile-le-Grand]
- 59 MRC de la Vallée-du-Richelieu (Sud) [Chambly, Carignan]
- 60 Mirabel

* partie de la Municipalité Régionale de Comté
[nom1, nom2...] municipalités constituantes

CARTE 7.2



MONTREAL METROPOLITAIN
RESEAU ROUTIER PRINCIPAL
(VOIR ANNEXE 7.B)

- Autoroutes
- Artères
- - - Liens projetés

Carte faite par GRAFON

7.B ANNEXE: Identification du réseau routier retenu - L(1)

no. Noms des tronçons

- 1 Transcanadienne (A-40), Métropolitaine (A-40)
De la Rive Nord (A-40)
- 2 Montréal - Sainte-Anne (A-20), Ville-Marie (A-20¹)
Tunnel L.-H.-Lafontaine (A-20¹), Jean-Lesage (A-20)
- 3 Laval (A-440)
- 4 Mille-Iles (A-440)
- 5 Côte-de-Liesse (A-520)
- 6 Chomedey (A-13)
- 7 Des Cantons de L'Est (A-10), Pont Champlain (A-10 A-15, A-20²)
Décarie (A-15), Laurentien / Labelle (R-117)
- 8 Des Laurentides (A-15)
- 9 Papineau / Leblanc (A-19)
- 10 Pie-IX (R-125)
- 11 L.-H.-Lafontaine (A-25), Montréal - Terrebonne (A-25)
- 12 Marie-Victorin (A-15 R-132), Marie-Victorin (A-20² R-132)
Varennnes (A-430)
- 13 De l'Acier (A-30)
- 14 Pont Mercier (R-138), St-Jean-Baptiste (R-138 R-132)
- 15 Rue Bridge / Pont Victoria (R-112), Sir Wilfrid Laurier (R-112 R-116)
- 16 Taschereau (R-134 R-116), Pont Jacques-Cartier (R-116)
- 17 Bonaventure (A-10)
- 18 Raccordement Cavendish/ Jean-Talon
- 19 Vaudreuil - Dorion (A-540)
- 20 Kahnawake - Laprairie (R-132)

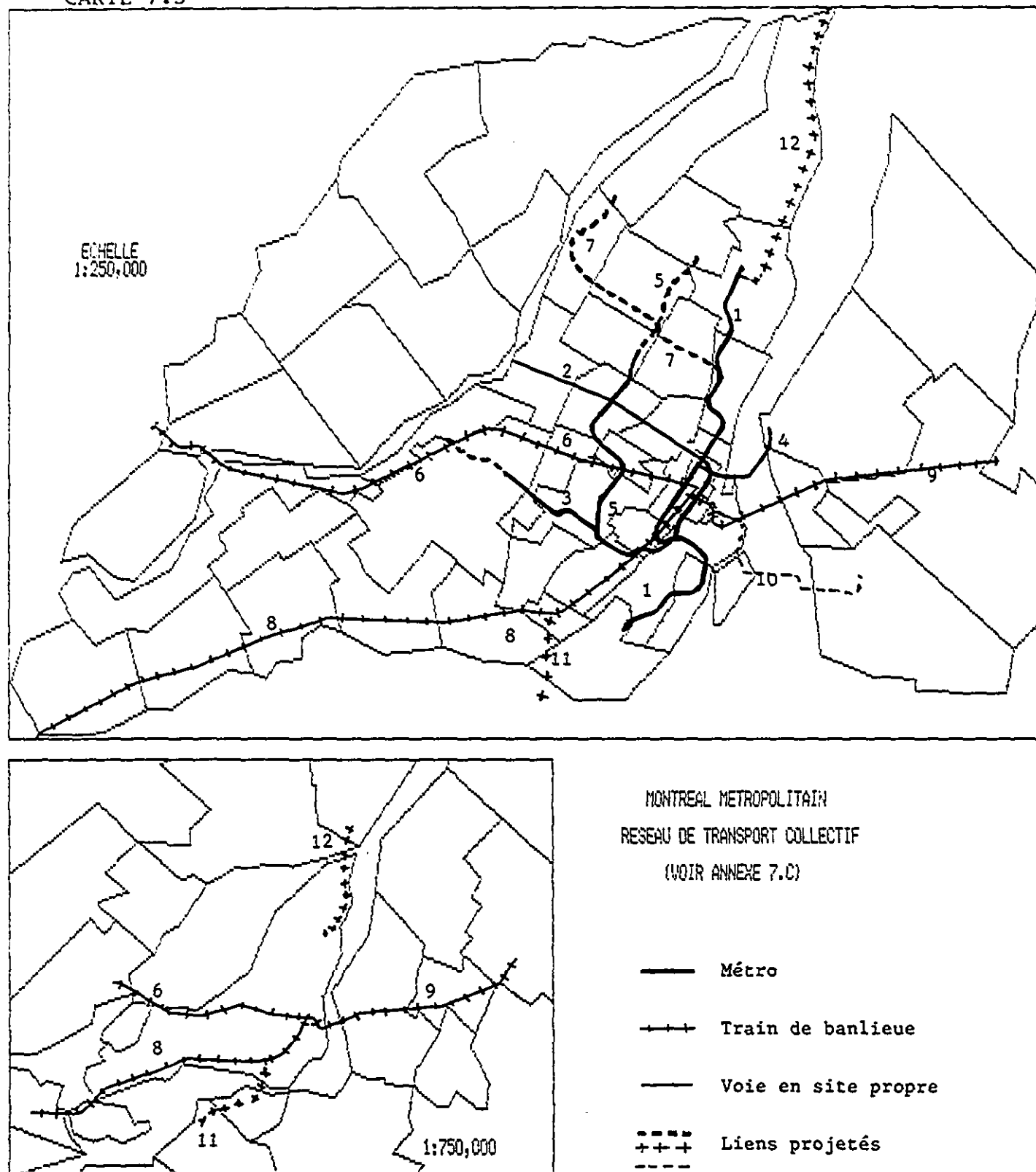
(A-nn): autoroute

(R-nn): route provinciale (peut avoir des caractéristiques autoroutières)

1: tracé éventuel de l'A-20 lorsque l'autoroute Ville-Marie sera complétée

2: trace temporaire de l'A-20 via le pont Champlain et l'autoroute
Marie-Victorin

CARTE 7.3



Carte faite par GRAFON

7.C ANNEXE: Identification du réseau de transport collectif retenu - L(2)

no. Noms des lignes

- 1 Ligne de metro no. 1 [Angrignon - Monore-Beaugrand]
- 2 Ligne de metro no. 2 [Henri-Bourassa - Bonaventure]³
- 3 Ligne de metro no. 2 [Bois-Franc - Bonaventure]³
- 4 Ligne de metro no. 4 [Berri-UQAM - Longueuil]
- 5 Ligne de metro no. 5 [Snowdon - Anjou⁴]
- 6 Ligne ferroviaire [Gare Centrale - Deux-Montagnes]
- 7 Ligne de metro no. 7 [Pie-IX - Maurice-Duplessis]⁵
- 8 Ligne ferroviaire [Gare Windsor - Dorion]
- 9 Ligne ferroviaire [Gare Centrale - St-Hilaire]⁶
- 10 Estacade du Pont Champlain [Terminus Bonaventure - Brossard]⁷
- 11 Ligne ferroviaire [Montréal-Ouest - Châteauguay]⁸

3: La ligne no. 2 est consideree comme deux lignes distinctes car sa forme en "U" occasionne deux correspondances avec les lignes transversales no.1 et no.5. Cela compliquait l'enregistrement des itinéraires.

4: Section St-Michel - Anjou projetée

5: Ligne projetée

6: Ligne en operation en 1981 mais abandonnee en 1988

7: Voie d'autobus en site propre

8: Ligne à l'étude

Chapitre 8

APPLICATION A LA REGION DE MONTREAL

8.1 LES TROIS SCENARIOS DE SIMULATION

8.1.1 Introduction

Les scenarios de simulation se basent sur des hypotheses distinctes de developpement des reseaux de transport dans la region de Montréal. De prime abord, il avait été question d'en tester quatre mais deux scenarios ont pu se fusionner car ils donnaient des résultats équivalents. En effet, les trois premiers scenarios consistaient à optimiser les coûts de transport pour les voyageurs, l'un en modifiant periodiquement les deux reseaux, l'autre, le reseau de transport en commun et le dernier, le reseau routier. Selon le type de mesure de coût, les deux premiers scénarios s'avéraient identiques car il semble que l'amélioration du reseau de transport collectif fait davantage baisser les coûts totaux que celle du reseau routier. Il y a trois explications possibles:

(1) Pour chaque paire O-D, les parcours effectués initialement en autobus sont écourtés davantage lorsque remplacés par un parcours usant d'un système rapide (métro, train).

(2) Les tronçons d'un système rapide sont plus fréquentes que les autoroutes car il sont moins nombreux et parcourent des zones plus denses.

(3) Il n'y a pas d'effet de congestion au sein du reseau public. Les gains de clientèle occasionnés par une expansion du reseau n'affectent donc pas les temps de trajet.

Le quatrième scenario, désormais troisième, devient notre scénario de référence. Il consiste à reproduire le «Plan d'action 1988-1998» proposé par le Ministère des transports du Québec [M.T.Q., 1988] concernant le transport dans la région de Montréal. Il ne s'agit pas d'un plan définitif mais il donne des orientations probables quant à la politique actuelle du gouvernement québécois en matière de transport. Il faut se rappeler que le gouvernement provincial est responsable du financement des autoroutes et des systèmes comme le métro et certains trains de banlieue. L'exploitation des systèmes collectifs n'est toutefois pas de son ressort.

Tout au long de l'analyse les trois scenarios retenus seront identifiés de la manière suivante:

Scenario A: developpement du reseau de transport en commun selon un principe d'optimisation des coûts totaux des voyageurs (tout mode);

Scenario B: developpement du reseau routier selon un principe d'optimisation des coûts totaux des voyageurs (tout mode);

Scenario C: developpement des deux reseaux selon le Plan d'Action 1988-1998 du M.T.Q. tel que défini en août 1988.

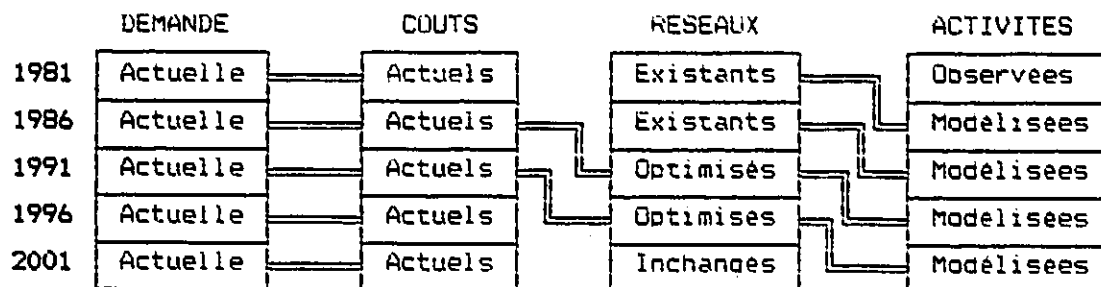
La periode d'analyse s'échelonne de 1981 à 2001. Pour l'année de référence, 1981, on utilise généralement des données observées. Seule l'affectation sur les reseaux est modelisee. Pour 1986, on connaît la population et les modifications des reseaux de transport. Les emplois sont donc alloués selon l'accessibilite de 1981 tandis que la demande est modelisee en regard du reseau existant 1986. Ainsi, pour 1981 et 1986, il n'y aucune difference de resultats au sein des trois scenarios.

En fait, les reseaux ne sont modifiés qu'en 1991 et 1996. Cette limitation à deux periodes s'explique par les decalages quinquennaux (fixes par hypothèse), d'une part, entre l'évaluation du svstème de transport et la realisation d'un projet et, d'autre part, entre la nouvelle configuration de l'accessibilité et ses effets sur la localisation des activités. Par ailleurs, le plan de transport du M.T.Q. ne prévoit que deux echeances, 1993 et 1998, qu'on ramènera pour les besoins de l'analyse à 1991 et 1996.

Pour les besoins de l'aplication, les budgets de developpement sont limites à 350 millions de dollars par cinq ans (valeur de 1988). En réalité, il se dépensera d'avantage d'argent pour l'amelioration du transport des personnes dans la région metropolitaine. Mais une part substantielle ira au chapitre de la refection et l'elargissement des routes, du réaménagement d'installations fixes, d'achat de materiel roulant. Si bien qu'en définitive, le montant qu'on octroie ici pour l'expansion seulement des reseaux semble vraisemblable.

Le schéma suivant montre la séquence de modélisation sur l'axe du temps. Un note pertinemment les effets réducteurs des décalages. En effet, seuls la demande et les coûts de transport procèdent à l'intérieur d'une même période. En fait, il se passe 10 ans entre le moment où une modification de réseau est prescrite et son effet sur la distribution spatiale des activités.

Figure 8.1 Séquence de modélisation



8.1.2 Description du scénario A

Le scénario A consiste à modifier le réseau de transport en commun pour 1991 et 1996 de manière à minimiser respectivement les totaux 1986 et 1991 des coûts de transport. Le sous-ensemble des liens projetés (LP) est alors restreint aux tronçons de métro et de train de banlieue. Même si aucune limite de capacité a été imposée au réseau public, le modèle de conception de réseau est employé intégralement. En effet, le sous-problème de demande doit être résolu afin d'établir l'affectation sur les deux réseaux et les coûts conséquents (par exemple, le transfert modal occasionné par l'amélioration du réseau L(1) peut réduire la congestion du réseau L(2)).

Le tableau suivant énumère les segments de lignes compris dans le sous-ensemble des liens projetés. Les lignes sont représentées sous forme de segments indépendants signifiant qu'elles peuvent être réalisées en partie. On y trouve une évaluation grossière des coûts d'immobilisation selon des dollars de 1988. La plupart des chiffres proviennent du sommaire des coûts contenu dans le plan ministériel de transport [M.T.O., 1988]. Pour les tronçons de métro exclus du plan, on a appliqué un coût/kilomètre semblable. Quant au métro régional desservant Pointe-aux-Trembles, l'évaluation provient du rapport du B.T.M [1984] et a été actualisée pour 1988.

Tableau B.1 Liste des liens projetés de transport en commun L⁽²⁾

- Ligne de métro 2	(Côte-Vertu à Bois-Franc)	100 M\$
- Ligne de métro 5	(St-Michel à Boulevard)	40 M\$
- Ligne de métro 5	(Boulevard à Anjou)	140 M\$
- Ligne de métro 7	(Pie-IX à Boulevard)	150 M\$
- Ligne de métro 7	(Boulevard à Leger)	160 M\$
- Ligne de métro 7	(Leger à Maurice-Duplessis)	80 M\$
- Ligne ferroviaire	(Montréal-Ouest à Châteauguay)	35 M\$
- Estacade du pont Champlain	(voie en site propre du terminus STRSM de Bonaventure à Brossard)	85 M\$
- Ligne ferroviaire	(Gare Centrale à St-Bruno)	70 M\$
- Ligne ferroviaire	(St-Bruno à Beloeil)	70 M\$
- Métro régional	(Radisson à Pointe-aux-Trembles)	100 M\$

La ligne ferroviaire «Gare Centrale - Beloeil» était en opération jusqu'en août 1988 et est aujourd'hui abandonnée: elle fait partie des "projets" puisque sa remise en service reste plausible. La ligne de banlieue «Gare Centrale -Deux-Montagnes», bien qu'elle requiert aujourd'hui un investissement majeur pour sa modernisation, est cependant exclue de la liste puisqu'elle existe déjà.

Le tronçon de la ligne de métro no.7 entre les stations «Boulevard» et «Léger» serait converti en ligne no.5 advenant le cas où la section «Pie-IX - Boulevard» de la ligne no.7 soit rejetée. Il se rabouterait à la section «St-Michel - Boulevard» de la ligne no.5. Dans cet éventualité la station «Boulevard» perdrait son titre de lieu de correspondance.

Les hypothèses potentiellement vérifiables par une simulation de ce type apparaissent comme suit:

a) Le transfert modal enregistré sur certains trajets O-D bénéficiant de l'expansion du réseau peut être annulé par des pertes réalisées ailleurs à cause d'une baisse de congestion du réseau concurrent.

b) Le développement du transport en commun entrepris à l'instar d'un gel du développement des routes peut ralentir substantiellement l'étalement urbain.

c) Le rythme de l'étalement urbain supprime les gains de clientèle suscités par l'expansion des systèmes rapides de transport collectif.

La vérification de l'hypothèse "a" tient à la différence entre gain marginal de clientèle (causé par une baisse de coût sur le réseau L(2)) et la réduction marginale des coûts variables sur le réseau L(1) causée par une diminution du trafic automobile. La relation entre les deux apparaît

extrêmement complexe puisqu'il peut s'agir d'un nouvel équilibre, non seulement au sein d'un petit nombre de trajet O-D, mais à travers la région. Cette relation dépend, outre de la sensibilité des fonctions de répartition modale et de congestion, des différences initiales et résultantes entre coûts modaux des paires O-D directement concernées ainsi que de celles dont les parcours profitent de la décongestion d'une partie du réseau routier. En fait, l'affectation routière de la région peut être complètement modifiée.

Les hypothèses comme "b" et "c" semblent intimement liées car toutes deux traitent de la "confrontation" entre l'étalement urbain et le transport collectif. L'une d'elle stipule que les systèmes de transport public améliorent surtout la desserte des zones densément urbanisées où d'ailleurs ils peuvent offrir un service plus compétitif. Cette hausse de l'accessibilité locale favorise alors une consolidation du territoire. C'est enfin que souhaitent les responsables de l'aménagement de la C.U.M. [1989] et de la Ville de Montréal. Le succès de cette lutte repose également sur un moratoire des prolongements autoroutiers.

La dernière hypothèse aborde par contre l'étalement urbain sous l'angle d'un phénomène peu contrôlable par des mesures de transport. Pire encore, sa vigueur est telle qu'elle anéantit les hausses de clientèle que devrait normalement susciter le développement des systèmes collectifs de transport. Cette éventualité semble très plausible puisque les zones périphériques sont généralement moins bien desservies (à cause de leurs faibles densités) et privilégient l'usage de l'auto. Par surcroît, l'étalement que connaît la région de Montréal ne touche pas que l'activité résidentielle. Les lieux d'emplois se décentralisent si bien que les trajets O-D se multiplient. Le transport en commun, surtout compétitif dans les corridors achalandés et dans les trajets destinés au centre-ville, ne peut plus concurrencer les véhicules privés dans les parcours inter-suburbains.

8.1.3 Description du scénario B

Le scénario B semble complémentaire au précédent. Il consiste à optimiser les améliorations du réseau routier seulement. Tout comme dans l'autre cas, le problème de conception reste néanmoins bi-modal. Aussi, le

réseau optimal est obtenu selon la demande de la période courante tandis que les modifications prescrites ne deviennent effectives que cinq ans plus tard.

Les projets routiers potentiels s'avèrent particulièrement de grande envergure car ils semblent pouvoir répondre à certains besoins précis. L'agglomération montréalaise est aujourd'hui quadrillée par un important réseau autoroutier. Cependant, plusieurs raccordements stratégiques manquent encore si bien que le réseau semble inégalement utilisé. En dépit de son étendue, le réseau souffre de congestion chronique à cause principalement de la sous-capacité de certains nœuds et tronçons névralgiques. Enfin, les camionneurs partagent tant bien que mal la chaussée avec les navetteurs. Plusieurs d'entre eux, pourtant, ne sont que de "transit" mais doivent se résoudre à circuler dans le cœur industriel de Montréal car aucune voie de contournement n'est encore complétée.

Tableau 8.2 Liste des liens routiers projetés: L*(1)

Autoroute de l'Acier, A-30 (A-10 à A-15M)	40 M\$
Autoroute de l'Acier, A-30 (A-15M à R-138)	15 M\$
Autoroute Laval, A-440 (R-335 à A-13) ¹	14 M\$
Autoroute Laval, A-440 (A-13 à Ile-Bizard) ²	60 M\$
Autoroute Laval, A-440 (Ile-Bizard à A-20) ³	85 M\$
Autoroute L-H-Lafontaine, A-25 (A-40 à H-Bourassa)	10 M\$
Autoroute L-H-Lafontaine, A-25 (H-Bourassa à A-440) ⁴	115 M\$
Autoroute Chomedey, A-13 (A-640 à Mirabel)	78 M\$
Autoroute Papineau, A-19 (A-440 à Dagenais)	5 M\$
Autoroute Papineau, A-19 (Dagenais à A-640) ⁵	45 M\$
Boulevard Ville-Marie, A-20 (Frontenac à A-25) ⁶	65 M\$
Raccordement Cavendish/Jean-Talon (A-520 à A-15D) ⁷	40 M\$

1) aménagement des voies rapides (voies de service existantes)
 2) comporte un pont sur les rapides Lalemant (près de l'île Bigras)
 3) comporte un pont sur la rivière des Prairies
 4) comporte un pont sur la rivière des Prairies
 5) comporte un pont sur la rivière des Mille-Îles
 6) transformation de la rue Notre-Dame en boulevard urbain
 7) comporte un viaduc enjambant la cour de triage Côte-St-Luc

Dans notre sous-ensemble de liens projetés, les élargissements et réfections de la chaussée sont exclus. Aussi, pour simplifier la sélection, les alternatives sont agrégées en tronçons puisque l'existence de certains liens est conditionnée par celle de son ou de ses voisins.

Les autoroutes de l'Acier (A-30) et Laval (A-440) représentent des voies possibles de contournement. L'autoroute 30 serait prolongée de l'autoroute des Cantons de l'Est (A-10) à l'autoroute Montréal - Toronto (A-20). Elle permettrait aux véhicules provenant du sud de l'Ontario

d'éviter l'île de Montreal pour rejoindre l'est du Québec. La section entre Chateauguay et Vallevfield est exclue puisqu'en dehors de la région métropolitaine.

Le prolongement de l'autoroute Laval joindrait directement l'A-13 à l'A-40 et l'A-20. Elle permettrait de dévier la circulation est-ouest de l'île de Montreal vers le territoire lavallois. Toutefois, le retour sur l'île se ferait via l'autoroute projetée L-H-Lafontaine (A-25). Cette voie de contournement fait craindre une concurrence accrue des parcs industriels de Laval envers ceux de l'Ouest de la CUM.

Il faut noter tout de suite que la fonction de contournement ne pourra être mise en évidence dans notre étude. En effet, les données de déplacement ne comprennent pas le trafic interurbain. Seule la circulation inter-suburbaine pourra être affectée sur ces prolongements. Le soulagement des axes de pénétration se révélera donc moins.

Le raccordement de l'autoroute 25, actuellement discontinue, sollicite une enjambée de la rivière des Prairies. Les navetteurs originant de la MRC des Moulins (Terrebonne, Mascouche) pourrait aisément joindre l'île de Montreal en évitant le pont Pie-IX. Le nouveau tronçon faciliterait aussi le passage entre Rive Sud et Rive Nord car l'A-25 aboutit au pont-tunnel L-H-Lafontaine. Les autorités de la C.U.M. craignent cependant que ce projet compromette l'expansion du quartier Rivière-des-Prairies en encourageant l'étalement résidentiel au profit de l'est de Laval (St-François) et de la MRC des Moulins.

Le projet de prolongement de l'autoroute Chomedey (A-13) vise essentiellement à améliorer l'accès à l'aéroport international de Mirabel. Il prétend remédier à la liaison aéroportuaire déficiente avec Dorval et le centre-ville de Montreal. Aussi, l'itinéraire actuel empruntant l'autoroute des Laurentides (A-15) semble fort congestionné puisqu'il confond le trafic suburbain et celui de villégiature (très important en fin de semaine).

L'autoroute Papineau (A-19) pourrait être prolongée afin de relier les autoroutes 440 et 640. Elle offrirait un chemin nord-sud alternatif à l'autoroute des Laurentides et au boulevard homonyme. Cependant, une autoroute 19 allongée connaîtrait une augmentation de son trafic qui deviendrait préjudiciable à sa portion située sur le territoire montréalais où la voie rapide aboutit sur un boulevard urbain.

La tristement célèbre autoroute Ville-Marie (A-20) devrait être complétée sous forme d'un boulevard urbain à haute capacité. Un court tronçon entre les rues St-Denis et Frontenac fut mis en service en 1987. Il reste à joindre l'autoroute L.-H.-Lafontaine en suivant principalement l'axe de la rue Notre-Dame. Le boulevard/autoroute Ville-Marie est vu comme une voie alternative à la Métropolitaine.

La dernière alternative consiste à raccorder la rue Jean-Talon Ouest au boulevard Cavendish. Actuellement, celui-ci reste incomplet, la section sud (Côte-St-Luc) ne rejoint pas la section nord (St-Laurent). Le prolongement de la rue Jean-Talon se ferait en surface tandis que le raccordement du boulevard Cavendish nécessiterait la construction d'un long viaduc d'au moins 500 mètres pour franchir une cour de triage. L'intérêt de ce raccordement est d'offrir une voie d'évitement à l'échangeur «Décarie/Métropolitain» fortement congestionné. Cette déviation est cependant contestée par les autorités municipales de Côte-St-Luc qui appréhendent un accroissement du trafic sur le boulevard Cavendish.

8.1.4 Description du scénario C: le plan "Côté"

L'orientation générale du ministère des Transports est d'améliorer la mobilité des citoyens en privilégiant trois types d'intervention soit:

- la conservation des infrastructures et du matériel en place lorsque leur état nécessite des travaux pour maintenir le service offert et accroître la sécurité des usagers;
- l'amélioration des réseaux pour en réduire le coût d'exploitation ou d'entretien et donner un meilleur service;
- le développement de nouvelles infrastructures pour améliorer l'équilibre des réseaux et rendre les modes de transport privé et collectif complémentaires.

Dans son plan d'action, le ministre Côté se défend de favoriser le développement de l'un des deux modes de transport ni même de privilégier un certain schéma d'aménagement du territoire. En fait, le plan apparaît à première vue comme un "plan neutre". D'une part, les dépenses accordées aux réseaux routiers et de transport collectif se comparent (842 M\$ vs 742 M\$). D'autre part, tous les secteurs de la région métropolitaine bénéficient de la "manne" gouvernementale.

Une lecture plus approfondie du plan nous apprend toutefois que cette équité est superficielle et trompeuse. D'abord, la parité des budgets à l'égard des réseaux paraît peu significative. L'argent dépensé pour le réseau routier se répartit dans trois chapitres différents, la conservation des infrastructures (réfection des chaussées et des ouvrages d'art), les améliorations (élargissement des chaussées, réaménagement des échangeurs) et le développement (prolongement routier, aménagement des carrefours). Les projets de conservation n'affectent pas l'accessibilité tandis que ceux d'amélioration ont un impact très ponctuel sur la circulation proprement dite. Seuls les projets dits de développement modifient suffisamment la mobilité pour que la structure urbaine puisse être changée.

Les dépenses attribuées aux transports publics se répartissent aussi dans des chapitres semblables mais le contenu des projets se compare difficilement avec celui du mode concurrent. La conservation des équipements concerne surtout le matériel roulant (ce qui ne s'applique pas dans le cas du réseau routier) alors que les améliorations se manifestent par des aménagements de gares et de garages d'autobus. Les projets de développement se traduisent en prolongement de lignes de métro et en implantation de voie réservée.

L'allocation des ressources financières et les effets sur la mobilité sont peu liés si bien qu'il devient difficile de parler d'équité au sein des réseaux. En effet, le coût du kilomètre de métro (40 à 50 M\$) semble prohibitif lorsqu'on le compare avec un prolongement autoroutier dans une zone périurbaine. Aussi, les obstacles à franchir (cours d'eau par exemple) font dramatiquement augmenter les coûts d'immobilisation sans pour autant étendre le réseau en question.

Quant à la répartition territoriale des projets, il s'agit d'un concept purement politique qui a peu de sens en matière de transport. D'abord, certaines infrastructures, surtout autoroutières, ont la particularité de desservir davantage les zones périphériques que celles dans lesquelles les projets sont implantés. Ensuite, si une répartition territoriale a tout de même lieu, elle doit se fonder sur les besoins spécifiques des sous-régions et non en respect de leurs poids démographiques et économiques.

L'apparente neutralité du plan d'action du M.T.Q. devint la cible des critiques. Pour les tenants du transport en commun, investir de façon équivalente dans les deux modes revient à privilégier l'usage de l'auto puisque celle-ci jouit d'une popularité croissante (les données préliminaires de l'enquête Q-D 1987 en font foi). Une baisse relative de la clientèle augure mal pour les sociétés de transport déjà au prises avec d'importants déficits d'exploitation. Mais le plan fut davantage dénoncé pour son manque de considération en matière de développement urbain. Les responsables de la C.U.M. et de la ville de Montréal [Barcelo et al. 1988] sont très soucieuses du phénomène de l'étalement en particulier de celui de l'activité résidentielle. Pour eux, le moindre prolongement autoroutier élargit le périmètre urbain davantage et facilite l'exode des citadins vers la périphérie. Même si le réseau de métro prend simultanément de l'expansion, les autoroutes paraissent plus persuasives car elles "mettent sur le marché", en leur facilitant l'accès, de vastes territoires vacants. Le langage ministériel reconnaît implicitement cet état de fait en considérant les autoroutes comme des outils de développement régional et en admettant que "le transport collectif remplit une mission sociale importante, assurant une mobilité de base à la population de la région" [M.T.Q., 1988 p.15].

Le scénario C ne retient dans le plan de transport du M.T.Q. que les développements et améliorations de réseau qui peuvent conditionner le développement urbain. Le plan prévoit sommairement un échéancier de réalisation étalé sur dix ans (1988-1998) et y distinguent deux années de mise en service, soit 1993 et 1998. Pour fin de comparaison avec les scénarios A et B, nous avons devancé de deux ans l'échéancier même si, techniquement, les projets ne pourraient être réalisés si hâtivement.

Voici un tableau énumérant les projets retenus dans le scénario C et la période de mise en service.

Tableau 8.3 Projets de transport du «Plan d'Action 1988-1998»

PROLONGEMENTS ROUTIERS	1991	1996	M\$
Boulevard Ville-Marie, A-20 (Frontenac à A-25)		X	65
Autoroute Chomedev, A-13 (A-640 à Mirabel)	X		78
Autoroute Papineau, A-19 (Dagenais à A-640)	X		45
Autoroute L-H-Lafontaine, A-25 (A-40 à H-Bourassa)	X		10
Autoroute L-H-Lafontaine, A-25 (H-Bourassa à A-440)		X	115
Autoroute de l'Acier, A-30 (A-10 à A-20)		X	167
Autoroute Laval, A-440 (R-335 à A-13)	X		14
AMELIORATIONS ROUTIERES			
Boul. Sir Wilfrid Laurier, R-112-116	X		23
Autoroute des Laurentides, A-15	X		12
Autoroute Jean-Lesage, A-20	X		4
PROLONGEMENTS, RESEAU DE TRANSPORT COLLECTIF			
Métro Ligne 5 (St-Michel à Boulevard)	X		50
Métro Ligne 5 (Boulevard à Leger)		X	150
Métro Ligne 2 (Côte-Vertu à Bois-Franc)		X	100
Train de banlieue de Châteauguay	X		35
Voie réservée, estacade du Pont Champlain	X		85

8.2 COMMENTAIRES SUR LES SIMULATIONS

8.2.1 Développement des réseaux

Dans les scénarios A et B, les réseaux sont prolongés selon un principe d'optimisation des coûts de transport inscrit dans le modèle de conception. Les seules distinctions résident au sein de leurs sous-ensembles de liens projetés. Quant au scénario de référence, C, il décrit un développement des réseaux selon les vœux du plan d'action du Ministère des transports du Québec en 1988.

Comme il a été dit précédemment, les réseaux ne sont modifiés que sur deux périodes soit 1986-1991 et 1991-1996. La demande correspondant à l'année débutant le quinquennat sert à évaluer les alternatives d'expansion des réseaux tandis que celle de l'année finissante est celle de la mise en service.

Scénario A

Selon la demande de transport estimée en 1986 et un budget de 350 M\$, la combinaison des projets suivants surpasse tous les autres quant à la réduction des coûts de transport:

1) Prolongement de la ligne de métro no.5 (St-Michel à Boulevard)	50 M\$
2) Implantation de la ligne de métro no.7 (Pie-IX à Léger)	300 M\$
Total	350 M\$

Le premier projet modifie très peu l'accessibilité puisqu'il s'agit de l'ajout d'une seule station. Il autorise par contre une correspondance indispensable avec la nouvelle ligne nord-sud no.7. Mis ensemble, ces deux projets de transport collectif semblent améliorer substantiellement les conditions de déplacement dans l'est de la CUM.

Les projets optimaux selon la demande de 1991 concernent encore une fois le réseau de métro:

1) Prolongement de la ligne de métro no.2 (Côte-Vertu à Bois-Franc)	100 M\$
2) Prolongement de la ligne de métro no.5 (Boulevard à Anjou)	150 M\$
3) Prolongement de la ligne de métro no.7 (Léger à M.-Duplessis)	80 M\$
Total	330 M\$

La sélection exclusive des projets de prolongement de métro ne semble pas le fruit du hasard. D'abord, le réseau de métro ne sillonne que des secteurs densément occupés. Les gains de temps qu'il occasionne profitent donc à un grand nombre de voyageurs potentiels. Ensuite, en recrutant une clientèle auparavant automobiliste, le métro permet d'alléger la circulation routière de la ville centrale.

Le tableau 8.4 expose les résultats globaux entourant la modélisation du réseau de transport en commun.

Tableau 8.4 Statistiques globales des simulations: scénario A

DEMANDE DE TRANSPORT	1981	1986	1991	1996	2001	
Nombre de déplacements	2417834	2545481	2545481	2692505	2800067	2890916
Partage modal (%): automobile	62.23	61.80	60.96	61.60	61.10	61.68
Partage modal (%): transport public	37.77	38.20	39.04	38.40	38.90	38.02
Coût total du système ¹	28117171	29640678	29304075	31304997	31108202	32691638
Coût moyen par voyageur	11.62	11.64	11.51	11.63	11.55	11.68
Coût moyen par voyageur: mode 1	11.23	11.21	11.02	11.16	11.03	11.13
Coût moyen par voyageur: mode 2	12.28	12.33	12.28	12.37	12.38	12.47
RESEAUX DE TRANSPORT (selon la période de mise en service)	1981	1986	1991	1996		

Chacune des colonnes du tableau se rattache soit à une demande affectée sur un réseau de la même période, soit à une demande affectée sur le réseau prescrit pour la prochaine période. Dans le second cas, il s'agit des résultats issus de la solution du problème de conception de réseau: les nombres en caractères gras désigne la valeur de la fonction objective minimisée.

Sans surprise, si on reste au sein d'une même période, l'expansion du réseau de transport public tend à réduire les coûts totaux de transport. Cependant, certains résultats étonnent comme le fait que les prolongements de métro font davantage décroître le coût moyen des automobilistes que celui des usagers du transport collectif. Cette anomalie peut s'expliquer par la conjugaison de deux facteurs:

- sur certains trajets O-D, usage plus grand du transport public;
- sur d'autres (plus périphériques), on profite d'une baisse de la congestion.

Si le coût marginal de la congestion est plus élevé que le bénéfice occasionné par la venue du métro, il en résultera (théoriquement) un gain de temps plus grand pour les automobilistes. Il ne s'agit pas d'une entorse à la relation offre/demande mais plutôt d'une manifestation de la complexité du transport urbain. En effet:

- le partage modal s'effectue au sein de chaque trajet O-D;
- pour les deux modes en présence, l'affectation des itinéraires s'appuie sur des principes différents;
- les liens du réseau routier sont empruntés par un nombre variable d'itinéraires: indirectement, la demande sur tous les trajets O-D est interreliée.

Un autre type de résultat peut aussi nous confondre. On constate peu d'améliorations au sein du transport collectif lorsqu'on compare les résultats évalués lors de la "recommandation" du réseau avec ceux obtenus lors de la mise en service. Il faut comprendre que d'une période à l'autre la demande est conditionnée par une organisation spatiale des activités évoluant vers un certain étalement. Celui-ci semble annuler les gains réalisés en terme de partage modal et même accroître les coûts moyens des usagers.

Sans doute, l'étalement persistant agit à la baisse sur la demande des trajets O-D favorables au transport en commun et augmente la part relative des trajets non-compétitifs. En surcroît, les distances parcourues tendent globalement à allonger même sur les parcours faits en automobiles.

Scénario B

Avec un même budget de 350 M\$ et selon la même demande de transport de 1986, les tronçons autoroutiers prescrits sont énumérés ci-dessous:

1) Boulevard Ville-Marie, A-20 (Frontenac à A-25)	65 M\$
2) Autoroute L.-H.-Lafontaine, A-25 (A-40 à A-440)	125 M\$
3) Autoroute Laval, A-440 (construction des voies rapides)	14 M\$
4) Autoroute Laval, A-440 (A-13 à A-20)	145 M\$
Total	349 M\$

On remarque que les projets retenus coïncident, comme dans le cas précédent, avec ceux qui parcourent les zones les plus centrales. Le boulevard Ville-Marie longe des anciens quartiers montréalais alors que l'autoroute Louis-Hyppolyte-Lafontaine joint les îles de Montréal et Jésus. L'autoroute Laval, bien qu'à vocation suburbaine, dessert néanmoins plusieurs zones relativement urbanisées.

La sélection prioritaire des projets plus "centraux" s'explique d'abord par une plus forte demande potentielle venant des zones adjacentes. De plus, ils offrent de nouvelles alternatives aux tronçons les plus congestionnés. Par leur centralité au sein du réseau, ils conviennent enfin à un grand nombre d'itinéraires.

Le problème de conception de réseau nous a offert une solution "triviale" pour 1991 puisque le budget accordé excédait la valeur totale des projets restants. Voici les projets ayant eu droit d'entrée:

1) Autoroute de l'Acier, A-30 (A-10 à R-138)	55 M\$
2) Autoroute Chomedey, A-13 (A-640 à Mirabel)	78 M\$
3) Autoroute Papineau, A-19 (A-440 à A-640)	50 M\$
4) Raccordement Cavendish/Jean-Talon	40 M\$
Total	223 M\$

Pour des raisons de comparabilité avec le scénario rival (A), nous avons choisi de dépenser les millions restants dans trois projets d'amélioration inscrits dans le "plan Côté". Il s'agit d'élargir les autoroutes 15 (d'Henri-Bourassa à St-Martin) et 20 (de Mortagne à l'autoroute 30) et de transformer la route 116 en voie autoroutière à la hauteur de la ville de St-Hubert (voir section 8.1.4). Ces projets totalisent 39 millions de dollars.

Tableau 8.5 Statistiques globales des simulations: scénario B

DEMANDE DE TRANSPORT	1981	1986	1991	1996	2001		
Nombre de déplacements	2417834	2545481	2545481	2692505	2890067	2890916	
Partage modal (%): automobile	62.23	61.80	62.37	62.37	63.10	63.21	63.43
Partage modal (%): transport public	37.77	38.20	37.63	37.63	36.90	36.79	36.57
Coût total du système ¹	28117171	29640678	29453075	31141963	31032653	33137638	33934487
Coût moyen par voyageur	11.62	11.64	11.57	11.57	11.53	11.83	11.74
Coût moyen par voyageur: mode 1	11.23	11.21	11.11	11.06	10.99	11.44	11.28
Coût moyen par voyageur: mode 2	12.28	12.33	12.34	12.41	12.41	12.51	12.53
RESEAUX DE TRANSPORT (selon la période de mise en service)	1981	1986	1991	1996			

La structure du tableau est la même que celle du tableau 8.4 i.e. la période de la demande de transport est indiquée au-dessus tandis que le réseau sujet à l'affectation est identifié au-dessous.

Comme il fallait s'y attendre, la part de marché du transport en commun est en diminution constante. Déjà la tendance à l'étalement urbain défavorise l'usage du service collectif. L'expansion du réseau routier avantage d'une part l'automobiliste en écourtant ses temps de parcours (introduction de nouveaux itinéraires, augmentation globale de la capacité du réseau) et, d'autre part, contribue au développement de la couronne sub-urbaine.

En examinant de plus près les chiffres du tableau 8.5, on décèle une accélération du phénomène de l'étalement dont les effets sur les coûts dépassent les bénéfices encourus par les développements routiers. Comme indices, il suffit de comparer la hausse des coûts moyens par voyageurs en auto du scénario B avec celle du scénario précédent. Les améliorations routières font étonnement grimper plus vite ce coût que ne le font les prolongements du métro! L'hypothèse la plus plausible serait que l'étalement accéléré accroît la fréquentation de trajets O-D plus longs. Ceux-ci, moins sujets à la concurrence du mode public, assurent une clientèle captive à l'automobile mais augmentent les temps moyens de déplacements.

Ces observations mettent en évidence le décalage entre le processus de développement du réseau de transport et celui des activités urbaines. Même si le premier vise à diminuer "optimalement" les coûts de déplacement, il met plutôt en place une organisation spatiale favorisant une hausse des dépenses de voyageant. De l'autre côté, même si les activités se reloca-

lisent en tirant profit de l'accessibilité, elles ne peuvent considérer leurs impacts sur la circulation et les changements qu'elles provoquent sur la configuration des voyages O-D.

Scenario C

Les projets de développement de transport inscrits dans le scénario C ne font pas l'objet d'une sélection par optimisation et sont énumérés au tableau 8.6. La valeur totale des projets ne se compare avec les budgets alloués dans les scénarios précédents puisqu'il s'agit d'une sélection des projets ayant le plus d'impact sur l'organisation urbaine. La facture des projets retenus n'est donc pas représentative du plan d'action du MTQ dont les volets de conservation et d'amélioration des infrastructures, omis dans nos simulations, comptent néanmoins pour une grande part des dépenses.

Malgré l'exogénéité du procédé de sélection des projets, un tableau équivalent à 8.4 et 8.5 a été dressé. Seules les colonnes relatives aux réseaux prescrits manquent.

Tableau 8.6 Statistiques globales des simulations: scénario C

DEMANDE DE TRANSPORT	1981	1986	1991	1996	2001
Nombre de déplacements	2417834	2545481	2692505	2800067	2890916
Partage modal (%): automobile	62.23	61.80	62.24	62.68	62.81
Partage modal (%): transport public	37.77	38.20	37.76	37.32	37.19
Coût total du système ¹	28117171	29640678	31460637	32856911	33954444
Coût moyen par voyageur	11.62	11.64	11.68	11.73	11.75
Coût moyen par voyageur: mode 1	11.23	11.21	11.22	11.26	11.23
Coût moyen par voyageur: mode 2	12.28	12.33	12.44	12.53	12.61
RESEAUX DE TRANSPORT (selon la période de mise en service)	1981	1986	1991	1996	2001

A certains égards, le plan ministériel de transport semble un scénario "intermédiaire" en comparaison des deux autres. Alors que ceux-ci n'envisagent que des projets exclusifs à un seul mode, le plan du MTQ prévoit des développements au sein de chaque réseau. Cette tentative d'équité modale se répercute au niveau du partage de la clientèle; elle demeure relativement stable quoique avec un léger avantage à l'auto. La faible augmentation de la part du mode privé doit s'expliquer en partie par la tendance lourde à l'étalement.

Le coût moyen par voyageur augmente sensiblement bien que le niveau de 2001 se compare aisément avec celui obtenu dans le scénario B. Par contre, son évolution par mode diffère dans la mesure où le coût des automobilistes augmente faiblement alors que celui des usagers du transport collectif connaît une hausse plus importante. Alors comment expliquer le quasi maintien de la clientèle du 2^e mode? La réponse se trouve encore dans la distribution spatiale des déplacements. Si les trajets gagnant en fréquentation se situent près des corridors mieux desservis par ce mode mais sont plus longs que la moyenne antérieure, une croissance simultanée de la clientèle et du coût moyen peut coexister.

Les observations faites à partir des statistiques globales paraissent avant tout hypothétiques. Elles recourent à des explications faisant appel à des phénomènes spatiaux sans s'appuyer sur une quelconque analyse géographique. Dans les prochaines sections, les résultats simulatoires seront examinés soit cartographiquement, soit selon une tabulation spatialisée. Les résultats devraient être plus probants sous ces formes qu'au sein de statistiques globales dans lesquelles ils apparaissent un peu trop dilués.

La section suivante abordera le thème de l'évolution des activités urbaines. Il s'agit de voir comment la localisation de la population et des emplois est affectée par la mise en service des projets en lice. Les analyses devraient permettre, espère-t-on, d'apprécier le rythme de déconcentration spatiale et de discriminer les scénarios à cet égard. Ensuite, on examinera les implications en terme de répartition modale. Les représentations cartographiques localiseront les gains et pertes de clientèle modale. L'analyse se révélera intéressante puisqu'il n'est pas sûr que les projets de transport affectent le partage modal seulement là où ils se situent. Dans la dernière section, on se penchera sur les conséquences sur les coûts et la congestion. Etant donné la complexité des réseaux et l'organisation métropolitaine évolutive, ces effets sont loins d'être prévisibles et pourraient même devenir contre-intuitifs.

8.2.2 Localisation future des activités

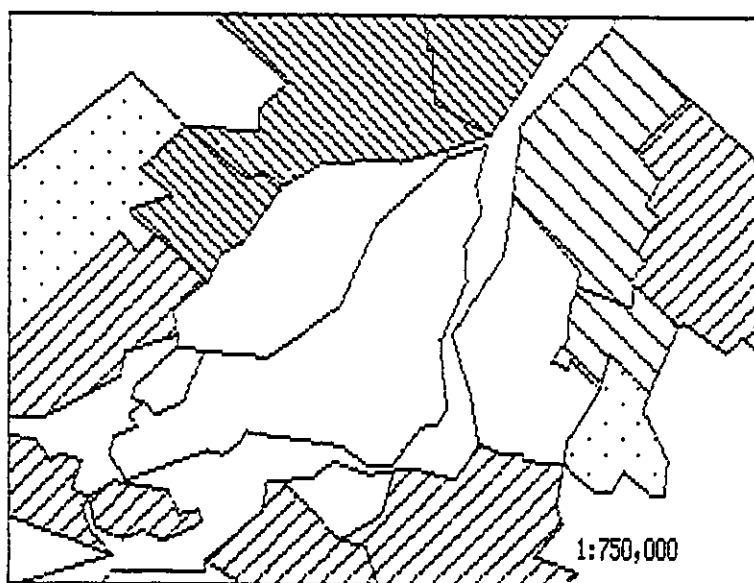
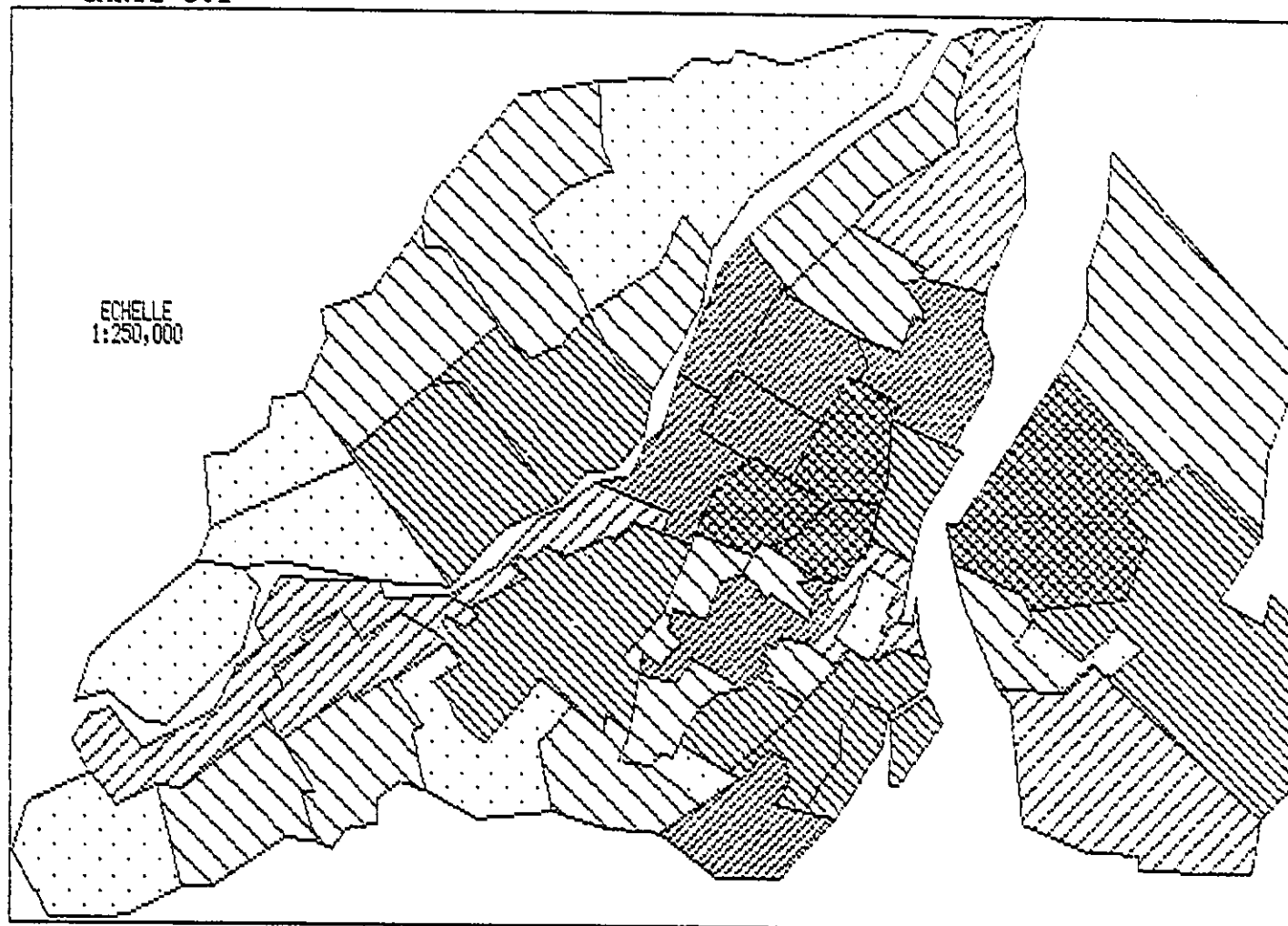
Le premier volet de l'analyse comparative des simulations concerne la distribution spatiale des activités car celle-ci conditionne passablement les résultats relatifs à la demande de transport.

La répartition géographique de la population et des emplois pour l'année de départ (1981) est illustrée sur les cartes 8.1 et 8.2. Les valeurs sont exprimées en part relative afin de comparer les deux distributions.

Sans aucune surprise, les zones les plus peuplées sont presque toutes situées en deça d'un certain périmètre au centre de l'île de Montréal. Ce périmètre déborde quelque peu sur la Rive Sud et dans ville de Laval. Les principaux foyers se groupent principalement à l'est (Villeray, Plateau Mont-Royal, Rosemont, Longueuil) tandis que les banlieues enclavées (Westmount, Outremont, Ville Mont-Royal) forment des "dépressions" dans le centre-ouest. Il faut garder en tête que la carte 8.1 n'est pas une carte de densité et que, par conséquent, les grandes superficies tendent à une sur-représentation des valeurs. Malgré tout, les municipalités régionales de comté (en cartouche) comptent, en dépit de leurs grandes tailles, peu de zones peuplées.

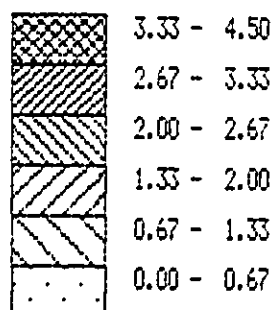
La carte suivante représente la part relative des zones dans l'emploi. Cette fois-ci, la concentration apparaît plus évidente et tend plus vers l'ouest. Les grandes zones d'emploi comptent outre le centre-ville et sa périphérie, les quartiers de Côte-des-Neiges, de Villeray et d'Achuntsic et la ville de St-Laurent. La carte dissimule cependant la prédominance du centre-ville qui accapare plus de 17% des emplois de la région métropolitaine. D'autres zones comme Ville Mont-Royal, Lachine, Dorval et Pointe-Claire affichent des valeurs plus élevées à l'égard de l'emploi qu'à celle de la population. Pour le reste, hormis les banlieues précitées, les zones suburbaines correspondent davantage à des lieux de résidence qu'à des lieux de travail.

CARTE 8.1



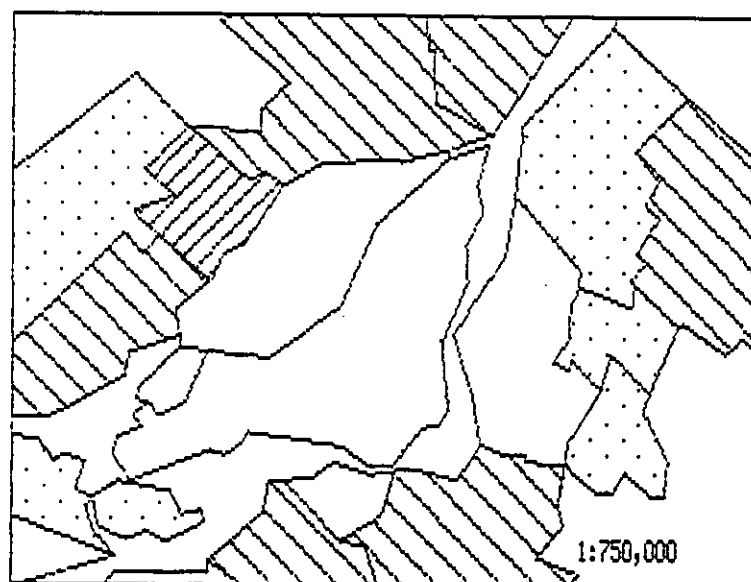
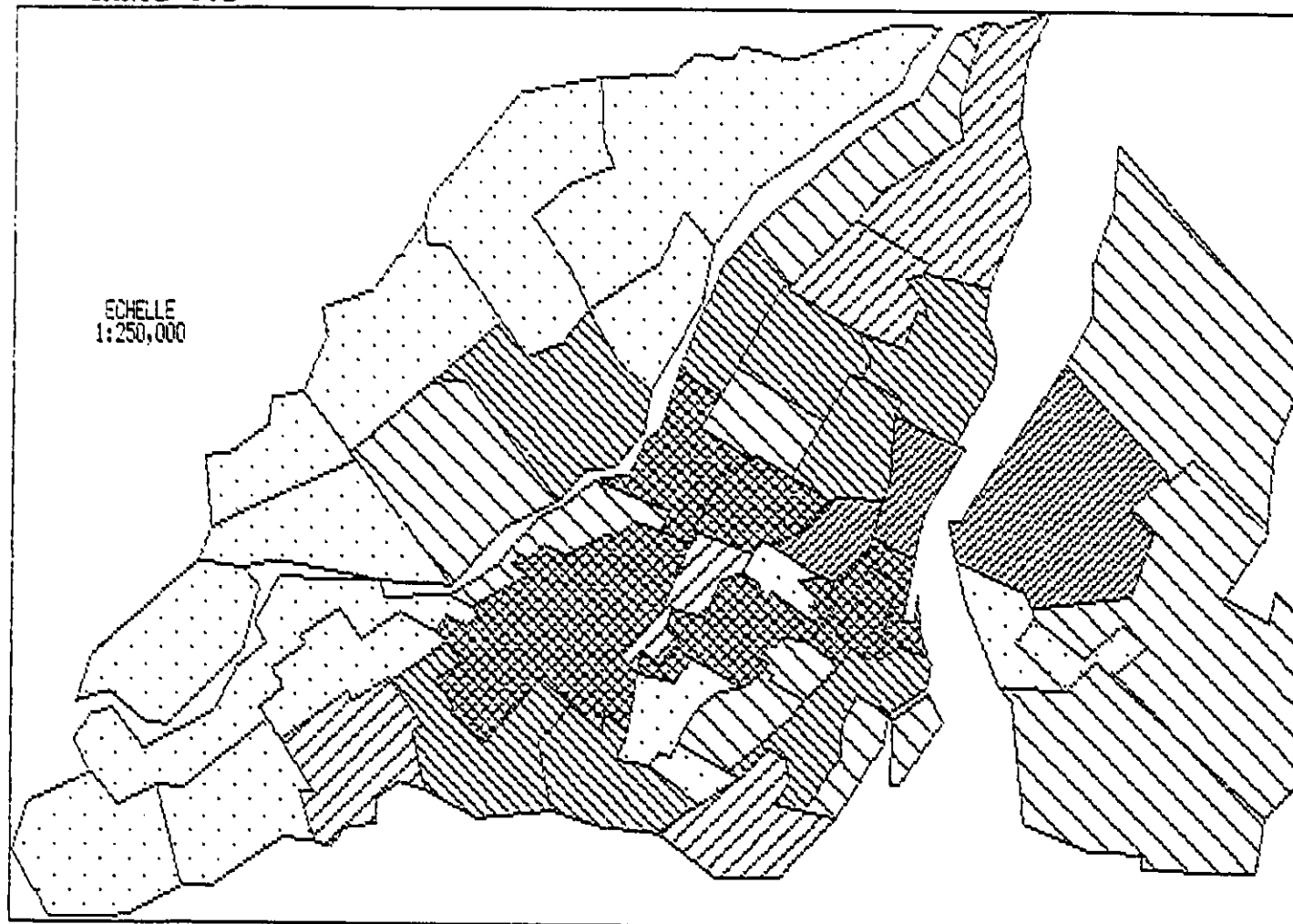
POPULATION 1981

PART RELATIVE (%)



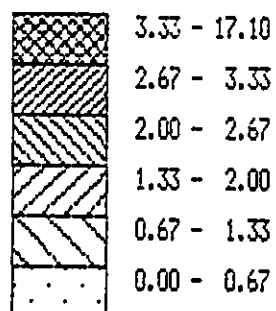
Carte faite par GRAFON

CARTE 8.2



EMPLOI TOTAL 1981

PART RELATIVE (%)



Carte faite par GRAFON

Les comparaisons entre les simulations se feront premièrement à l'aide de tableaux dressant les effectifs projetés jusqu'à 2001 pour huit sous-régions. Ensuite, une analyse cartographique sera faite afin d'identifier clairement les zones bénéficiaires des développements de réseaux.

Les sous-régions désignées correspondent au découpage classique de la CUM (voir section 7.2.6) bien que nous y avons agrégé les secteurs "Centre-Ouest" et "Sud-Ouest" puisque une de nos zones, Montréal-Ouest/St-Pierre, se partage dans chacune d'elles. Le découpage distingue aussi la ville de Laval, la Rive Nord périphérique, la Rive Sud proche et la Rive Sud périphérique.

Le premier tableau, 8.7, montre l'évolution projetée de l'emploi exogène pour les années 1981-2001. Comme on sait, ces projections sont invariables selon les scénarios et proviennent du rapport de Lamonde et al. [1989]. La majorité des emplois de ce type se situe dans la Communauté Urbaine de Montréal (83% de la RMM en 1981). Toutefois, ils connaîtront un déclin constant jusqu'à la fin de la période de prévision principalement à cause de la chute des emplois manufacturiers. À l'intérieur de la CUM, les différences sont notables; la sous-région CUM-Centre, l'ancien cœur manufacturier, semble responsable de la totalité des pertes. Les autres sous-régions, spécifiquement ceux à l'ouest, affichent une certaine vigueur. Partout hors de la CUM, les emplois dits exogènes sont appelés à croître durant la période mais surtout durant la première moitié.

L'évolution démographique simulée est exposée dans le tableau suivant. Cette fois, les valeurs sont ventilées pour les trois scénarios à partir de 1996 i.e. cinq ans après les premières modifications de réseau. Le profil général est que les sous-régions tendent à conserver leur dynamique du départ mais que les rythmes se modèrent en avançant dans le temps. Ainsi, les sous-régions les plus anciennement urbanisées perdent peu à peu de leurs poids démographiques tandis que celles de type suburbain sont toujours en croissance.

Tableau 8.7 Evolution de l'emploi exogène⁽¹⁾ des sous-régions selon les prévisions de Lamonde et al. [1988]

SOUS-REGIONS	Obs.		Prévisions ⁽²⁾			Taux annuel (%)	
	1981	1986	1991	1996	2001	1981-91	91-2001
CUM Centre (1-13, 21-22)	204.7	182.1	173.1	158.2	145.6	-1.664	-1.715
CUM Est (14-18)	46.1	45.2	46.0	47.3	47.2	-0.006	0.258
CUM Centre/Sud-Ouest (19-20, 23-27)	94.1	98.9	107.4	113.4	115.7	1.329	0.751
CUM Ouest (28-34)	32.3	35.0	39.8	43.0	44.6	2.108	1.143
Proche Rive Sud (35-40)	26.4	27.4	28.5	29.6	29.9	0.772	0.470
Ville de Laval (41-48)	15.1	17.4	19.6	21.1	21.9	2.655	1.133
Rive Nord Périph. (49-52, 60)	18.8	21.3	23.6	25.4	26.2	2.293	1.063
Rive Sud Périph. (53-59)	16.8	18.7	19.6	20.4	20.8	1.559	0.566
Communauté Urbaine de Montreal	377.2	361.2	366.3	361.9	353.1	-0.292	-0.365
RMM hors CUM	77.1	84.8	91.3	96.6	98.8	1.707	0.790
Total RMM	454.3	446.1	457.6	458.5	452.0	0.074	-0.125

(1) Emplois manufacturiers, des services publics et de commerce de gros

(2) Pour le secteur manufacturier, les données de 1986 sont issues d'une banque de l'INRS-Urbanisation

Tableau 8.8 Evolution démographique des sous-régions selon les scénarios A, B et C

SOUS-REGIONS	Scén.	Observations		Simulations			Taux annuel (%)	
		1981	1986	1991	1996	2001	1981-91	1991-2001
CUM	A				942.6	933.5		-0.271
Région Centre (1-13, 21-22)	B	1005.7	981.6	959.2	936.3	924.7	-0.473	-0.365
	C				940.3	929.2		-0.317
CUM	A				304.1	316.0		0.916
Région Est (14-18)	B	273.2	283.2	286.4	299.8	296.9	0.545	0.289
	C				294.0	304.4		0.538
CUM Région Centre/Sud-Ouest (19-20, 23-27)	A				298.1	298.9		-0.018
	B	306.2	302.5	299.4	295.8	298.6	-0.224	-0.028
	C				297.0	295.5		-0.133
CUM	A				204.9	216.3		1.037
Région Ouest (28-34)	B	175.0	184.8	195.1	207.7	217.8	1.091	1.107
	C				204.6	215.3		0.992
	A				338.3	347.2		0.394
Proche Rive Sud (35-40)	B	312.1	324.2	333.8	337.3	345.2	0.676	0.338
	C				344.6	349.8		0.468
	A				313.2	325.3		0.827
Ville de Laval (41-48)	B	268.3	284.2	299.6	318.7	337.8	1.108	1.209
	C				313.8	334.2		1.101
	A				337.7	359.6		1.415
Rive Nord Périphérique (49-52, 60)	B	258.8	285.2	312.5	342.3	368.3	1.904	1.658
	C				340.3	368.8		1.670
	A				313.6	332.6		1.208
Rive Sud Périphérique (53-59)	B	255.3	274.4	294.9	314.4	335.3	1.452	1.291
	C				317.8	336.6		1.330
Communauté Urbaine de Montréal	A				1749.7	1764.6		0.128
	B	1760.1	1752.1	1742.1	1739.6	1742.4	-0.103	0.002
	C				1735.8	1739.8		-0.013
	A				1302.8	1364.7		0.956
Région Métropolitaine hors CUM	B	1094.5	1168.0	1240.8	1312.8	1386.7	1.263	1.118
	C				1316.6	1389.4		1.137
Total RMM		2854.6	2920.1	2982.9	3052.4	3129.2	0.441	0.480

Tableau B.9 Evolution de l'emploi de service⁽¹⁾ des sous-régions selon les scénarios A, B et C

SOUS-REGIONS	Scen.	Observations		Simulations			Taux annuel (%)	
		1981	1986	1991	1996	2001	1981-91	1991-2001
CUM	A				439.1	449.1		0.602
Région Centre (1-13, 21-22)	B	364.0	393.6	423.0	439.2	448.3	1.514	0.583
	C				439.0	449.4		0.608
CUM	A				60.4	64.5		1.453
Région Est (14-18)	B	42.4	48.9	55.9	60.5	64.5	2.792	1.439
	C				60.7	64.0		1.372
CUM Région Centre/Sud-Ouest (19-20, 23-27)	A				86.3	90.0		1.039
	B	65.2	73.1	81.2	86.4	90.2	2.211	1.060
	C				86.3	89.9		1.025
CUM	A				46.0	49.7		1.821
Région Ouest (28-34)	B	30.3	35.6	41.5	46.1	50.2	3.171	1.936
	C				46.0	49.6		1.816
	A				63.0	67.1		1.575
Proche Rive Sud (35-40)	B	42.4	49.6	57.4	63.0	67.2	3.087	1.583
	C				62.9	67.3		1.606
	A				56.4	61.4		1.956
Ville de Laval (41-48)	B	36.2	42.8	50.6	56.2	61.6	3.390	1.998
	C				56.6	61.7		2.004
	A				52.5	59.6		2.902
Rive Nord Périphérique (49-52, 60)	B	27.1	35.4	44.7	52.3	59.4	5.142	2.876
	C				52.4	59.4		2.877
	A				50.4	56.2		2.535
Rive Sud Périphérique (53-59)	B	28.1	35.5	43.7	50.4	56.2	4.535	2.538
	C				50.3	56.2		2.540
Communauté Urbaine de Montréal	A				631.8	653.4		0.830
	B	502.0	551.3	601.5	632.3	653.2	1.825	0.828
	C				631.9	653.0		0.825
Région Métropolitaine hors CUM	A				222.3	244.3		2.201
	B	133.8	163.3	196.5	221.8	244.4	3.919	2.208
	C				222.2	244.6		2.217
Total RMM		635.8	714.6	798.0	854.2	897.6	2.299	1.184

(1) Emplois de type 1, 2 et 3

Les différences entre scénarios peuvent paraître à première vue peu évidente. C'est le cas notamment des sous-régions au sud et à l'ouest de la région métropolitaine dans lesquels se trouvaient peu de projets de transport majeurs. Par contre, dans la CUM-Est, Laval, la couronne Nord et, dans une certaine mesure, la CUM-Centre, les écarts en l'an 2001 nous semblent significatifs. Les disparités les plus claires se manifestent entre les scénarios A et B. En gros, le scénario A de développement exclusif du transport collectif favorise le maintien de la population au sein de la CUM. Le scénario alternatif, même s'il contient des projets sis sur l'île, profite davantage aux zones suburbaines. Des résultats intermédiaires sont obtenus pour le scénario C de référence.

Les trois catégories d'emplois dont la répartition a été subseqüemment modélisée sont ici regroupées sous le nom d'emplois de service. Leur évolution est décrite dans le tableau 8.9. Un simple coup d'oeil permet de constater la croissance soutenue de ce secteur à travers la région métropolitaine. Les sous-régions périphériques témoignent cependant d'un dynamisme supérieur à celles plus centrales. Il faut remarquer que les croissances les plus fortes coïncident là où les effectifs de départ étaient les moins nombreux. Malgré tout, il s'agit ni plus ni moins d'une déconcentration de l'emploi.

Le phénomène d'étalement de l'emploi de service reste compréhensible puisque sa localisation est en bonne part conditionnée par celle de la population. On remarquera par ailleurs que l'utilisation des scénarios affecte peu les valeurs simulées. Alors comment interpréter l'étalement observé? Les paramètres inscrits dans les différents modèles de localisation permettent d'expliquer l'indifférence vis-à-vis les scénarios:

1) Contrairement au modèle résidentiel, les trois modèles d'emploi sont dotés de paramètres de transport relativement élevés; seule la contribution des zones très proches compte dans le processus d'allocation. Les modifications de l'accessibilité deviennent alors peu significantes.

2) Il y a une inertie géographique du fait qu'il se trouve beaucoup de variables d'emploi parmi les variables "localisatrices".

3) La convergence vers l'équilibre favorise un étalement persistant mais spatialement indifférencié.

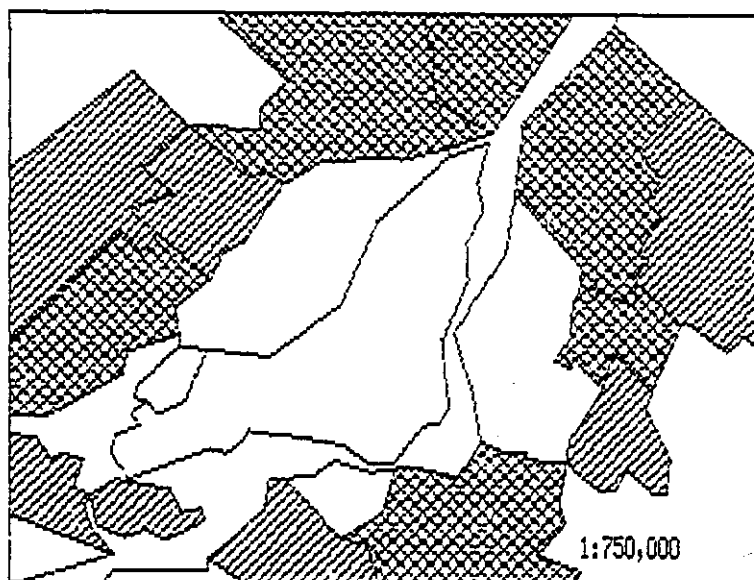
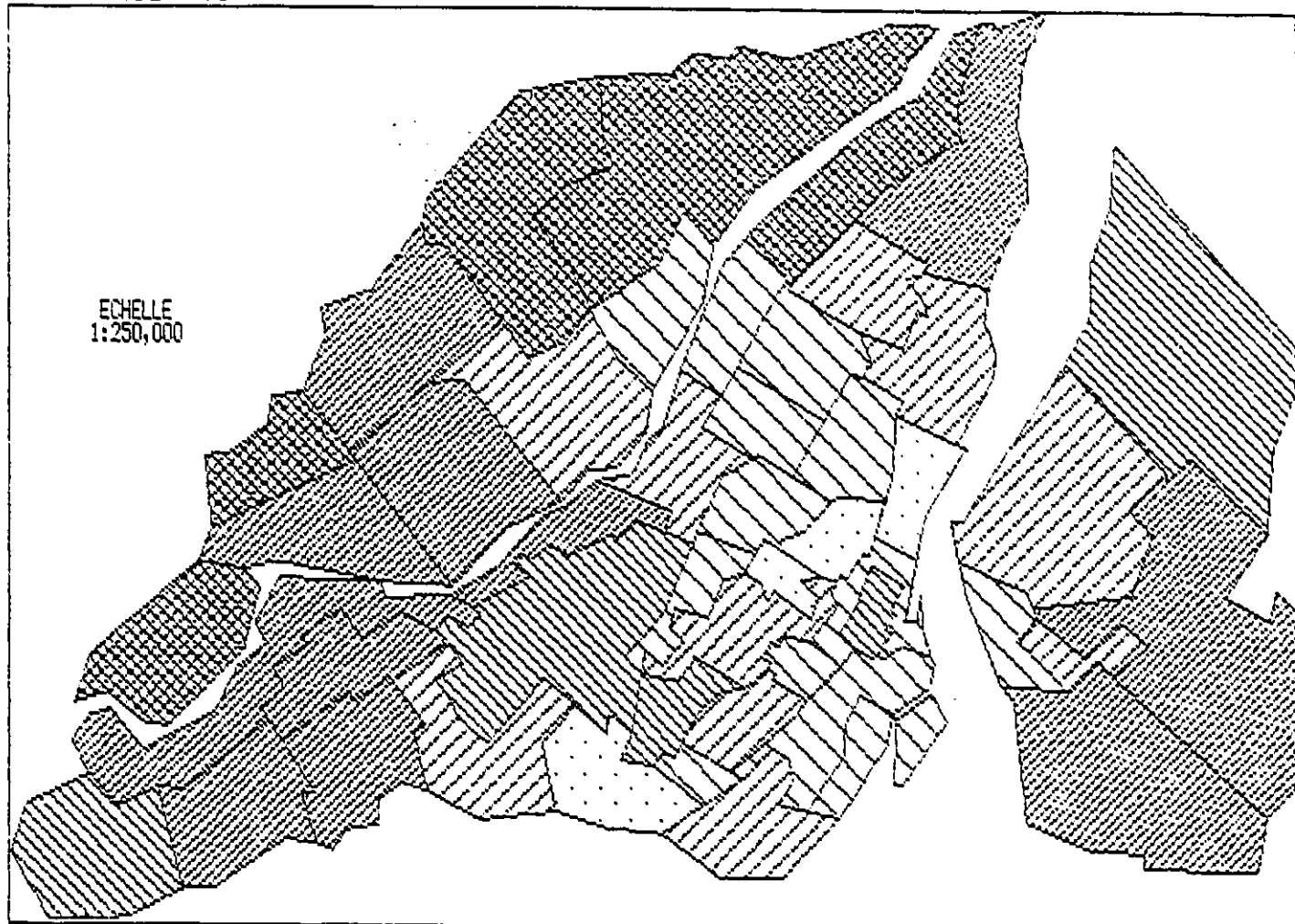
L'analyse cartographique permettra néanmoins d'appréhender, à une échelle plus désagrégée, les différences évolutives de la population et de l'emploi de service. Les cartes 8.3 et 8.4 illustrent les taux annuels moyens de croissance de la période complète 1981-2001 pour le scénario de référence "C". Quant aux résultats des autres scénarios, ils sont exprimés en écart du scénario de référence. Cette représentation a été choisie afin de faire ressortir les différences entre scénarios de transport.

L'examen de la carte 8.3 montre clairement que la dispersion de la fonction résidentielle continuera de plus belle. Les zones les plus centrales connaîtront un déclin de la population quoique les taux apparaissent plutôt modérés. À l'instar, les couronnes périurbaines accaparent le gros de la croissance démographique. À cause des nombreux projets de transport localisés dans l'est de la région métropolitaine (ligne 5 de métro, raccordement de l'A-25), les zones comme Rivière-des-Prairies, St-François, Vimont-Auteuil et la MRC des Moulins bénéficient d'une meilleure accessibilité à l'emploi. Toutefois, cette hausse semble peu profiter aux zones du secteur est dont les projets sillonnent pourtant le territoire.

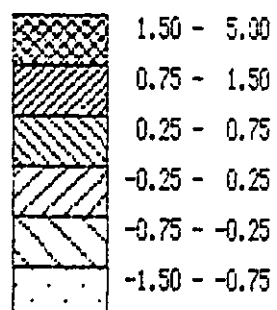
La carte représentant la variation des emplois de service n'a pas, contrairement à la carte précédente, de valeur négative. Elle montre néanmoins un aspect semblable i.e. une croissance plus forte en périphérie que dans la partie centrale. Les zones correspondant au centre-ville et sa périphérie, forte concentration d'emplois tertiaires, affichent les plus faibles performances. L'asymétrie favorable au côté est devient plus apparent bien qu'elle touche peu les zones de l'île de Montréal.

La paire de cartes 8.5-8.6 expriment les écarts entre le scénario "pro-transport collectif" A et celui proposé par le MTQ. En examinant la première carte, il semble bien que l'expansion du métro secondée par un moratoire des prolongements autoroutiers inhibent la déconcentration résidentielle. La majorité des zones de la CUM affiche un écart nettement positif. Les grands bénéficiaires sont situés à même les prolongements proposés dans le scénario A: St-Laurent avec la ligne 2-Ouest, Montréal-Nord et Rivière-des-Prairies avec la ligne 7, Anjou avec la ligne 5 et St-Léonard avec les deux dernières.

CARTE 8.3

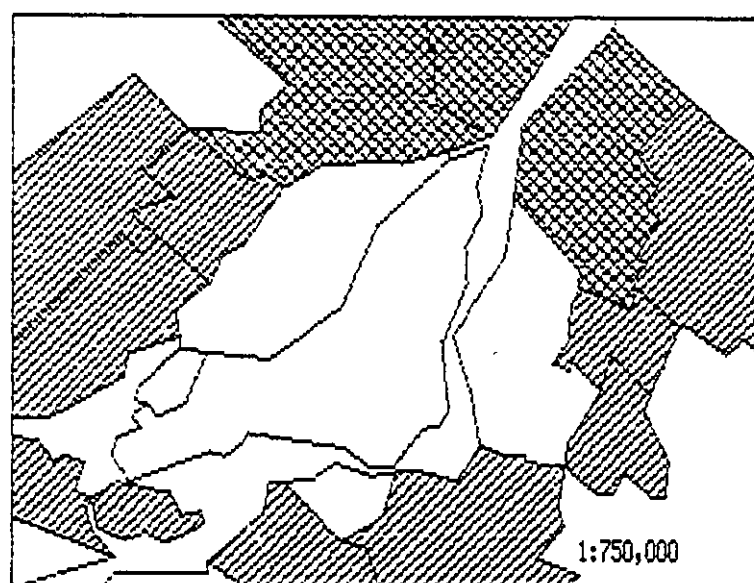
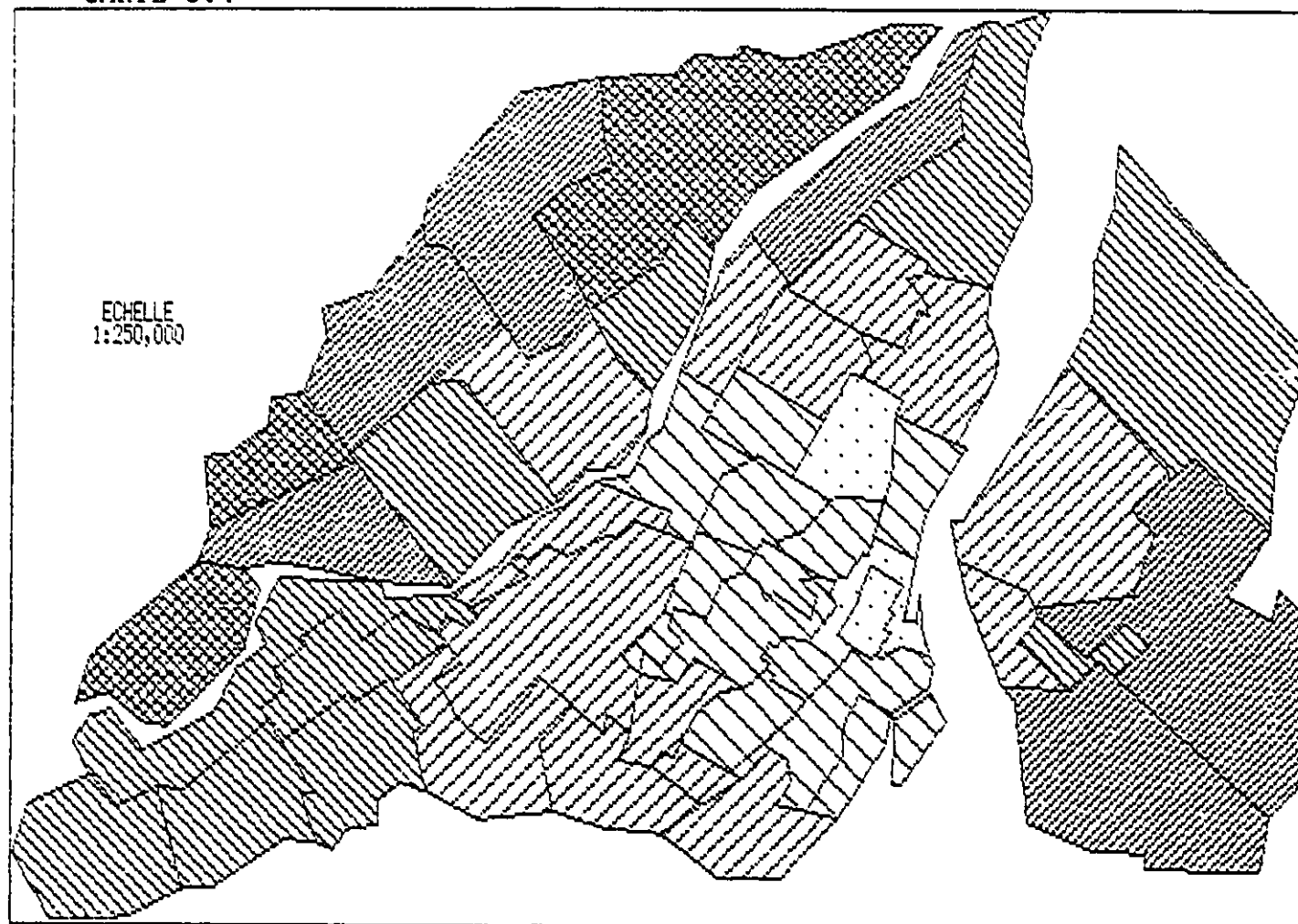


POPULATION 1981-2001
SCENARIO DE REFERENCE (C)
TAUX ANNUEL MOYEN

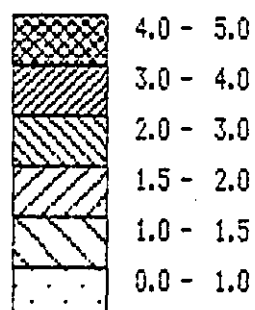


Carte faite par GRAFON

CARTE 8.4

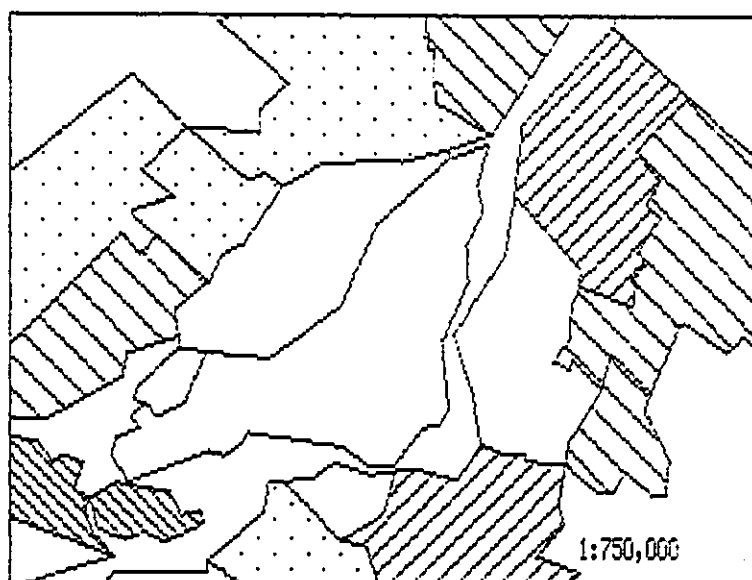
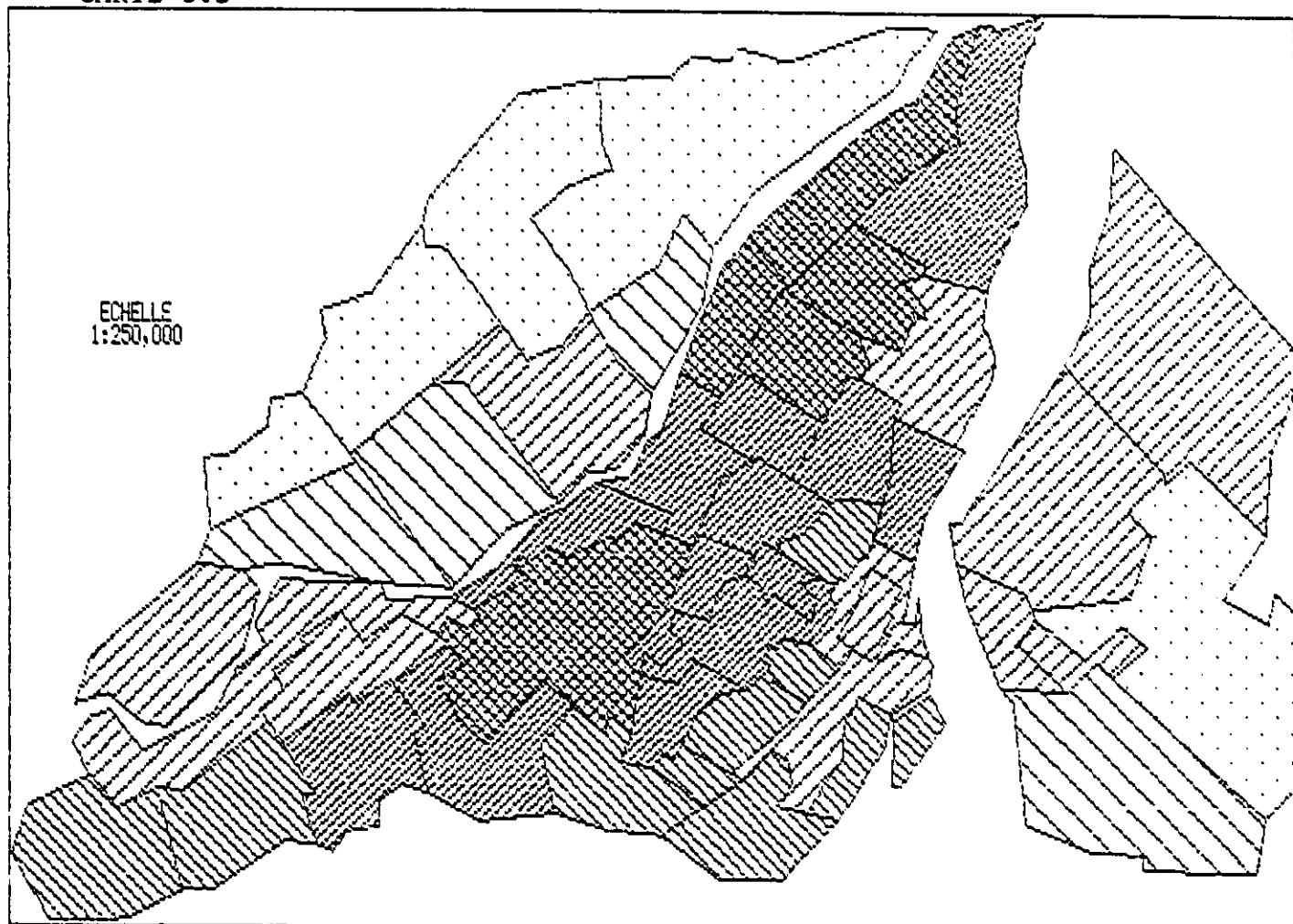


SERVICES 1981-2001
SCENARIO DE REFERENCE (C)
TAUX ANNUEL MOYEN



Carte faite par GRAFON

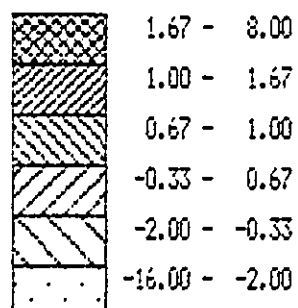
CARTE 8.5



POPULATION 2001

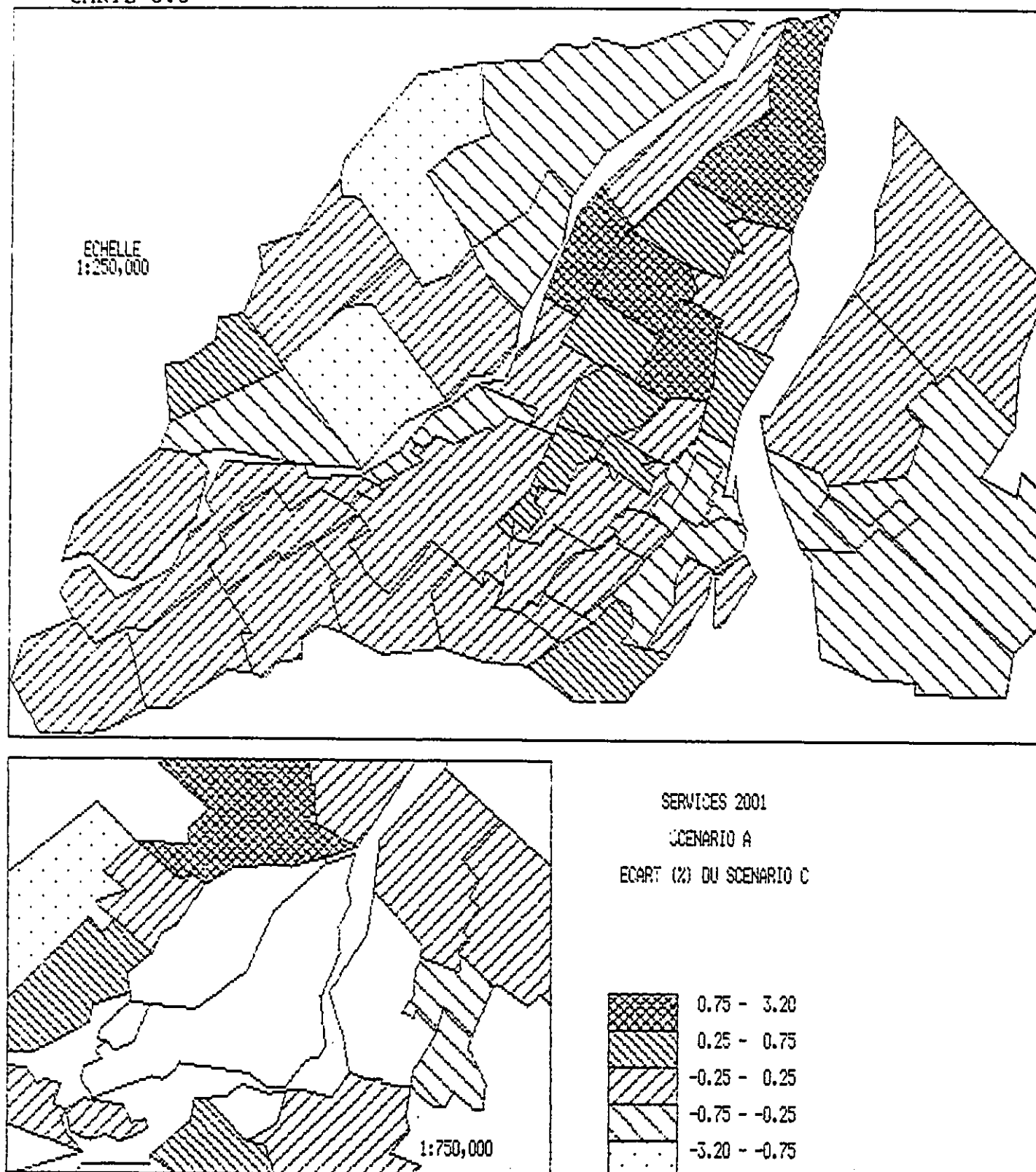
SCENARIO A

ECART (%) DU SCENARIO C



Carte faite par GRAFON

CARTE 8.6



Carte faite par GRAFON

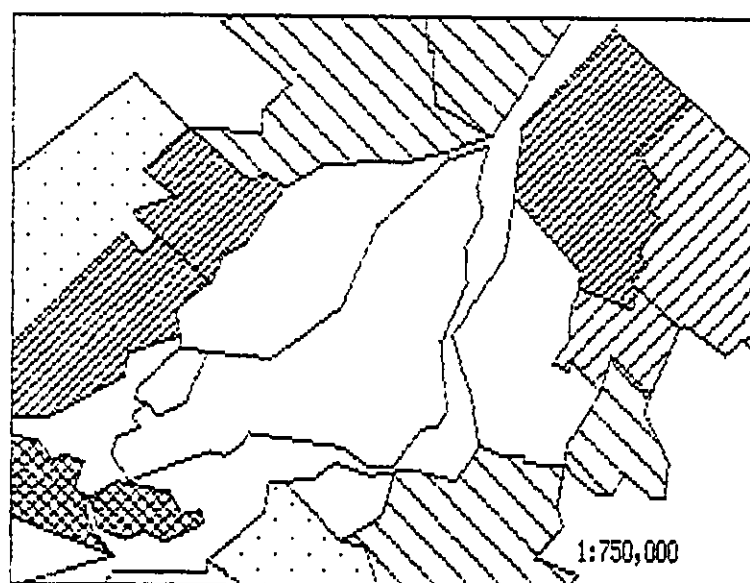
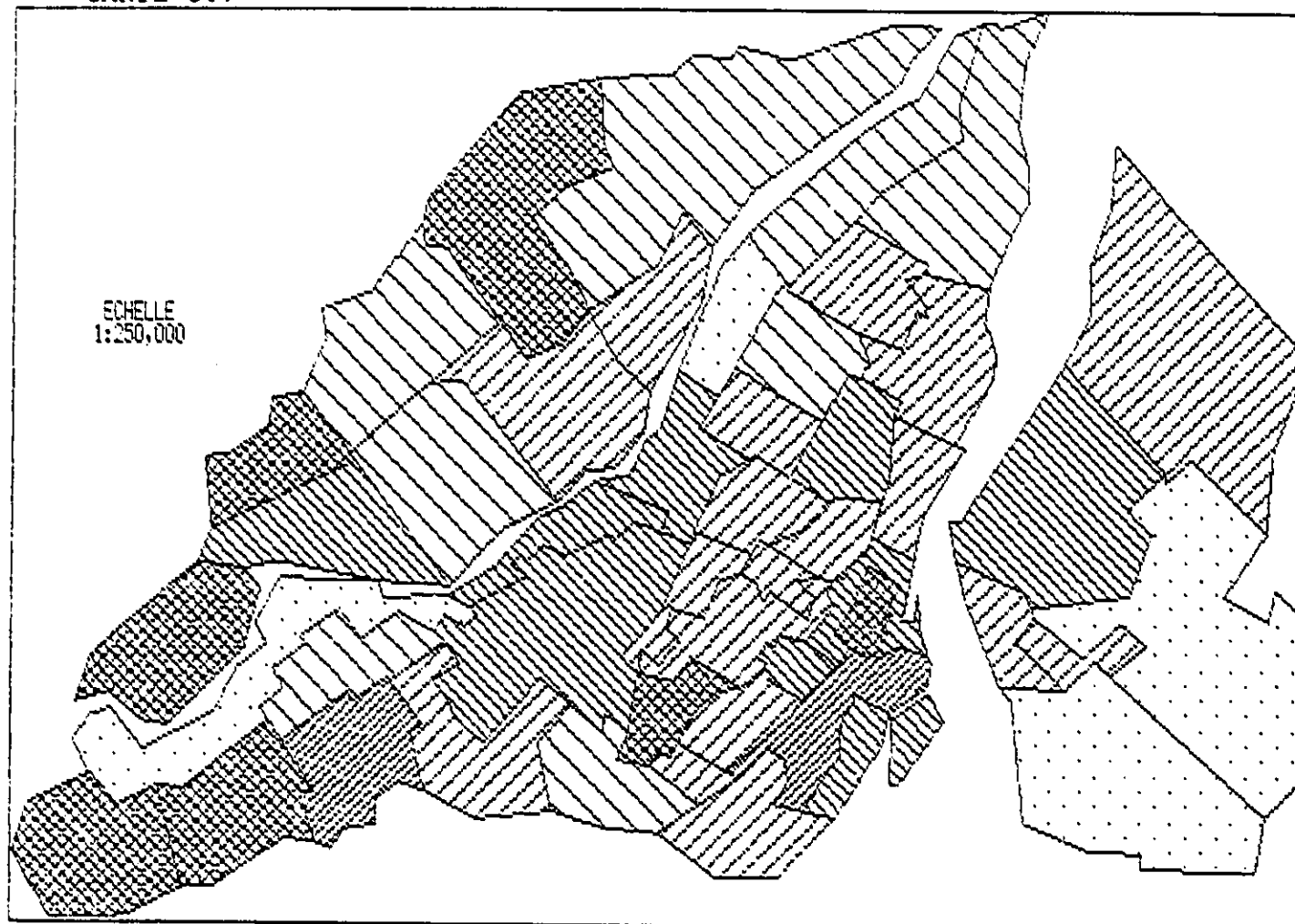
Quant aux emplois de service, l'image cartographique tend plutôt vers une mosaïque et se démarque de l'allure uniforme de la précédente. La plus forte friction de la distance dans le processus de localisation de l'emploi explique pourquoi les gains sont concentrés très près des infrastructures prolongées. Par surcroît, puisqu'il n'y a aucun projet dans l'ouest de l'île, les écarts entre scénarios A et C se révèlent imperceptibles. Deux anomalies (ou variations inexpliquées) sont observées dans la MKU des Moulins et le quartier Pointe-aux-Trembles; les forts écarts positifs ne peuvent se justifier sur la base des projets recommandés dans A.

Les écarts entre scénarios B et C sont cartographiés dans la paire de cartes 8.7-8.8. À l'égard de la population, il semble peu y avoir d'écart significatif dans le centre de l'île. L'exception serait sa partie meridionale qui bénéficierait dès 1991 de la mise en service du boulevard Ville-Marie (au lieu de 1996 dans C). Par rapport au scénario de référence, la partie est de la CUM y semble moins favorisée en ce qui concerne la complétion hâtive de l'autoroute 25; le statu quo du transport collectif doit faire la différence.

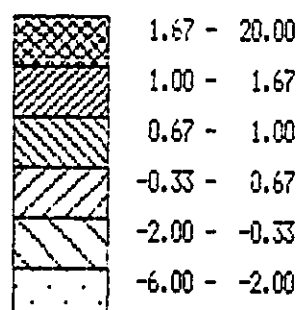
Le parachèvement de l'autoroute Laval (A-440) a un impact moins grand que prévu sur l'expansion résidentielle de la ville homonyme (il faut faire attention toutefois car il s'agit toujours d'écart à la carte 8.3 laquelle montre bien la forte croissance lavalloise). Les zones bénéficiaires se situent surtout en marge du tracé de l'autoroute (Vimont-Auteuil, Laval-Ouest) ou près de son extrémité occidentale (Ouest de la CUM à l'exception de Pierrefonds où l'accès à l'A-440 serait peu utilisée). Le cas de Vimont-Auteuil s'explique aussi par le prolongement de l'autoroute Papineau (A-19) réalisé en 1996.

Une image passablement différente nous est présentée dans la carte 8.8 montrant les écarts relatifs à l'emploi de service. Cette fois, le dessin paraît net. Une partie centrale plutôt déprimée et deux secteurs, Ouest et Nord-Est, affichant une croissance supérieure à celle connue selon l'hypothèse de référence. Le cas de l'Ouest de l'île tient à la combinaison de deux facteurs, la concentration initiale d'emploi et la venue de l'autoroute 440. Dans le Nord-Est, seuls les projets autoroutiers (A-20, A-25, A-440 Est) semblent faire la différence.

CARTE 8.7

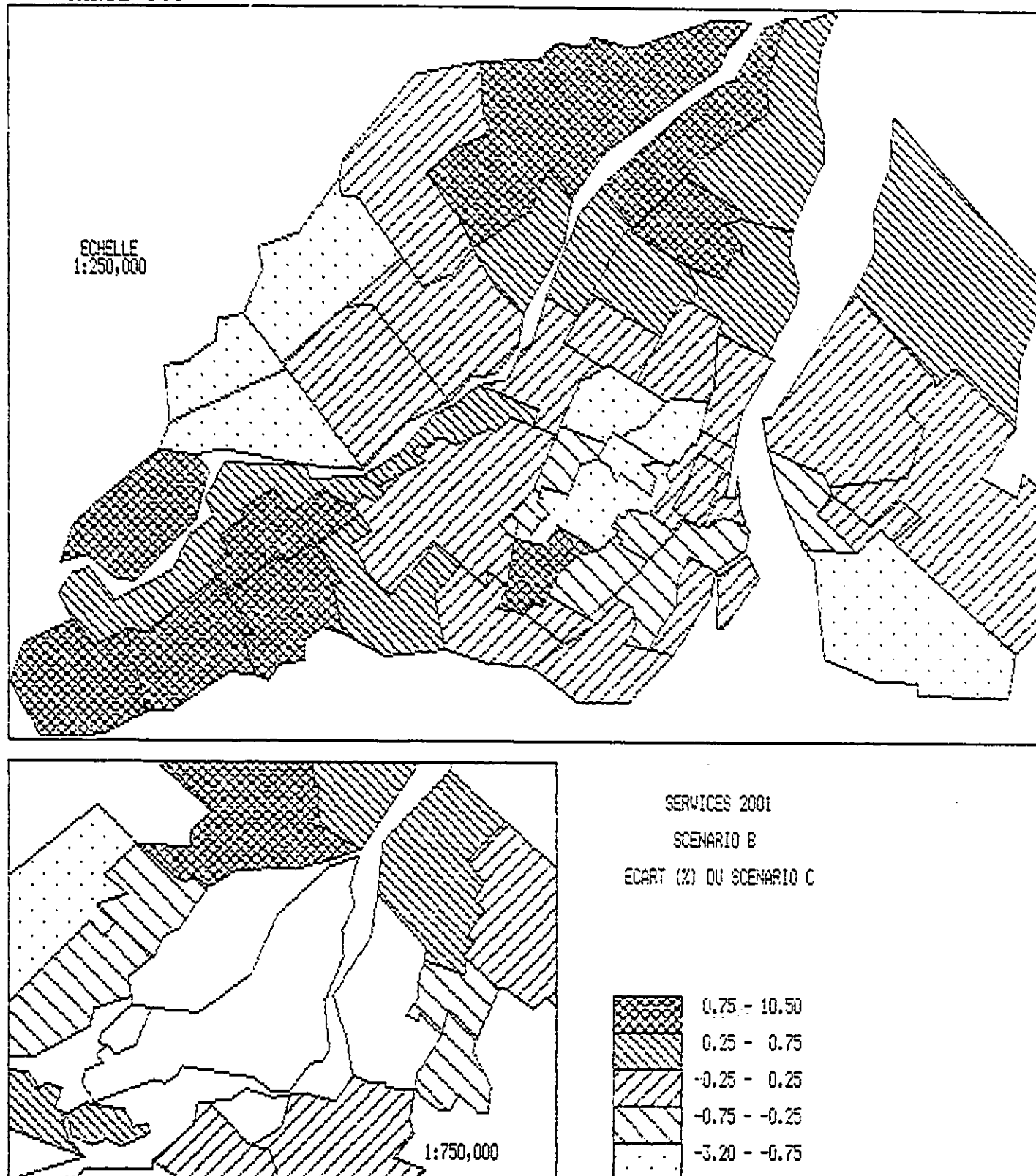


POPULATION 2001
SCENARIO B
ECART (%) DU SCENARIO C



Carte faite par GRAFON

CARTE 8.8



Carte faite par GRAFON

Les cartes du scénario de référence 8.3 et 8.4 ont montré que l'étalement persistait tant dans la population que dans l'emploi de service. Les cartes comparatives ont mis en évidence la spécificité du développement selon des projets de transport alternatifs. Dans un cas, ceux-ci freinent un peu l'étalement en favorisant les secteurs parcourus par les transports collectifs. Dans l'autre, les effets sont plus difficiles à cerner car les impacts se situent plus ou moins près des projets réalisés: la mobilité plus grande de l'automobile se manifeste sûrement par cette ubiquité du développement urbain. La section suivante abordera conséquemment des impacts sur les choix modaux.

8.2.3 Répartition modale

Dans ces simulations, on distingue deux types d'impact sur la répartition modale. Le premier type concerne les gains (ou pertes) qu'entraîne directement une amélioration (ou détérioration) d'une partie du réseau. Dans notre analyse, seules les variations relatives des coûts de transport modaux permettent un transfert de clientèle au sein d'un trajet O-D. Le second type réfère aux changements de la demande O-D occasionnés par le déplacement des activités urbaines. Le partage modal global est affecté du fait que chaque paire origine-destination affiche un niveau de concurrence modale différent.

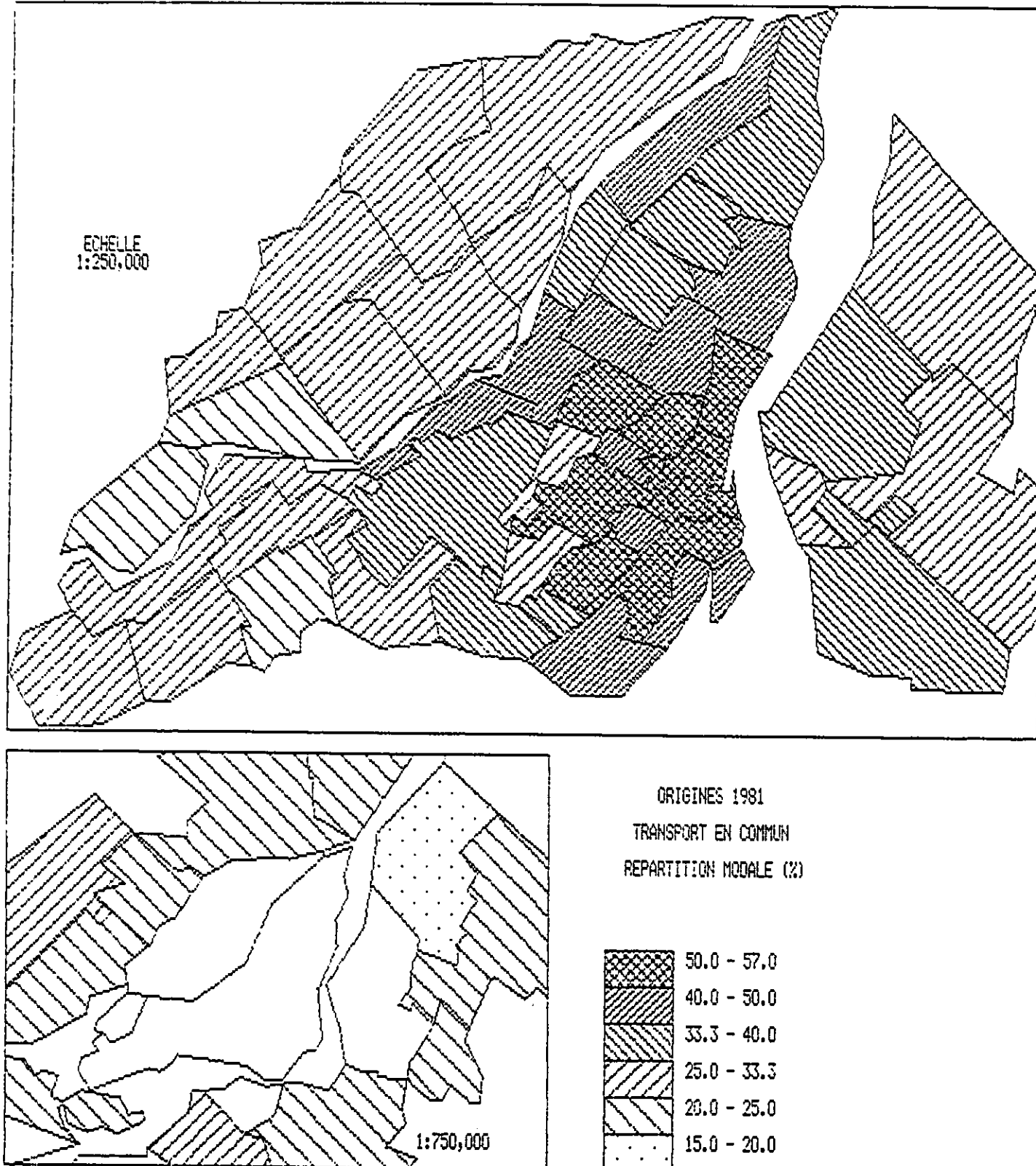
A cause des effets de congestion, les transferts de demande entre trajets O-D peuvent modifier les écarts de coût si bien que les deux types d'impact peuvent survenir en simultanéité. Cependant, ce jeu complexe échappe à la simple équation causale «amélioration du réseau = nouvelle clientèle». Seule une analyse cartographique permet d'apprécier les modulations du partage modal en fonction des scénarios de développement du transport.

Situation en 1981

Les cartes 8.9 et 8.10 montrent respectivement la proportion d'usager du transport collectif par origine et par destination en 1981. Les patrons géographiques se distinguent clairement. Dans le cas du partage modal par origine, les zones où la demande excède les 40% coïncident grossièrement au territoire municipal montréalais et à quelques municipalités adjacentes (Verdun, LaSalle, Westmount, Outremont). En fait, la structure spatiale paraît nettement concentrique, l'usage de l'automobile croissant avec la distance du centre-ville. Les banlieues huppées (Ville Mont-Royal, Côte-St-Luc/Hampstead, St-Lambert) représentent des exceptions notables mais attendues.

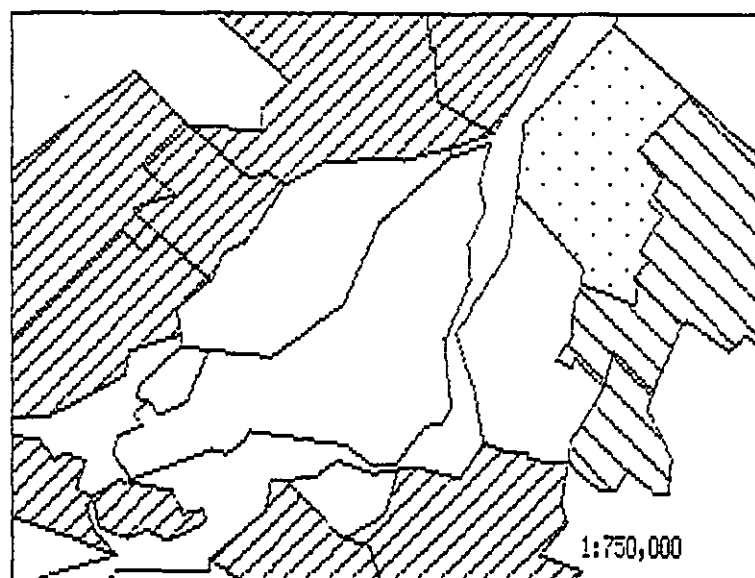
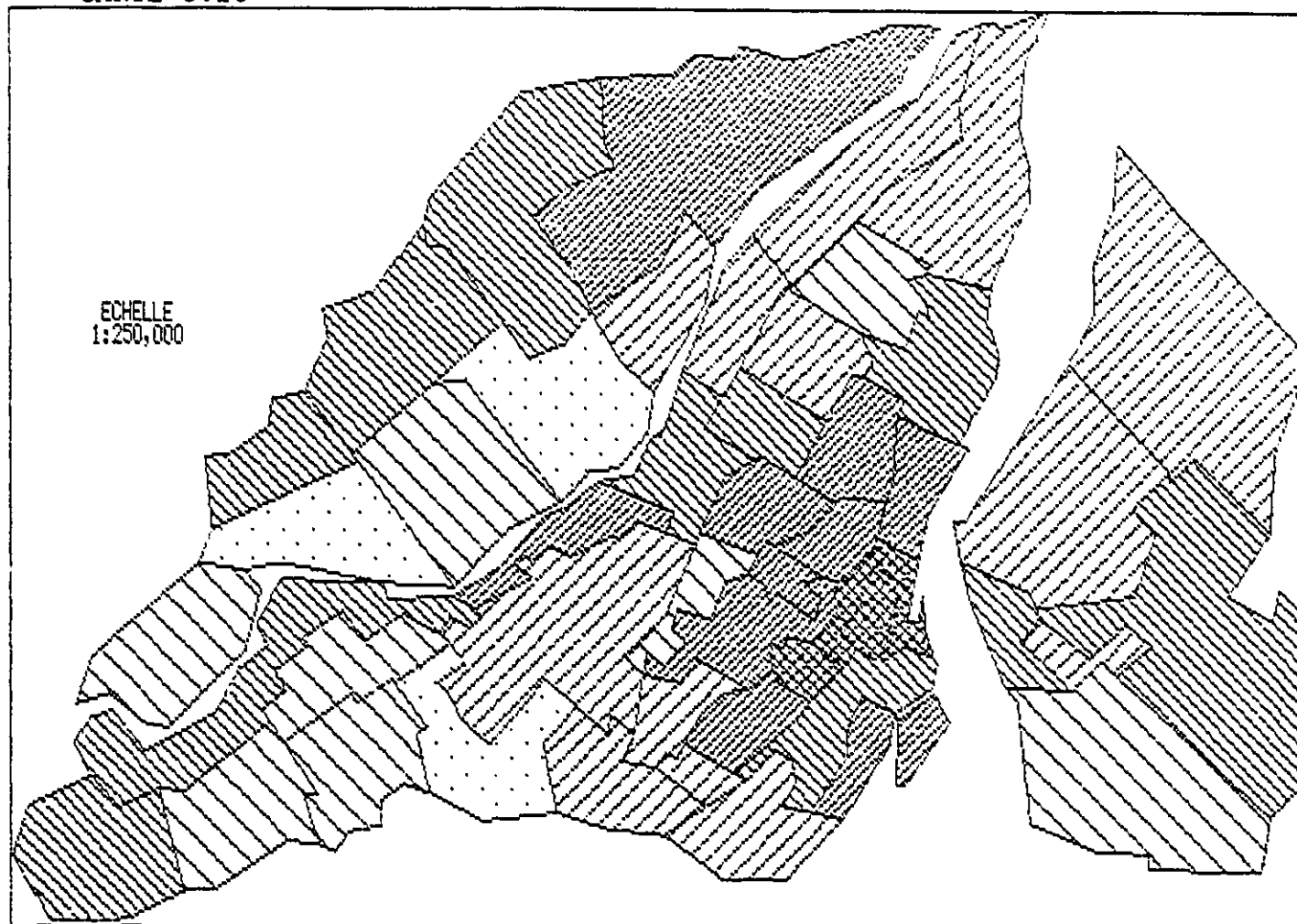
Les déterminants classiques de la répartition modale par origine tiennent à des caractéristiques de la population comme le niveau socio-économique et le taux de possession de véhicules. La distribution géographique de ces facteurs permet à elles seules d'expliquer une grande

CARTE 8.9

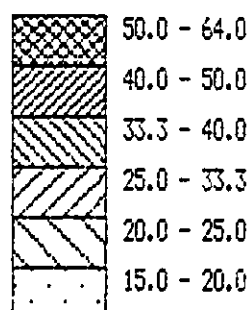


Carte faite par GRAFON

CARTE 8.10



DESTINATIONS 1981
TRANSPORT EN COMMUN
REPARTITION MODALE (%)



Carte faite par GRAFON

part du patron concentrique. Toutefois, la composition des lieux de destination contribue à moduler la proportion d'usager par origine. Par exemple, si les déplacements venant d'une zone sont majoritairement destinés vers le centre-ville, lieu réputé par sa grande accessibilité par transport collectif, le partage à l'origine sera favorable à ce mode. La proportion de voyageurs convergeant vers le centre-ville et sa périphérie décroît d'ailleurs avec la distance. Elle reproduit donc une forme concentrique.

La répartition modale par destination arbore également un modèle concentrique. Le déclin de l'usage apparaît néanmoins plus abrupte. Les seules zones où le pourcentage dépasse 50% sont le centre-ville et sa périphérie. Ces deux taux doivent être considérés avec égards puisque ces zones totalisent près de 450 milles destinataires motorisés, soit près de 20% du total métropolitain.

Le déclin de la clientèle transport collectif du centre vers la périphérie s'explique d'abord par celui du niveau de service et par la densité des aménagements urbains. Le niveau de service n'est pas nécessairement équivalent dans une même zone si l'on parle d'origine ou de destination. La structure convergente du réseau de transport public permet de desservir un vaste territoire lorsque les déplacements s'accommodent à cette structure. Par contre, le service convient habituellement moins bien aux trajets transversaux à l'intérieur de ce même territoire.

La densité d'occupation du sol intervient dans le choix modal parce qu'elle affecte, d'une part, la distance moyenne de déplacement et, d'autre part, le niveau d'encombrement du réseau routier. Il semble inutile de démontrer la convenance de l'automobile dans les milieux suburbains. Par surcroît, la disponibilité des sites de stationnement est étroitement liée à la densité de l'utilisation du sol. Les cas les plus probants concernant l'impact du stationnement sur le choix modal sont les zones de Dorval et de Laval-des-Rapides/Pont-Viau; dans l'une, c'est l'aéroport, aménagement horizontal par excellence, qui constitue le pôle d'attraction tandis que dans l'autre, les centres d'achat régionaux et le parc industriel lavallois mettent à la disposition des automobilistes beaucoup d'espaces pour se garer.

Les partages modaux surprenamment élevés qui sont observés dans la moitié septentrionale de Laval et dans la zone suburbaine de Pierrefonds/Roxboro relèvent indirectement des faibles nombres de destinataires. La quasi-absence d'activité commerciale et d'emploi fait ressortir le nombre d'écoliers, habituellement négligeable au sein des destinations, utilisant les autobus scolaires (comptes dans "autres autobus" dans l'enquête O-D de la STCUM).

Transferts modaux

L'analyse cartographique des transferts modaux repose sur l'écart du pourcentage de part de marché du transport en commun entre 1981 et 2001. La durée de vingt ans permet de mettre en relief les différences au sein des trois scénarios. Pour parvenir à une image synthétique de l'évolution géographique de la répartition modale, nous avons choisi de représenter les variations zonales à la fois pour les origines et les destinations. Il s'agit d'une moyenne pondérée des départs et des arrivées. Le pourcentage de répartition modale combinée par zone se formule ainsi:

$$MT = \frac{O_i + D_j}{O_i + D_j} \cdot 100$$

où MT: part (%) du mode m au sein des origines et des destinations de i
 O_i , D_j : déplacements originant de i et se destinant à j effectués par un mode m.

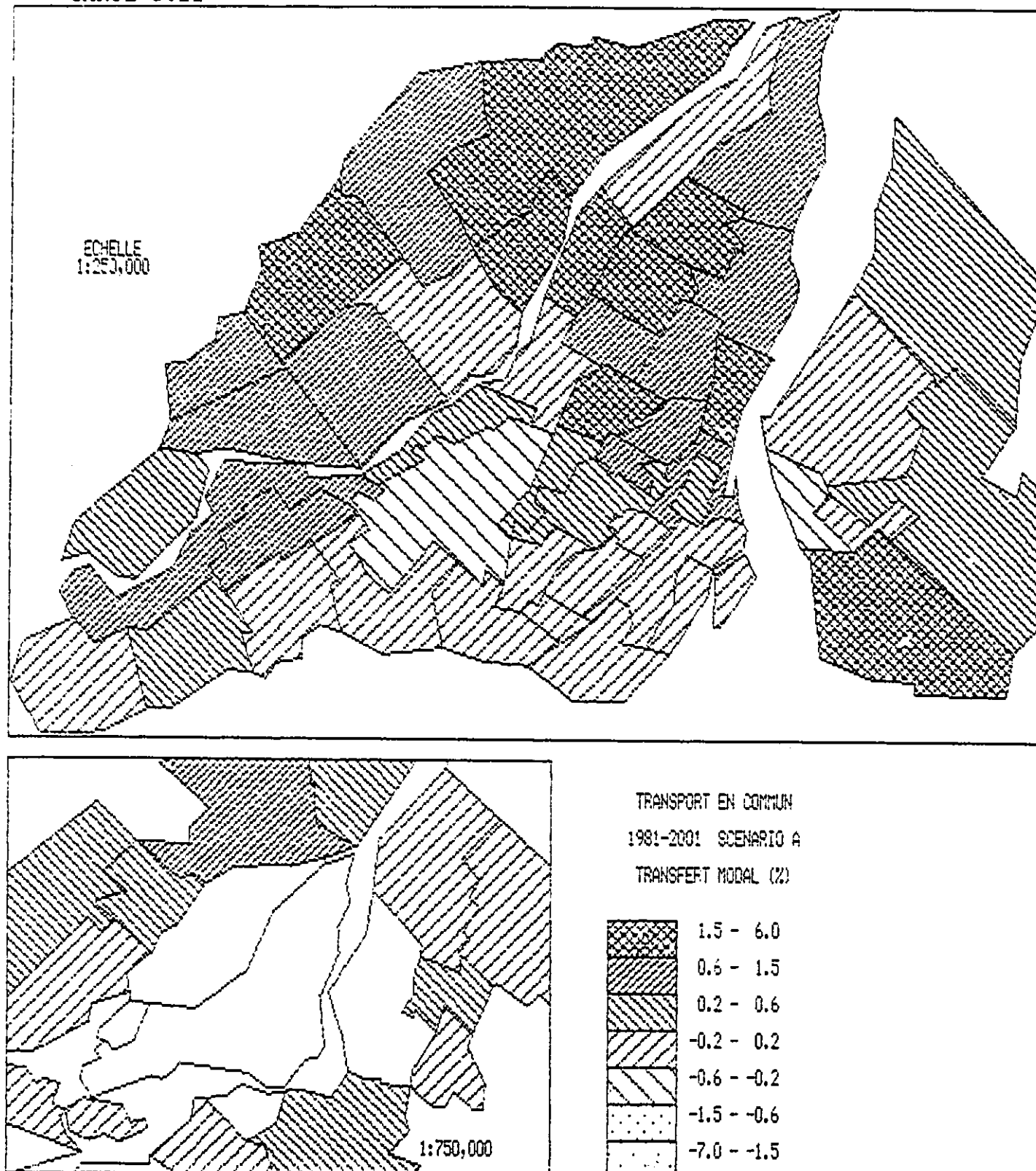
L'écart 1981-2001 est tout simplement calculé par soustraction:

$$ET(1981-2001) = MT(2001) - MT(1981)$$

Il faut se rappeler qu'il s'agit d'écart de point de pourcentage plutôt que d'écart traduit en pourcentage. Par conséquent, un écart de deux points s'avère très significatif puisqu'il indique un transfert de 2% de la part de marché.

La carte 8.11 illustre le transfert modal occasionné par la réalisation du scénario A. La grande majorité des zones montrent soit un gain de clientèle, soit une stabilité (classe centrale). Les zones où se produit un transfert positif se trouvent pour la plupart dans la moitié orientale de l'archipel. Les gains touchent non seulement les lieux affectés direc-

CARTE 8.11



Carte faite par GRAFON

tement par le prolongement de la ligne de metro no.5 et l'implantation de la ligne no.7 (notamment Montréal-Nord, St-Leonard, Anjou, Sud-est, Rosemont) mais aussi ceux qui beneficie d'un acces ameliore a ceux-la (Villeray, Plateau Mont-Royal, Outremont, Ville Mont-Royal et Côte-des-Neiges qui sont proches de la ligne 5 initiale: Centre-Ville et sa peripherie qui, en tant que destinations majeures, profitent des gains enregistres ailleurs). On observe egalement des transferts positifs dans l'est de l'île Jesus qui doivent s'expliquer par le rabattement presume des autobus lavallois sur la ligne no.7.

Des gains significatifs sont aperçus dans le nord-ouest de la CUM (Pierrefonds/Roxboro, Dollard-des-Ormeaux, Cartierville) et dans l'ouest de ville de Laval (Ste-Rose/Fabreville, Chomedey, Ste-Dorothee, Laval-Ouest). Il semble que le court prolongement de la branche occidentale de la ligne de metro no.2 soit responsable de l'augmentation de la part de marche. La zone de Cartierville et les quatre zones lavalloises beneficient d'un meilleur acces au reseau de metro (en 1981, l'interface entre le metro et les autobus de Laval se faisait uniquement a la station Henri-Bourassa de la ligne 2-est) tandis que la possibilite de correspondre au metro a la ligne de train de banlieue «Deux-Montagnes» permet d'attirer une clientele supplementaire dans le nord-ouest de l'île de Montreal.

Le cas de ville St-Laurent reste irregulier car il affiche une perte de clientele alors que le tronçon prolonge de la ligne 2-Ouest est situe a même son territoire. La seule explication envisageable tient au fait que la zone est une forte concentration d'emplois et de destinations en general. De par l'etalement urbain, une part de plus en plus grande des voyageurs provient des secteurs plus peripheriques mal desservis en transport en commun.

Sur la Rive Sud, l'implantation d'une voie reservee sur l'estacade du pont Champlain en 1996 semble accroître la part des usagers du transport collectif dans les zones de Brossard et de St-Hubert. Dans les couronnes peripheriques, on observe soit une augmentation, soit une stabilisation de la clientele. Les zones gagnantes semblent coïncider aux secteurs disoosant d'un nouveau point d'entree a l'île de Montréal comme les lignes 7 et 2-Ouest et l'estacade.

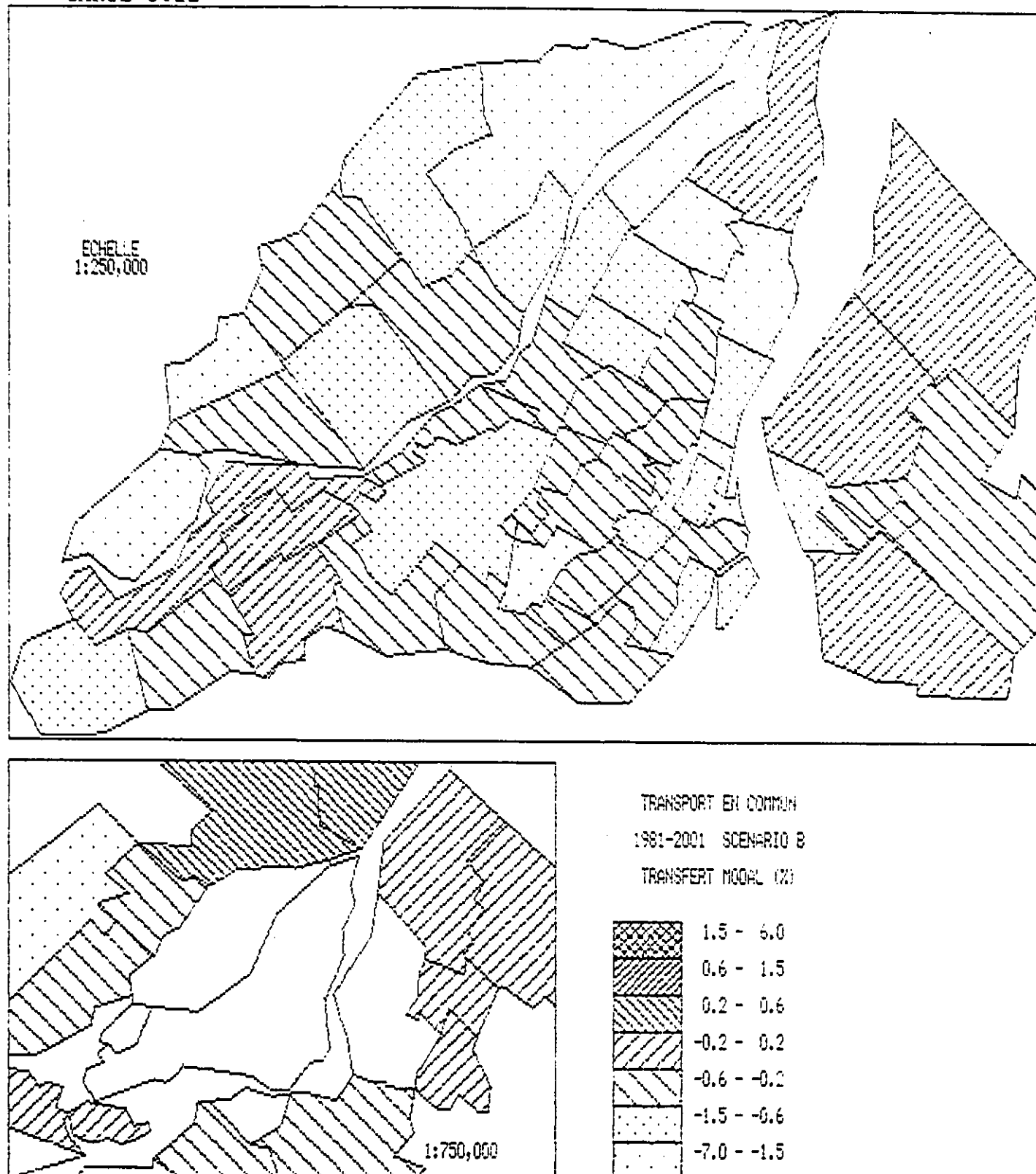
Les transferts modaux occasionnés par la réalisation du scénario R se manifestent négativement. En effet, selon la carte 8.12, il n'y a que très peu de zones où s'enregistre un gain de clientèle du transport en commun. La bonification du réseau autoroutier, le "statu quo" au sein du réseau concurrent et la tendance résultante à une forte suburbanisation favorise évidemment l'usage de l'automobile.

Les territoires directement touchés par les développements autoroutiers les plus précoces de notre période d'analyse, soit 1991, affichent les pertes d'usager les plus lourdes. Par exemple, les zones qui longent l'autoroute 25, raccordée en 1991, comme Mercier, Anjou, Rivière-des-Prairies et St-François de Laval accusent des déficits marqués. De plus, le parachèvement du boulevard Ville-Marie affectent les zones riveraines comme le Centre-Ville, Westmount, Sud-Est et Mercier. Enfin, l'autoroute Laval (440) quasi-périphérique semble bénéficier aux automobilistes des îles Jésus et Bizard.

Fait à noter, les transferts modaux sont généralement faibles dans les zones des couronnes périphériques. On remarque bien des chutes plus prononcées de clientèle dans les zones désormais desservies par l'autoroute Chomedey (Mirabel surtout et les MRC de Deux-Montagnes et de Ste-Thérèse) et dans le secteur sud-ouest nouvellement doté de la voie de ceinture A-30 (MRC de Roussillons Est et Ouest), mais les conversions restent modestes en regard de celles observées dans les territoires dits "non-périphériques". L'explication la plus évidente serait que les parts modales du transport public étaient en 1981 si tenues que les modifications de réseau n'ont que peu d'impact sur les points de pourcentage. D'ailleurs, la relation logistiquue que dessine la fonction de répartition modale (un modèle logit) est asymptotique à ses extrêmes; sa sensibilité aux différences de coût paraît forte lorsque les clientèles sont partagées mais faible si les parts sont fortement inégales.

Le cas particuliers des MRC des Moulins et de l'Assomption méritent de l'attention. Ces zones authentiquement périurbaines affichent paradoxalement des gains d'usager de transport en commun. L'une d'elle devrait même bénéficier des parachèvements des autoroutes 25 et 440. L'hypothèse la plus vraisemblable serait que le développement accélère de ces zones à

CARTE 8.12



Carte faite par GRAFON

accentue la congestion des axes autoroutiers qui les desservent. En effet, selon le scénario B, la croissance démographique 1981-2001 des deux MRC atteindrait de plus de soixante-deux mille habitants alors que l'emploi n'augmenterait moins de vingt mille. Dans la section suivante qui traite de la variation des coûts de transport, nous verrons si le facteur congestion peut jouer un tel rôle dissuasif, du moins théoriquement, sur l'usage de l'automobile.

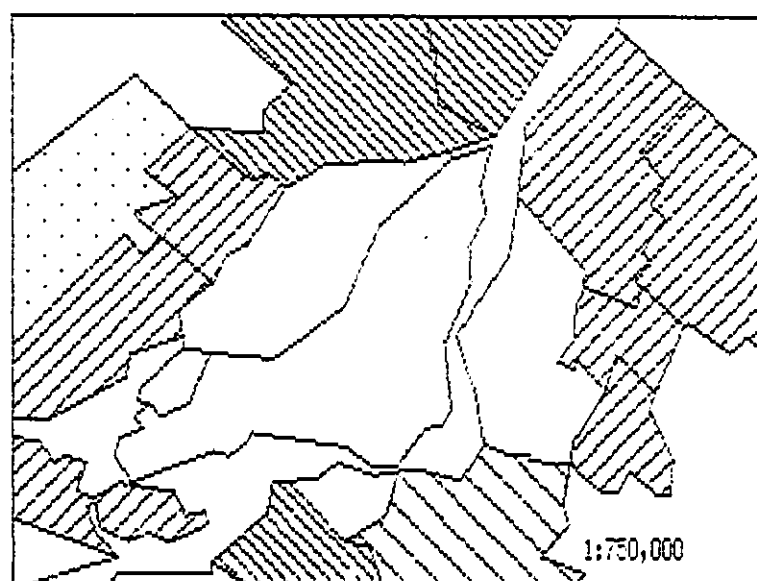
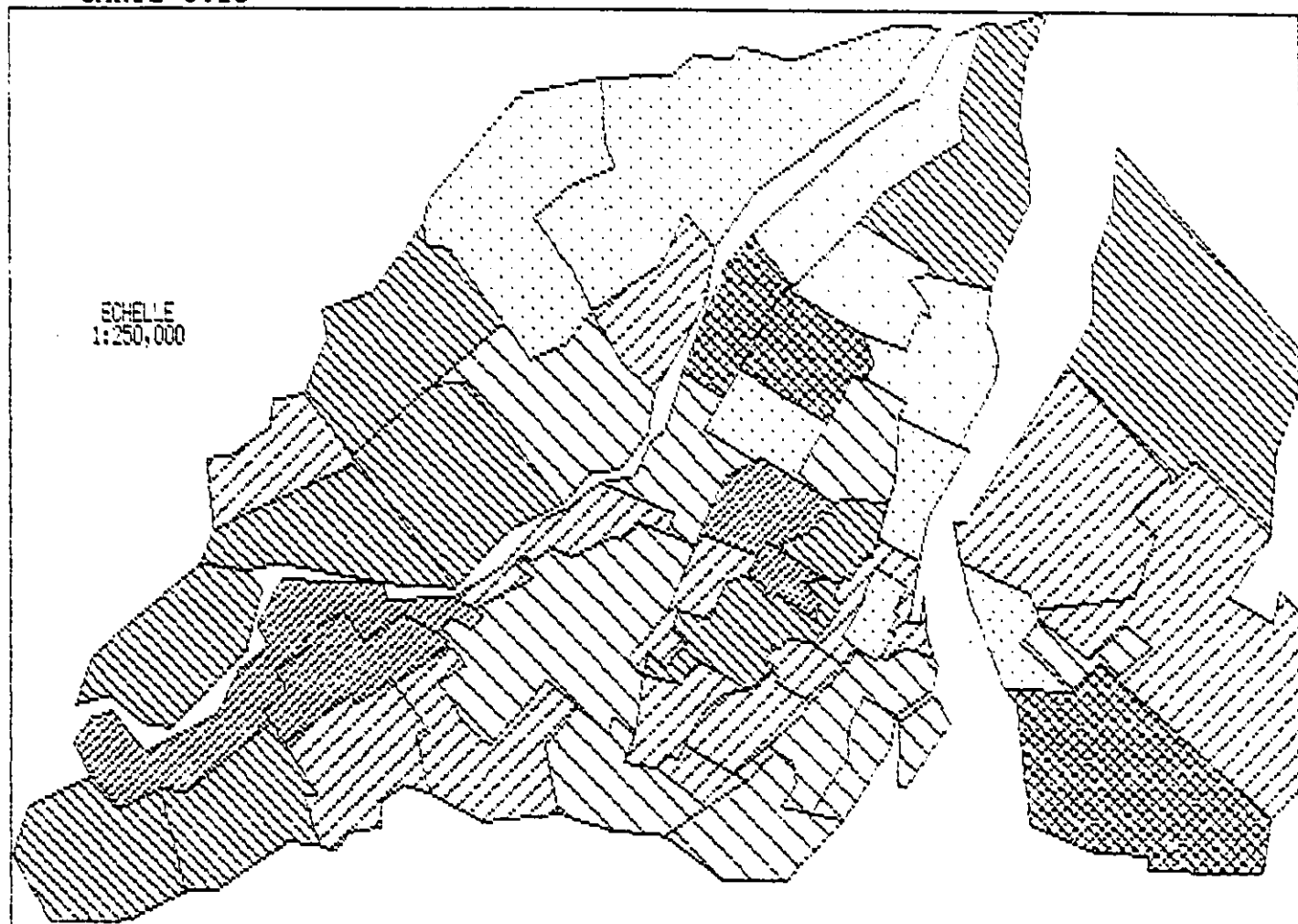
La carte de transfert modal relative au scénario C a une apparence beaucoup plus bigarrée. La mise en service simultanée de projets autoroutiers et de transport collectif implique des effets plus différenciés dans l'espace métropolitain.

Des gains de clientèle du transport en commun se font dans le nord-ouest de la CUM et dans la partie est de Laval à cause notamment du prolongement de la ligne 2-ouest. Comme conséquence du prolongement de la ligne 5 vers Montréal-Nord, des transferts positifs se produisent dans la municipalité du même nom et dans la St-Leonard. Les zones longeant le tronçon initial de cette ligne (Villeray, Outremont, Côte-des-Neiges) profitent de son extension vers le nord-est. Par ailleurs, la réalisation des voies d'autobus en site propre sur l'estacade du pont Champlain et d'un train de banlieue vers Châteauguay permet de gonfler la clientèle dans les zones directement desservies comme Brossard et la MRC de Roussillon (partie Ouest).

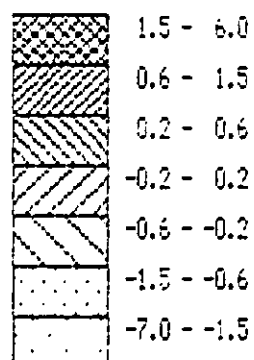
Les baisses relatives de clientèle sont surtout observées là où de nouvelles infrastructures routières ont été développées. Les exemples les plus frappants sont situés à même les corridors du boulevard Ville-Marie et de l'autoroute L.-H.-Lafontaine (A-25). À la différence du scénario précédent, les effets sur le transfert modal sont plus localisés car ils semblent endigués par l'amélioration synchronique du réseau de transport en commun.

Certaines zones semblent connaître un effritement de la part modale sans qu'aucune modification des réseaux n'intervienne directement. C'est le cas des zones du centre-ouest de l'île de Montréal comme Verdun, LaSalle, Lachine, Sud-Ouest et St-Laurent. Celles-ci restent d'importants

CARTE 8.13



TRANSPORT EN COMMUN
1981-2001 SCENARIO C
TRANSFERT MODAL (%)



Carte faite par GRAFON

lieux de destinations dans la région. Il se peut donc que l'étalement urbain, plus ou moins latent selon les scénarios, favorise l'usage de l'auto dans la zone d'origine et, par le fait même, affecte le partage modal à la destination. Ce constat devrait s'appliquer au centre-ville dont le transfert modal est toujours inférieur à la moyenne, sinon négatif, dans tous les scénarios.

8.2.4 Congestion routière et coût de transport

Le développement des réseaux de transport a avant tout pour but d'améliorer les conditions de déplacement ou, en d'autres mots, de réduire les coûts de transport. Les réductions se réalisent de trois manières, l'une en augmentant la capacité de transit du réseau, l'autre en offrant de nouveaux itinéraires et la dernière en autorisant des hausses de vitesse de déplacement.

Puisque le réseau de transport collectif ne présente pas, selon notre définition, de relation coût/trafic, ses usagers ne peuvent se prévaloir du premier type de réduction. Par contre, l'implantation d'une nouvelle ligne de métro peut à la fois procurer un parcours plus approprié et un gain de vitesse par rapport au service initial d'autobus.

Les automobilistes peuvent profiter des trois formes d'amélioration des conditions de transport. Par exemple, un élargissement de la chaussée augmente la capacité et, selon une demande fixe, diminue le coût relié à la congestion. La transformation d'une artère en autoroute permet d'accroître à la fois la capacité et la vitesse permise. Le prolongement ou le raccordement d'un tronçon autoroutier introduit de nouveaux itinéraires et hausse la capacité totale du réseau.

Dans les tableaux synthèses de la section 8.2.1, des statistiques globales de coût ont été dressées. Elles permettent d'apprécier l'impact général de l'expansion des réseaux. Mais elles occultent les effets locaux qui surviennent de façon plus prononcée. Les effets locaux que nous analyserons concernent le niveau de congestion sur les routes et l'accessibilité par zone.

Trois tableaux relatifs au niveau simulé de congestion sur les principaux éléments du réseau routier ont été dressés. Pour faciliter la lecture sans trop perdre d'information, la congestion est représentée par le rapport «débit/capacité» sur des tronçons bi-directionnels agrégés. L'agrégation a été faite selon un critère d'homogénéité de la capacité et du volume de circulation. Pour calcul le rapport débit/capacité moyen, il faut d'abord calculer le coût d'équilibre par lien selon la fonction de congestion en viqueur, soit:

$$s_i = s_0 \cdot (1 + \alpha(v_i/q_i)^\eta)$$

où s_i : coût d'équilibre du lien i ; s_0 : coût minimal de déplacement;
 v_i : volume sur le lien i ; q_i : capacité pratique du lien.

Dans les simulations les paramètres α et η prenaient respectivement les valeurs de 0.3 et 5. Un tronçon bi-directionnel agrégé est représenté par un sous-ensemble Λ auquel un certain nombre de liens adjacents, de même orientation et dans les deux directions. Le coût agrégé du tronçon (C_Λ) devient:

$$C_\Lambda = \sum_{i \in \Lambda} s_i(v_i)$$

Le rapport débit/capacité moyen (R_Λ) est alors estimé en inversant la fonction de congestion de manière à trouver le ratio équivalent au coût C_Λ en question:

$$R_\Lambda = \left[\frac{C_\Lambda / S_0 - 1}{\alpha} \right]^{(1/\eta)}$$

$$\text{où } S_0 = \sum_{i \in \Lambda} s_0$$

La série de tableaux 8.10 à 8.12 montre l'évolution de la congestion de 1981 à 2001 au sein du réseau routier. Selon le scénario A, qui n'implique aucun changement du réseau, le rapport débit/capacité augmente généralement de façon régulière d'une période à l'autre. Ceci s'explique par la croissance de la population et par une hausse de la mobilité causée par des facteurs de destinations (les activités non-résidentielles) davantage croissants.

Tableau B.10 Niveau simulé de congestion du réseau routier: scénario A

No.	Nom du tronçon	Extrémités	Rapport «débit/capacité» moyen				
			1981	1986	1991	1996	2001
A-40	Transcanadienne	(A-540 - Sources)	0.384	0.409	0.444	0.468	0.485
A-40	Transcanadienne	(Sources - A-150)	1.030	1.064	1.132	1.204	1.249
A-40	Métropolitaine	(A-150 - A-19)	1.179	1.301	1.229	1.305	1.258
A-40	Métropolitaine	(A-19 - A-25)	1.112	1.231	1.125	1.168	1.114
A-40	Métropolitaine	(A-25 - A-640)	0.598	0.690	0.716	0.757	0.759
A-20	Montréal/Ste-Anne	(A-540 - A-13)	0.262	0.276	0.289	0.299	0.310
A-20	Montréal/Ste-Anne	(A-13 - A-15)	0.995	1.022	1.027	1.052	1.047
A-20	Ville-Marie	(A-15 - St-Denis)	1.061	1.092	1.052	1.085	1.094
A-20	Ville-Marie	(St-Denis - A-25)	0.624	0.589	0.592	0.621	0.611
A-20	Tunnel L.H.Lafontaine	(A-25 - A-430)	0.662	0.680	0.738	0.742	0.775
A-20	Jean-Lesage	(A-430 - A-30)	0.432	0.447	0.479	0.488	0.514
A-440	Laval	(A-20 - A-13)	--	--	--	--	--
A-440	Laval	(A-13 - A-25)	0.782	0.746	0.799	0.895	0.769
A-640	Mille-Îles	(Deux-Mts - A-40)	0.379	0.404	0.446	0.481	0.509
A-520	Côte-de-Liesse	(A-20 - A-40)	0.698	0.715	0.718	0.720	0.698
A-13	Chomedey	(A-20 - A-40)	0.511	0.552	0.562	0.569	0.561
A-13	Chomedey	(A-40 - Samson)	0.768	0.804	0.864	0.923	0.955
A-13	Chomedey	(Samson - A-640)	0.483	0.505	0.546	0.587	0.624
A-10	Cantons de l'Est	(A-35 - A-15)	0.367	0.381	0.429	0.474	0.466
A-10	Pont Champlain	(R-134 - A-108)	1.224	1.284	1.383	1.420	1.498
A-150	Décarie	(A-108 - A-20)	0.963	0.955	1.044	1.092	1.171
A-150	Décarie	(A-20 - A-40)	0.896	0.886	0.946	1.017	1.110
R-117	Laurentien/Labelle	(A-40 - A-640)	0.420	0.428	0.492	0.598	0.587
A-15L	Des Laurentides	(A-40 - A-440)	0.842	0.914	0.932	0.972	0.918
A-15L	Des Laurentides	(A-440 - A-640)	0.589	0.632	0.667	0.674	0.685
A-19	Papineau	(A-40 - A-440)	0.841	0.996	1.100	0.890	0.941
A-19	Papineau	(A-440 - A-640)	--	--	--	--	--
R-125	Pie-IX	(A-20 - A-440)	0.933	1.001	1.024	0.934	1.131
A-25	L.H.Lafontaine	(A-20 - A-40)	0.763	0.781	0.853	0.878	0.866
A-25	L.H.Lafontaine	(A-40 - A-440)	--	--	--	--	--
A-25	Montréal/Terrebonne	(A-440 - A-640)	0.506	0.551	0.618	0.668	0.681
A-15M	Marie-Victorin	(A-30 - A-10)	0.181	0.195	0.216	0.227	0.237
A-430	Marie-Victorin	(A-10 - A-20)	0.402	0.411	0.445	0.460	0.483
A-430	Marie-Victorin	(A-20 - Varennes)	0.449	0.468	0.486	0.493	0.498
A-30	De l'Acier	(R-138 - A-10)	--	--	--	--	--
A-30	De l'Acier	(A-10 - A-20)	0.326	0.340	0.387	0.426	0.422
R-138	Pont Mercier	(A-30 - A-20)	1.107	1.142	1.228	1.284	1.320
R-112	Pont Victoria	(Well. - R-134)	0.505	0.496	0.488	0.473	0.467
R-116	Sir Wilfrid Laurier	(R-134 - A-30)	0.440	0.485	0.488	0.476	0.532
R-134	Taschereau	(A-15M - A-430)	0.408	0.404	0.448	0.469	0.469
R-116	Pont Jacques-Cartier	(R-134 - A-20)	1.221	1.307	1.290	1.299	1.262
A-108	Bonaventure	(A-150 - A-20)	0.662	0.788	0.749	0.726	0.734
--	Cavendish/Jean-Talon	(A-40 - A-150)	--	--	--	--	--
R-132	Kahnawake/Laprairie	(R-138 - A-30)	0.162	0.180	0.202	0.207	0.217

Tableau B.11 Niveau simulé de congestion du réseau routier: scénario B

No.	Nom du tronçon	Extrémités	Rapport «débit/capacité» moyen				
			1981	1986	1991	1996	2001
A-40	Transcanadienne	(A-540 - Sources)	0.384	0.409	0.426	0.446	0.463
A-40	Transcanadienne	(Sources - A-150)	1.030	1.064	1.129	1.190	1.167
A-40	Métropolitaine	(A-150 - A-19)	1.179	1.301	1.264	1.379	1.325
A-40	Métropolitaine	(A-19 - A-25)	1.112	1.231	1.079	1.120	1.111
A-40	Métropolitaine	(A-25 - A-640)	0.598	0.690	0.712	0.714	0.675
A-20	Montréal/Ste-Anne	(A-540 - A-13)	0.262	0.276	0.324	0.322	0.341
A-20	Montréal/Ste-Anne	(A-13 - A-15)	0.995	1.022	1.063	1.041	1.068
A-20	Ville-Marie	(A-15 - St-Denis)	1.061	1.092	1.138	1.152	1.158
A-20	Ville-Marie	(St-Denis - A-25)	0.624	0.589	0.870	0.944	1.017
A-20	Tunnel L.H.Lafontaine	(A-25 - A-430)	0.662	0.680	0.748	0.787	0.812
A-20	Jean-Lesage	(A-430 - A-30)	0.432	0.447	0.482	0.327	0.347
A-440	Laval	(A-20 - A-13)	--	--	0.240	0.256	0.278
A-440	Laval	(A-13 - A-25)	0.782	0.746	0.661	0.661	0.682
A-640	Mille-Îles	(Deux-Mts - A-40)	0.379	0.404	0.420	0.487	0.518
A-520	Côte-de-Liesse	(A-20 - A-40)	0.698	0.715	0.732	0.749	0.724
A-13	Chomedey	(A-20 - A-40)	0.511	0.552	0.563	0.573	0.583
A-13	Chomedey	(A-40 - Samson)	0.768	0.804	0.859	0.895	0.906
A-13	Chomedey	(Samson - A-640)	0.483	0.505	0.545	0.605	0.636
A-10	Cantons de l'Est	(A-35 - A-15)	0.367	0.381	0.451	0.381	0.400
A-10	Pont Champlain	(R-134 - A-108)	1.224	1.284	1.388	1.301	1.374
A-150	Décarie	(A-108 - A-20)	0.963	0.955	0.955	0.985	1.039
A-150	Décarie	(A-20 - A-40)	0.896	0.886	0.953	1.061	1.128
R-117	Laurentien/Labelle	(A-40 - A-640)	0.420	0.428	0.673	0.667	0.694
A-15L	Des Laurentides	(A-40 - A-440)	0.842	0.914	0.989	0.979	0.994
A-15L	Des Laurentides	(A-440 - A-640)	0.589	0.632	0.593	0.638	0.631
A-19	Papineau	(A-40 - A-440)	0.841	0.996	1.009	0.926	0.889
A-19	Papineau	(A-440 - A-640)	--	--	--	0.397	0.336
R-125	Pie-IX	(A-20 - A-440)	0.933	1.001	0.875	0.775	0.756
A-25	L.H.Lafontaine	(A-20 - A-40)	0.763	0.781	0.967	1.065	1.053
A-25	L.H.Lafontaine	(A-40 - A-440)	--	--	0.638	0.743	0.696
A-25	Montréal/Terrebonne	(A-440 - A-640)	0.506	0.551	0.824	0.872	0.868
A-15M	Marie-Victorin	(A-30 - A-10)	0.181	0.195	0.214	0.243	0.251
A-430	Marie-Victorin	(A-10 - A-20)	0.402	0.411	0.462	0.452	0.459
A-430	Marie-Victorin	(A-20 - Varennes)	0.449	0.468	0.491	0.500	0.505
A-30	De l'Acier	(R-138 - A-10)	--	--	--	0.120	0.125
A-30	De l'Acier	(A-10 - A-20)	0.326	0.340	0.404	0.380	0.401
R-138	Pont Mercier	(A-30 - A-20)	1.107	1.142	1.230	1.184	1.247
R-112	Pont Victoria	(Well. - R-134)	0.505	0.496	0.490	0.480	0.486
R-116	Sir Wilfrid Laurier	(R-134 - A-30)	0.440	0.485	0.448	0.367	0.388
R-134	Taschereau	(A-15M - A-430)	0.408	0.404	0.423	0.463	0.482
R-116	Pont Jacques-Cartier	(R-134 - A-20)	1.221	1.307	1.317	1.476	1.383
A-108	Bonaventure	(A-150 - A-20)	0.662	0.788	0.750	0.729	0.787
--	Cavendish/Jean-Talon	(A-40 - A-150)	--	--	--	0.476	0.507
R-132	Kahnawake/Laprairie	(R-138 - A-30)	0.162	0.180	0.202	0.227	0.232

Tableau 8.12 Niveau simulé de congestion du réseau routier: scénario C

No.	Nom du tronçon	Extrémités	Rapport «débit/capacité» moyen				
			1981	1986	1991	1996	2001
A-40	Transcanadienne	(A-540 - Sources)	0.384	0.409	0.450	0.468	0.473
A-40	Transcanadienne	(Sources - A-150)	1.030	1.064	1.134	1.195	1.179
A-40	Métropolitaine	(A-150 - A-19)	1.179	1.301	1.239	1.257	1.435
A-40	Métropolitaine	(A-19 - A-25)	1.112	1.231	1.113	1.020	1.094
A-40	Métropolitaine	(A-25 - A-640)	0.598	0.690	0.734	0.678	0.712
A-20	Montréal/Ste-Anne	(A-540 - A-13)	0.262	0.276	0.285	0.301	0.324
A-20	Montréal/Ste-Anne	(A-13 - A-15)	0.995	1.022	1.032	1.036	1.131
A-20	Ville-Marie	(A-15 - St-Denis)	1.061	1.092	1.058	1.150	1.142
A-20	Ville-Marie	(St-Denis - A-25)	0.624	0.589	0.523	0.954	0.982
A-20	Tunnel L.H.Lafontaine	(A-25 - A-430)	0.662	0.680	0.726	0.732	0.801
A-20	Jean-Lesage	(A-430 - A-30)	0.432	0.447	0.305	0.309	0.331
A-440	Laval	(A-20 - A-13)	--	--	--	--	--
A-440	Laval	(A-13 - A-25)	0.782	0.746	0.610	0.660	0.658
A-640	Mille-Iles	(Deux-Mts - A-40)	0.379	0.404	0.472	0.509	0.542
A-520	Côte-de-Liesse	(A-20 - A-40)	0.698	0.715	0.715	0.716	0.724
A-13	Chomedey	(A-20 - A-40)	0.511	0.552	0.575	0.563	0.634
A-13	Chomedey	(A-40 - Samson)	0.768	0.804	0.886	0.900	0.968
A-13	Chomedey	(Samson - A-640)	0.483	0.505	0.609	0.655	0.687
A-10	Des Cantons de l'Est	(A-35 - A-15)	0.367	0.381	0.371	0.404	0.404
A-10	Pont Champlain	(R-134 - A-108)	1.224	1.284	1.314	1.375	1.329
A-150	Décarie	(A-108 - A-20)	0.963	0.955	1.046	1.078	0.998
A-150	Décarie	(A-20 - A-40)	0.896	0.886	0.957	1.066	0.918
R-117	Laurentien/Labellé	(A-40 - A-640)	0.420	0.428	0.676	0.648	0.808
A-15L	Des Laurentides	(A-40 - A-440)	0.842	0.914	0.865	0.852	0.874
A-15L	Des Laurentides	(A-440 - A-640)	0.589	0.632	0.575	0.651	0.635
A-19	Papineau	(A-40 - A-440)	0.841	0.996	1.112	0.754	0.739
A-19	Papineau	(A-440 - A-640)	--	--	0.192 ¹	0.243 ²	0.274 ¹
R-125	Pie-IX	(A-20 - A-440)	0.933	1.001	1.106	0.981	0.794
A-25	L.H.Lafontaine	(A-20 - A-40)	0.763	0.781	0.882	0.901	1.027
A-25	L.H.Lafontaine	(A-40 - A-440)	--	--	0.414 ²	0.630	0.711
A-25	Montréal/Terrebonne	(A-440 - A-640)	0.506	0.551	0.688	0.838	0.881
A-15M	Marie-Victorin	(A-30 - A-10)	0.181	0.195	0.210	0.257	0.261
A-430	Marie-Victorin	(A-10 - A-20)	0.402	0.411	0.452	0.441	0.448
A-430	Marie-Victorin	(A-20 - Varennes)	0.449	0.468	0.485	0.489	0.498
A-30	De l'Acier	(R-138 - A-10)	--	--	--	0.123	0.127
A-30	De l'Acier	(A-10 - A-20)	0.326	0.340	0.361	0.394	0.400
R-138	Pont Mercier	(A-30 - A-20)	1.107	1.142	1.226	1.251	1.321
R-112	Pont Victoria	(Well. - R-134)	0.505	0.496	0.485	0.466	0.467
R-116	Sir Wilfrid Laurier	(R-134 - A-30)	0.440	0.485	0.364	0.392	0.396
R-134	Taschereau	(A-15M - A-430)	0.408	0.404	0.504	0.503	0.494
R-116	Pont Jacques-Cartier	(R-134 - A-20)	1.221	1.307	1.406	1.476	1.489
A-108	Bonaventure	(A-150 - A-20)	0.662	0.788	0.646	0.708	0.752
--	Cavendish/Jean-Talon	(A-40 - A-150)	--	--	--	--	--
R-132	Kahnawake/Laprairie	(R-138 - A-30)	0.162	0.180	0.250	0.229	0.233

1) autoroute mise en service jusqu'au boulevard Dagenais seulement

2) en 1991, autoroute construite jusqu'au boulevard Henri-Bourassa et complétée par la suite

Par contre, le faible déclin démographique et l'augmentation modérée de l'emploi dans le Centre de la CUM se traduisent par une quasi-stagnation des tronçons du quadrilatère central (autoroutes Décarie, Ville-Marie, Métropolitaine à l'ouest de l'A-25, L.-H.-Lafontaine au sud de l'A-40). L'oeuvre de l'étalement urbain, responsable d'une croissance soutenue des activités en banlieue, se manifeste surtout par une pression accrue sur les axes suburbains et certains ponts autoroutiers.

À propos des résultats du scénario B, la comparaison avec les résultats du scénario précédent semble ambiguë. En effet, la simulation B propose un accroissement de la capacité globale du réseau routier surtout en ajoutant de nouveaux liens. La reorganisation des flux automobiles qui s'ensuit fait qu'une bonne partie des tronçons, sans connaître une baisse de débit, affiche une hausse moins forte que dans l'application précédente. C'est le cas notamment de la route Transcanadienne (A-40) dans l'ouest de la CUM qui est soulagée par l'autoroute de contournement Laval (A-440) ou du boulevard Pie-IX (R-125) bénéficiant du raccordement de l'autoroute L.-H.-Lafontaine (A-25).

Mais il reste néanmoins que plusieurs tronçons subissent une augmentation plus forte dans le scénario B. Les tronçons qui ne profitent pas d'un ajout d'une route alternative, figure au sein de cette catégorie. Par exemple, le pont Jacques-Cartier (R-116) enregistre un encombrement croissant en général (mise à part de 1996 à 2001).

Certaines variations peuvent s'expliquer, mais incomplètement toutefois, par des facteurs plus généraux. La hausse générale de la clientèle automobile, en termes relatif et absolu, fait que plus de véhicules sont en circulation dans le réseau. Le niveau de congestion réussit à se maintenir du fait qu'une bonne part de cette hausse est absorbée par les nouveaux tronçons.

Par ailleurs, l'étalement géographique des résidences et des activités économiques introduit de nouveaux trajets O-D entre banlieues tout comme il favorise les déplacements à l'intérieur des zones suburbaines. Ce patron migratoire est engendré surtout par l'augmentation plus spectaculaire, car partant d'un niveau plus bas, des destinations dans les banlieues. Ce pro-

cessus d'«autonomisation suburbaine» a un impact particulier dans notre application puisque les déplacements internes et inter-banlieues sollicitent moins les grands axes autoroutiers traditionnels qui ont été conçus pour une ville centrale polarisant l'activité économique. De ce fait, en dépit d'un trafic à la hausse, la pression sur les axes majeurs est quelque peu atténuée. Dans l'analyse présente, le portrait est trompeur puisque les migrations intra-zonales ne sont pas affectées sur le réseau et que les liaisons inter-banlieues ne sont généralement pas autoroutières et deviennent, par le fait même, exclues du réseau routier étudié.

Les résultats du scénario C révèlent que le développement simultané du transport en commun et du réseau routier a un effet modérateur sur les niveaux de congestion. En fait, les rapports débit/capacité sont en général sensiblement moins élevés que dans la simulation précédente bien que le réseau routier a été moins étendu.

Parmi les raisons invocables, il faut noter que dans le scénario C, la part modale de l'auto et la poussée de l'étalement urbain y paraissent plus faibles en comparaison avec le scénario B. De plus, on y propose davantage de hausse de capacité des liens existants.

Il semble quand même difficile d'apprécier la contribution des projets de transport en commun. Le projet le plus important et le plus coûteux, l'extension de la ligne de métro no.5 vers Montréal-Nord, semble par exemple alléger la charge du boulevard Pie-IX. En regardant les périodes de mise en service, il devient par contre fort probable que le raccordement de l'autoroute L.-H.-Lafontaine concourt à la baisse de trafic du boulevard en question.

Néanmoins, il a toujours lieu de s'interroger sur l'opportunité de développer conjointement les réseaux routier et de transport en commun. Somme toute, le transport collectif perd globalement de la clientèle tandis que la hausse de la congestion n'est que ralentie (elle diminue que dans des axes précis).

Dans nos trois simulations, le mouvement de déconcentration urbaine persistait bien que le développement des systèmes de transport accélérât ou freinât le phénomène du moins localement. Le fait de tout miser sur le transport en commun (A) provoque avantageusement un gain de clientèle. Mais puisqu'on ne joule pas l'étalement résidentiel, les gains sont minces en comparaison de l'investissement et l'encombrement du réseau routier dont la capacité est gelée, continue de s'amplifier. La seule expansion du réseau routier (B) favorise davantage l'éclatement du périmètre urbanisé et, par conséquent, introduit beaucoup de véhicules dans le système. S'il faut conclure sur l'évolution de la congestion routière, nous devrions dire que le contrôle efficace de la circulation passe avant tout par celui de l'organisation urbaine. Sans cela, les efforts consentis au transport sont vite supprimés, voire contredits, par le patron migratoire généré qui conditionne davantage le choix modal et des itinéraires.

L'analyse de la congestion routière occulte cependant l'impact du développement des réseaux sur les conditions générales de déplacement. D'abord, elle n'informe pas sur les coûts de transport proprement dits qui sont assumés et sur les zones affectées. Ensuite, elle ignore les conditions de transport collectif.

La dernière lacune peut être comblée en s'intéressant à la matrice de coût composé, soit $\{X_{ij}\}$ (60-60). Celle-ci synthétise par définition les deux coûts modaux selon le processus de répartition modale en vigueur. Le problème inhérent à une telle variable est sa désagrégation élevée (3600 trajets) et sa difficulté d'agrégation spatiale. Si on parvient, par exemple, à faire une moyenne arithmétique des coûts par zone, comme $\sum_j (X_{1j} + X_{j1}) \div (2 \cdot 60)$, le résultat devient peu révélateur puisqu'une grande part des trajets considérés sont inutilisés.

Une moyenne pondérée par les déplacements interzonaux répondrait un peu mieux à nos attentes; elle résumerait par zone les coûts assumés par les voyageurs originant de et/ou de destination vers cette zone. Toutefois, la distribution spatiale des voyageurs étant issue d'un modèle doté d'une fonction déclinante du coût de transport, il tend à y avoir peu de variabilité entre les zones. En effet, lorsque les coûts de déplacement dans une zone sont généralement élevés, les trajets U-D usités se restreignent

aux environs. Inversement, les zones où il est facile de se déplacer montre un patron migratoire plus étendu. De ce fait, la moyenne pondérée varie peu d'une zone à l'autre.

Le coût composé de transport entre chacune des zones et le centre-ville pourrait être un indicateur significatif. Cependant, même en admettant le statut particulier du centre-ville, on ne peut considérer qu'il soit le foyer exclusif des déplacements. De plus, l'expansion des réseaux ne répond pas uniquement à une demande dirigée vers le centre des affaires montréalais.

En tenant compte de toutes les considérations précédentes, l'analyse de l'évolution des coûts devrait s'effectuer à travers d'un indice d'accessibilité. Un tel indice permet de résoudre le problème d'agregation zonale tout en évitant de recourir à une pondération par les déplacements. Il permet par surcroît de mesurer l'accessibilité par rapport à toutes les zones (et non seulement du centre-ville) selon leur importance relative.

Dans le contexte présent, cette "importance relative" pourrait se mesurer en terme d'emploi total par zone. A l'exception de l'activité résidentielle, toutes les activités urbaines ont été modélisées avec des variables d'emploi; celles-ci deviennent par conséquent de bons indicateurs de l'attraction respective des zones.

Un indice d'accessibilité est caractérisé par une fonction inverse de la distance ou du coût de déplacement. Dans un but de consistance, la fonction choisie est identique à celle qui a été estimée pour simuler le processus de localisation résidentielle. La fonction reproduit le déclin général des interactions "domiciles-lieux de travail" de la région et, partant de cette présomption, elle traduit une valeur perçue du coût de transport. L'indice d'accessibilité pour 1981 s'écrit comme suit:

$$AP^1 = \sum_j \exp(-\psi \chi_{1,j}) \cdot E_j^1 \cdot (100/E^0)$$

où AP^1 : indice 1981 de la zone i ; E_j^1 : emplois 1981 de la zone j ;

ψ : paramètre ($\psi = 1451$); $\chi_{1,j}$: coût composé bi-modal $i-j$ (1981);

E^0 : emplois totaux de la région.

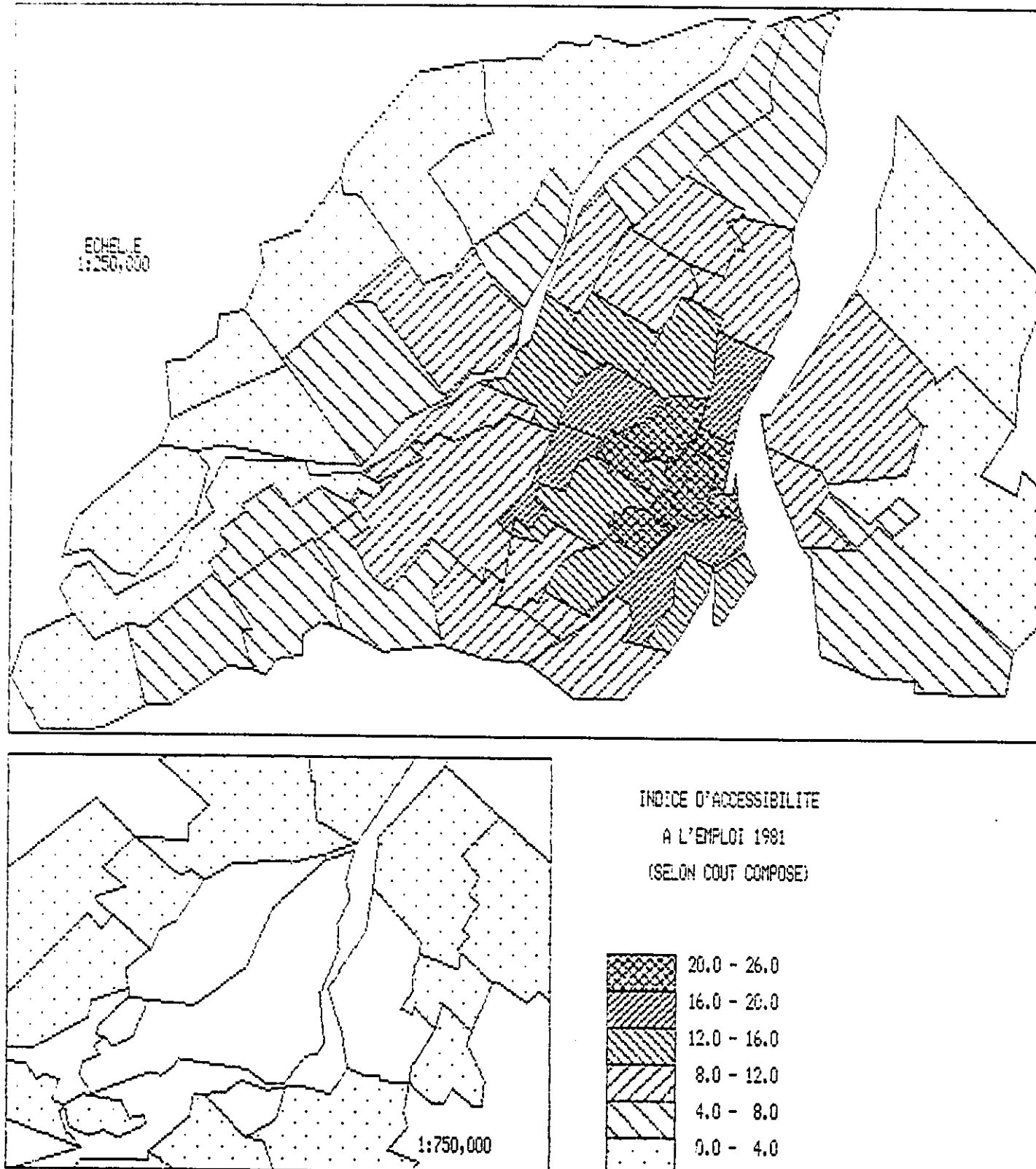
La multiplication par une constante " $100/\sum_{i=1}^n$ " reste facultative. Elle permet de standardiser les résultats de manière à qu'une zone obtienne un score parfait de 100 si tous les emplois se situent à un coût nul d'elle-même.

La prochaine carte (8.14) représente la distribution spatiale de l'indice d'accessibilité en 1981. Le patron général apparaît, comme il fallait s'y attendre, clairement concentrique. Il indique la forte concentration d'activités économiques dans le centre des affaires et les zones limitrophes. Néanmoins, les valeurs se dispersent géographiquement du fait que, d'une part, il existe plusieurs autres pôles d'activité dans la région et, d'autre part, la fonction de transport reste relativement lâche (le processus de localisation résidentielle étant dispersé).

Sur la carte suivante (8.15), est illustrée la variation de l'indice d'accessibilité sur vingt ans selon le scénario A. À cette fin, l'indice pour l'an 2001 a été calculé en prenant la matrice de coût composé (X_{ij}) de l'année correspondante mais en conservant le vecteur d'emplois de 1981. Il valait mieux pour fin de comparaison de reprendre la distribution initiale de l'emploi puisque le redéploiement spatial ensuivi, plus que les véritables variations de conditions de transport, aurait beaucoup trop affecté l'accessibilité. L'analyse présente se préoccupe uniquement des conditions de déplacement.

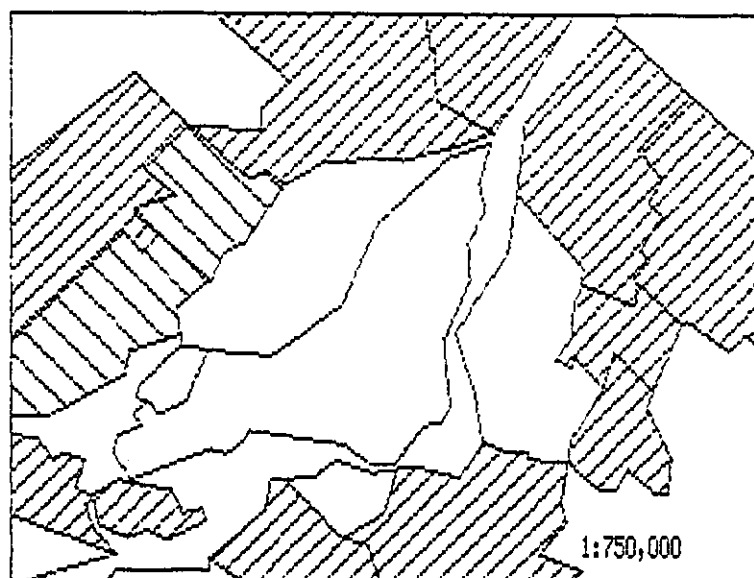
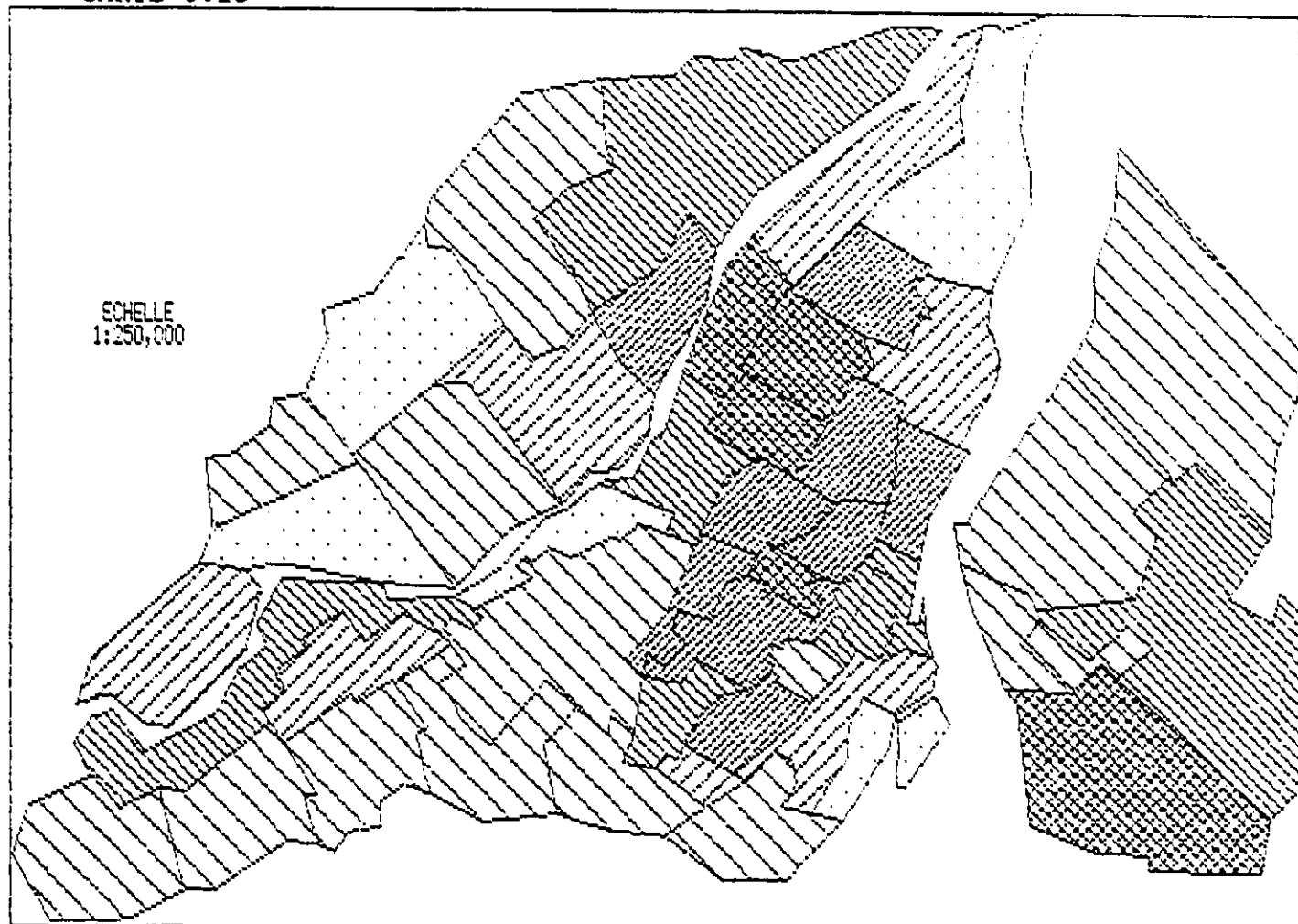
Sur cette carte, les plus forts gains d'accessibilité sont réalisés dans le nord-est de la CUM (Montréal-Nord, St-Michel, St-Leonard, Anjou), dans l'est de Laval (Duvernay/St-Vincent, St-François), dans la partie centrale de l'île de Montréal (Sud-Est, Rosemont, Villeray, Plateau Mont-Royal, Outremont, Ville Mont-Royal, Côte-des-Neiges, Notre-Dame-de-Grâce) et dans les villes de Brossard, St-Hubert et Pierrefonds/Roxboro. Les projets d'implantation des lignes de métro no.7 (mise en service jusqu'à Leger en 1991 et terminée en 1996) et no.5 (non-opérationnelle en 1981, prête pour la période suivante de Snowdon à St-Michel et complétée en 1991 et 1996) expliquent les variations observées dans le centre et dans l'est de l'île de Montréal ainsi que celle dans l'est de l'île Jésus. Les cas de Brossard et St-Hubert relèvent de l'implantation de la voie réservée sur l'estacade du Pont Champlain. La zone Pierrefonds/Roxboro profite plutôt de la correspondance du train de banlieue avec la ligne 2-Ouest prolongée

CARTE 8.14

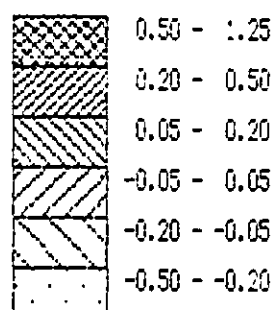


Carte faite par GRAFON

CARTE 8.15



VARIATION D'ACCESSIBILITE
1981-2001 SCENARIO A
(SELON L'EMPLOI 1991)



Carte faite par GRAFON

pour la période 1991.

Mais presque partout dans le reste de la CUM, de Laval et de la proche Rive Sud, on observe une baisse modérée de l'accessibilité. Étant donné que le scénario ne prévoit que des améliorations du transport collectif, on peut facilement attribuer cette dégradation des conditions de transport à la hausse régulière de la congestion routière. Autrement dit, l'augmentation de la circulation automobile, tributaire de l'organisation urbaine en expansion, tend à réduire l'accessibilité des banlieues à l'égard des foyers traditionnels d'activités économiques. Cela peut avoir pour effet de nuire à la croissance (résidentielle surtout) de certaines banlieues tout comme de favoriser l'étalement des emplacements de par la difficulté croissante d'accéder aux zones centrales.

La détérioration de l'accessibilité devient pire dans le cas d'un développement des transports orienté vers le réseau routier (voir carte 8.16). Sur l'île de Montréal notamment, les conditions de déplacement chutent dans la majorité des zones. Les hausses se concentrent dans le centre-sud-ouest de l'île qui bénéficie entre autres du raccordement Cavendish/Jean-Jacques (voie d'évitement du carrefour Decarie/Metropolitain) et du prolongement du boulevard Ville-Marie.

Dans la ville de Laval, les zones connaissent des augmentations substantielles d'accessibilité. Le parachèvement des autoroutes L.-H.-Lafontaine (A-25), Laval (A-440) et Papineau y est pour quelque chose. Cependant, la proche Rive Sud qui ne bénéficierait d'aucun nouveau lien routier avec l'île de Montréal, voit son accessibilité décroître de par la difficulté grandissante de traverser le fleuve. Ailleurs, dans les couronnes périphériques, la situation semble généralement stable. Cela s'explique par leurs indices d'accessibilité initiaux plutôt faibles et les marges du même ordre de grandeur.

Fait intéressant, ce sont surtout les banlieues de la CUM et de la proche Rive Sud qui écopent le plus dans la hausse de la congestion. En général, les quartiers montréalais (hormis Mercier et Pointe-aux-Trembles) affichent un bilan soit neutre, soit positif en raison de leur faible utilisation du réseau autoroutier. Au contraire, les voyageurs originant

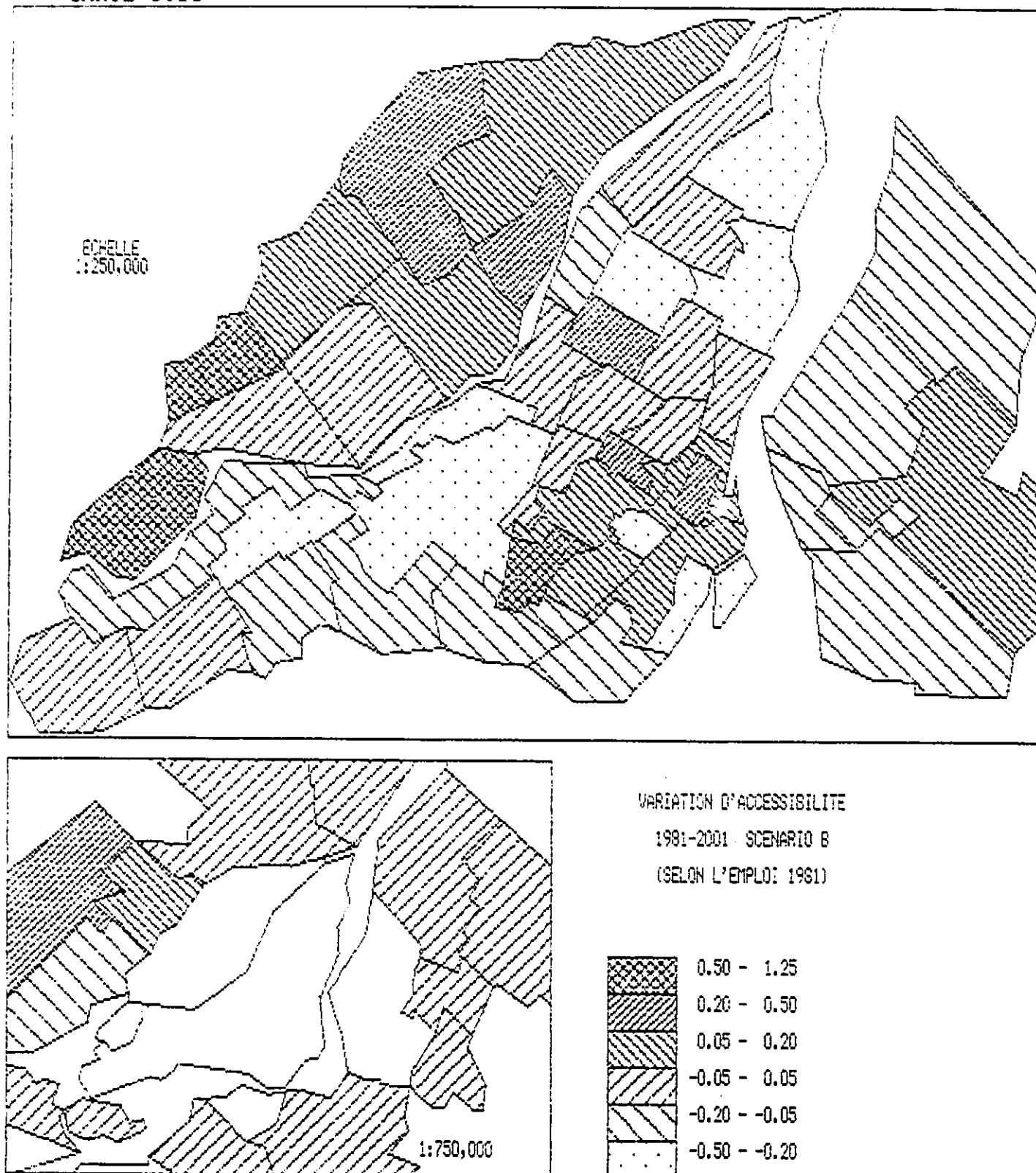
des banlieues de l'est et de l'ouest de DUM comorent beaucoup sur les autoroutes pour se déplacer.

L'image qui se dégage de la carte 8.16 est peu réjouissante du fait que la distribution des emplois est celle de 1961. Les mécanismes de relocalisation des activités économiques qui ont été modélisés ici, feront que les déséquilibres des conditions de transport s'atténueront. Face à la difficulté croissante s'accéder aux zones traditionnelles d'emplois et aux améliorations des autoroutes périphériques, une plus grande autonomisation des banlieues en résultera. Autrement dit, de plus en plus d'activités économiques qui se répandront en périphérie, s'adresseront aux banlieusards. L'histoire récente des régions métropolitaines américaines est éloquent à cet égard. On assiste à la formation de la "ville-beigne" caractérisée par un centre qui s'affaiblit et un couronne suburbaine dynamique et de plus en plus indépendante du centre.

Ce phénomène d'autonomisation a de grandes conséquences sur l'organisation urbaine des transports. Il devrait soulager la pression sur la partie centrale du réseau routier; mais cet effet peut être rapidement compensé par une hausse général de l'usage de l'automobile (le tableau 8.11 est révélateur sur ce point). Il rend par ricochet de plus en plus inappropriées les infrastructures lourdes de transport collectif qui convergent vers le centre-ville.

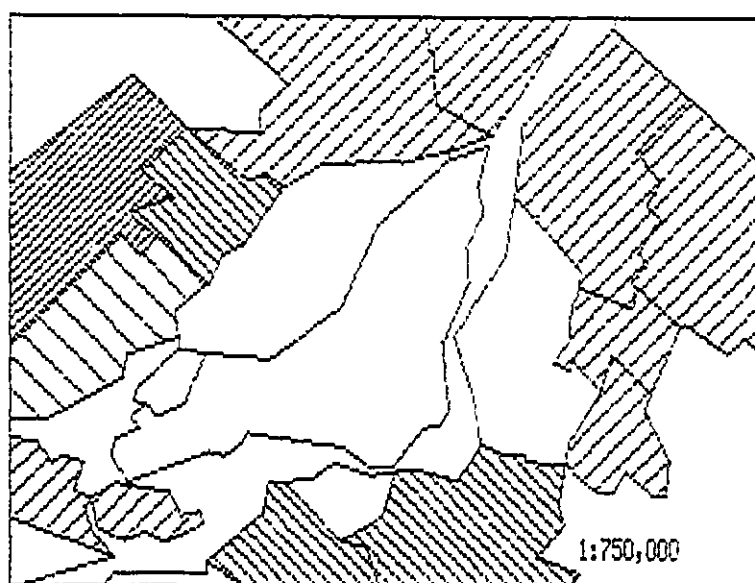
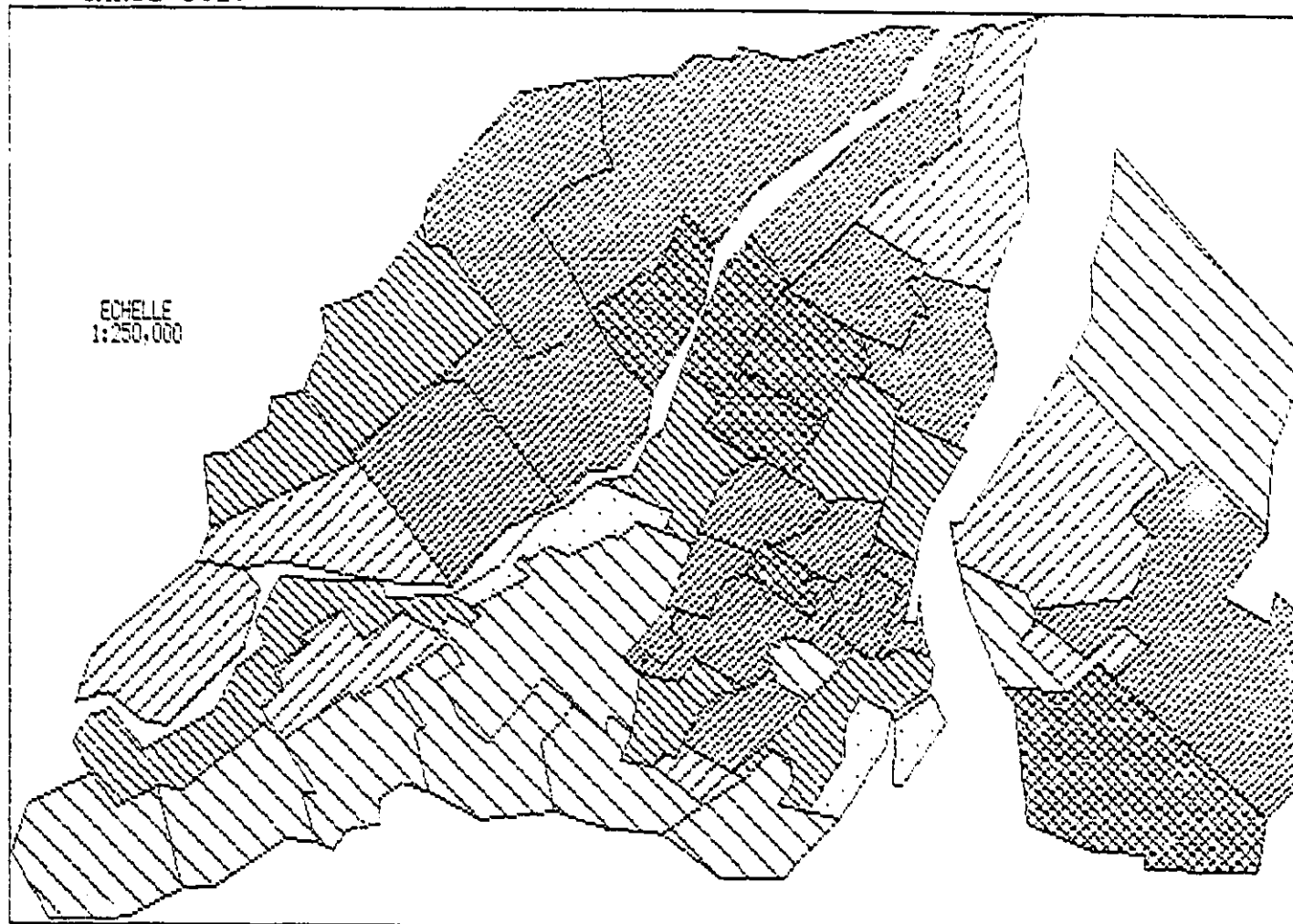
Quant aux variations d'accessibilité occasionnées par la réalisation du "plan Côte" (voir carte 8.17), elles se comparent davantage à celles observées sur la carte associée au scénario A. En effet, les zones du centre de la communauté urbaine restent favorisées bien que, cette fois, des gains substantiels d'accessibilité se produisent dans la majorité des zones de l'île Jésus. Il semble évident que les améliorations des infrastructures routières dans la ville de Laval et à ses abords permettent d'accéder plus facilement aux concentrations traditionnelles d'emplois.

CARTE 8.16

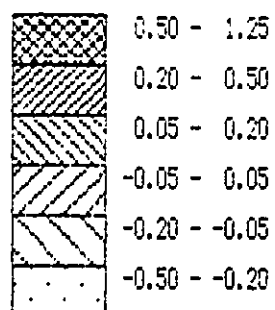


Carte faite par GRAFON

CARTE 8.17



VARIATION D'ACCESSIBILITE
1981-2001 SCENARIO C
(SELON L'EMPLOI 1981)



Carte faite par GRAFON

PAGINATION ERROR.

ERREUR DE PAGINATION.

TEXT COMPLETE.

LE TEXTE EST COMPLET.

NATIONAL LIBRARY OF CANADA.

BIBLIOTHEQUE NATIONALE DU CANADA.

CANADIAN THESES SERVICE.

SERVICE DES THESES CANADIENNES.

8.3 COMMENTAIRES SUR L'APPLICATION DU MODELE

Ce sous-chapitre nous invite à examiner les points forts et les lacunes d'une modélisation conciliant le développement urbain et les systèmes de transport. Parmi les éléments qui conditionnent la qualité de la modélisation et de son application, il faut noter les niveaux d'agregation des données utilisées. Dans le cas présent, ces préoccupations portent sur le découpage spatial, la grossièreté des réseaux de transport et de la périodicité des observations. Par ailleurs, il convient de s'intéresser aussi aux relations fonctionnelles des sous-modèles de transport et des sous-modèles de localisation. Ces deux aspects de la modélisation ont spécifiquement posé certains problèmes d'application.

8.3.1 L'agregation spatiale

Le niveau d'agregation spatiale correspond ni plus ni moins à la finesse du découpage zonal. Selon le point de vue, on parle d'agregation si on considère que l'unité spatiale représente tous les individus qui y vivent (ou y travaillent) ou de désagregation si on perçoit la zone comme une subdivision originale de la région qui l'englobe.

Par rapport au niveau individuel, on est tenté de croire que l'agregation en zone implique une forte perte d'information sur les comportements. Il y a perte d'information que si l'on compare des banques de données individuelles et zonales. Cependant, des observations au niveau de la personne (et tous les croisements d'information que cela suppose) sont rarement disponibles de par leur confidentialité. Les données de comportement économique ne sont qu'accessibles qu'avec une certaine forme d'agregation, spatiale ou non.

Au niveau de la modélisation, les implications de l'agregation spatiale relèvent peu de cette prétendue perte d'information. Il n'y a pratiquement pas de différence morphologique sinon conceptuelle entre des modèles de comportement individuel de transport [voir par exemple Domencich, McFadden 1975; Ben-Akiva, Lerman, 1985] et ceux de demande par zone. Les uns déterminent une probabilité de comportement (choix de destination, de mode, d'itinéraire) tandis que les autres concernent des fréquences attendues:

les comportements "probables" des voyageurs sont alors intéressés à une zone.

Les problèmes engendrés par l'agrégation spatiale qui a été préconisée ici concernent plutôt la représentativité des centroïdes et la grande proportion des déplacements interzonaux. En effet, le point central d'un territoire vaste correspond très peu à un lieu type de la zone. C'est le cas notamment des zones périurbaines de notre région à l'étude dont les superficies paraissent très étendues et dont les portions urbanisées ne sont pas forcément compactes. Une telle non-représentativité du centroïde occasionne certes des distorsions de l'affectation routière, les noeuds proches du point central bénéficiant d'une clientèle surestimée. Par surcroît, même si le calcul du coût de transport prend en compte les effets de l'aire de zone (voir section 7.2.3), il ne peut compenser la très forte variabilité de distance entre deux vastes zones adjacentes.

Dans la modélisation de la demande, l'usage de zones étendues mais peu denses fait qu'une trop forte proportion des déplacements restent à l'intérieur du périmètre zonal. Par exemple, en 1981, dans tous les MRC de la région métropolitaine, plus de moitié des voyages se destinaient dans la même zone. Cela ne tient pas au fait que ces zones ont plus d'opportunités que les autres (au contraire, les destinations sont déficitaires par rapport aux origines) mais plutôt au fait que les zones les plus proches restent relativement éloignées.

Toutefois, les zones périurbaines peuvent difficilement être davantage désagrégées, la base de données disponibles ne le permettant pas, et si cela s'avérait possible, leur faible densité n'assurerait pas un trafic significatif avec les autres entités du système.

Dans notre schéma de modélisation, la forte proportion de trafic intra-zonal apparaît fort nuisible car elle constitue en quelque sorte à une fuite. En effet, les déplacements internes, en étant automatiquement affectés à des liens auxiliaires, ne sollicitent aucun élément des réseaux existants comme projetés. Ils échappent donc au processus de développement de réseau accompli par le modèle de conception. Celui-ci faillit donc dans son appréhension du trafic suburbain en pleine croissance.

Jusqu'ici, seuls les problèmes d'agréation spatiale ont été examinés en regard de l'analyse du transport. Si on admet qu'une trop grande aggrégation masque beaucoup de caractéristiques de transport, il faut aussi reconnaître qu'un fin découpage spatial compromet grandement l'obtention de données sur les activités urbaines. Seule la variable "population" se prête facilement à des systèmes zonaux très morcelés car elle est recueillie à partir de secteurs très réduits. Par contre, les données sur les autres activités, exprimées ici en termes d'emplois, sont issues d'une base spatiale très grossière. La désaggrégation géographique à l'aide de données complémentaires (destinations ventilées par motif, personnel hospitalier par exemple) comprend certains de risques comme, entre autre, celui d'uniformiser les effectifs selon un patron connu et d'occulter les distributions spatiales spécifiques. De toute façon, la plupart des études quantitatives voulant concilier la demande de transport et le développement urbain se bute à cette déficience: soit elle abandonne tout projet de désaggrégation [Miller, Cubukgil, 1981], soit elle tente de réunir des informations partielles.

La désaggrégation géographique se voit contrainte aussi par des considérations de manipulation des données. Le type de modélisation proposé privilégie surtout les formats matriciels car ils permettent de représenter toute une gamme d'interactions entre lieux (matrices de coût de transport, de déplacement, de structure spatiale, de chemins admissibles et de leur incidence sur le réseau, etc...). Du fait, chaque zone supplémentaire fait accroître géométriquement la base de données. De plus, un système spatial finement découpé produit quantité d'informations insignifiantes; les matrices de déplacements sont "creuses", les matrices de coûts se remplissent d'éléments élevés dont les trajets équivalents n'enregistrent aucun trafic, les chemins admissibles se multiplient inutilement.

Le niveau d'agréation choisi doit également tenir compte de celui des réseaux. En effet, la grossièreté du réseau codé reste intimement tributaire de la densité de zones. Les zones petites mais nombreuses commandent un réseau plus raffiné sinon les parcours sollicités risquent d'être trop semblables; une telle inadéquation n'y gagnerait aucune information supplémentaire. Dans le cas présent, il y aurait été difficile de densifier davantage les réseaux.

8.3.2 L'agrégation des réseaux

Notre application repose sur une hypothèse à l'effet que seules les routes rapides (autoroutes et certaines artères) et les systèmes rapides de transport en commun (métro et trains) peuvent influencer le développement urbain. En fait, cette hypothèse se confirme par elle-même parce que les mesures de l'organisation urbaine sont prises à un niveau d'agrégation (système de 60 zones) équivalent à celui des réseaux.

La grossièreté des réseaux allait de pair avec l'imprécision des mesures de coût. En effet, bien que la distance kilométrique des segments ait été établi avec une bonne précision, les vitesses admises n'ont pu être attribuées que par catégorie de liens. Ainsi, les coûts minimaux de parcours sur les autoroutes se basaient sur les vitesses maximales légales tandis que sur artères sélectionnées, les vitesses ont été réduites aux deux tiers à cause des interruptions de circulation. Ces mesures peuvent être considérées comme passablement agrégées car elle ne peuvent traduire exactement les conditions pratiques d'écoulement du trafic. Malgré tout, la prise en compte des effets de congestion doit permettre une modulation plus réaliste des conditions de transport.

Sur le réseau de transport en commun, les vitesses commerciales ont été utilisées pour l'établissement des coûts. La négligence d'un effet de congestion s'explique par l'exploitation contrôlée du système (ajustement du service avec l'achalandage) et les voies en site propre dont jouissent les systèmes rapides. Par ailleurs, des temps de correspondance ont été additionnés alors que sur le réseau rival, il n'y avait aucune pénalité de virage sur les carrefours du fait qu'elles varient beaucoup trop selon le type d'intersection (échangeur, feux de circulation) et le côté du virage.

Le problème le plus apparent d'un tel niveau d'agrégation reste la considération des réseaux secondaires. Les trajets ne sollicitant pas le réseau primaire ainsi que les parties de trajet joignant la zone et le noeud d'accès au réseau principal ont été synthétisés par un parcours auxiliaire. Il s'agit d'une grille rectilinéaire imaginaire permettant de voyager à vitesse constante. Nous avons tenu compte des principaux obstacles soit en interdisant des trajets auxiliaires, soit en corrigeant

le coût de parcours ou soit en identifiant correctement les noeuds d'accès admissibles.

Les parcours auxiliaires sont effectués selon des conditions constantes de transport. Il n'y a aucune différenciation des vitesses ni même une considération de l'état de la circulation. Nous supposons alors que les voyageurs obtiennent généralement des résultats semblables, peu importe l'itinéraire opté au sein d'un réseau secondaire.

Du point de la modélisation, la mise à l'écart des réseaux auxiliaires facilite certes le traitement. Cela élimine un bon nombre d'itinéraires théoriquement possibles mais pratiquement inutilisés. Néanmoins, chacun des chemins auxiliaires, en agrégeant plusieurs parcours, attire une part non négligeable de la clientèle du trajet O-D correspondant. Dans notre application, cela constitue une "fuite" ou une perte d'information car dès qu'un flux est affecté à un chemin auxiliaire, nous perdons sa trace sur les réseaux à l'étude. Cette fuite a été sentie dans notre application à mesure que les activités urbaines se dispersaient dans la région métropolitaine. En effet, la croissance des trajets transversaux (en particulier inter-banlieues) lesquels sont peu desservis par des liens des réseaux primaires, fait qu'avec le temps le trafic disparaît peu à peu des réseaux de base.

Ce transfert (sans doute réel) rend plus difficile la prescription des éléments de réseau par le modèle optimisant leur conception. Bien que le modèle prend en compte dans sa fonction objective des coûts occasionnés sur les deux types de liens, il reste davantage sensible à l'affectation sur le réseau principal. D'ailleurs, les problèmes de surcharge de trafic ne peuvent être décelés sur un réseau qui ne possède aucune référence aux liens qui le composent.

Par ailleurs, il y a lieu de s'interroger sur la compatibilité des mesures de coût entre les trajets réguliers et auxiliaires. Dans les corridors urbains très achalandés, la négligence de la congestion ainsi que la présomption plutôt optimiste d'une géométrie régulière, donne un avantage certain aux parcours conceptuels. Ceux-ci profiteront vraisemblablement d'une affectation surestimée.

Un autre problème d'incompatibilité lié à l'agrégation a été observée mais, cette fois, parmi les parcours réguliers. Nous avons noté en effet une sur-utilisation du train comme moyen de transport. Bien que le train est le moyen collectif le plus rapide entre deux points, l'espacement des stations et les correspondances déficientes avec l'autobus et le métro en font un mode peu commode pour nombre de trajets O-D. Ces caractéristiques défavorables sont cependant masquées par l'agrégation des lignes d'autobus par une grille imaginaire qui sous-tend erronément que toutes les gares ferroviaires sont desservies à l'égal du métro. Dans les faits, les parties de trajet entre lieu de départ et gare et entre gare et lieu de destination ne s'effectuent pas aussi facilement.

8.3.3 Période d'étude et agrégation temporelle

Parmi les dimensions considérées, le temps a été sans aucun doute celle qui a imposé le plus de restrictions sur l'application. D'abord, étant donné l'impossibilité de constituer une base de données rétrospectives et désagrégées spatialement sur les activités urbaines, nous n'avons pu estimer les paramètres temporels du sous-modèle de localisation selon une procédure basée sur les séries temporelles. Ensuite, la périodicité irrégulière et l'inconstance de la base spatiale des données de demande de transport n'a pas permis d'effectuer une analyse évolutive ni même de lier sa dynamique antérieure avec celle des activités urbaines.

Cette pauvreté de l'information antérieure a interdit toute prétention à une analyse prévisionnelle. Pourtant, la méthodologie d'estimation exposée dans le chapitre 6 permet de procéder à des prévisions. Ces méthodes, il faut ici l'admettre, n'apparaissent pas appropriées pour la modélisation de variables d'aménagement spatialisées. D'une part, celles-ci sont rarement obtenues selon une base temporelle désagrégée alors que les analyses de séries temporelles ne s'alimentent que de chroniques finement découpées (mensuellement ou trimestriellement) et suffisamment longues. D'autre part, de telles analyses ne s'accroissent que de séries stationnaires i.e. dépourvues de tendances. Or les données d'activité n'exhibent aucune cyclicité et varient tendanciellement dans le temps; la différenciation des séries n'apporterait pas l'effet escompté puisque l'évolution d'une petite zone, si examinée à intervalle court, se fait par

saccades. La seule modélisation spatio-temporelle possible, selon une approche de Bennett [1979], convient donc mieux à des séries observées dans des grandes régions (où les variations accidentelles restent moins probables).

Il a fallu alors se résoudre à adopter une approche simulatoire basée sur une période de référence sur laquelle la plupart des paramètres ont été estimés. Dans le cas des modèles de demande de transport, il était déjà convenu d'employer des paramètres fixes calibrés à partir de l'année de base, 1981. Le modèle de localisation a pour sa part été estimé selon les distributions existantes en 1981; seuls les paramètres assignés à la variable "population", disponible en 1981 et 1986, ont pu être établis sur la différence des deux périodes. Les coefficients rattachés aux variables d'emplois ont par contre été mesurés comme si la distribution de 1981 parvenait à l'équilibre (elle n'y parvient pas car tous les régresseurs évoluent sans cesse). Cette simplification de la méthode prévue à l'origine semble quand même adéquate, les résultats de la simulation témoignant de la plausibilité des valeurs paramétriques.

Au cours de l'application, nous avons été confronté à un autre problème d'agrégation temporelle mais, cette fois, sur une échelle plus réduite. Il s'agit de la longueur de la période sur laquelle la fonction de congestion était appliquée. Usuellement, on s'intéresse aux conditions exceptionnelles des périodes de pointe si bien que l'affectation du trafic se fait sur une base horaire. Les données de demande de transport doivent par conséquent être dressées par heure spécifique.

Dans notre cas, le mariage du processus de développement urbain et des déplacements fait que la scission horaire ne convient plus. En effet, la distribution spatiale des activités génère toute sorte de trafics (magasinage, travail, loisir, etc...) qui prennent place à toute heure du jour. La réduction d'un trafic 24 heures à une seule heure par simple division aurait été trompeuse de par les fortes variations horaires de la circulation et de par la répartition différentielle de chaque type de déplacement.

Les matrices de déplacements produites ne comprenaient pas les migrations de retour. Ainsi dire, la moitié presque des flux était omise.

En tenant compte de la quasi-absence de circulation nocturne et la concentration en période d'affluence, il a été décidé d'établir le trafic sur six heures. Bien qu'il s'agisse d'un choix plutôt arbitraire, il a semblé adéquat selon les paramètres fixés préalablement de la fonction de congestion. La congestion estimée paraissait donner un portrait moyen des conditions moyennes de transport dans une journée de semaine.

8.3.4 Les sous-modèles de transport

En théorie, la forme fonctionnelle des sous-modèles de demande de transport est définie par la solution du programme d'optimisation sous-jacent. Toutefois, cette démarche apparaît un peu rigide si le modélisateur croit que des transformations seraient nécessaires pour se conformer aux relations décelées des données observées.

Transformation des variables

Pris un à un, les sous-modèles de demande (distribution spatiale, répartition modale, affectation routière de type logit), peuvent adopter des formes fonctionnelles alternatives si l'on prend soin de transformer préalablement la ou les variables connus du problèmes d'optimisation. Par exemple, l'affectation de la demande parmi les parcours de transport collectif a été faite en optimisant l'entropie de la distribution et le logarithme du coût de transport. Dans le cas-ci, il n'y avait pas de données sur la répartition qui permettait d'identifier la forme appropriée de la transformation. Il semblait sûr cependant que la mesure initiale du coût ne rendait pas compte des disparités des effets marginaux d'un trajet O-D à l'autre. En effet, les ordres de grandeur des coûts diffèrent tellement au sein des paires O-D, qu'il semble inconcevable de présumer que chaque kilomètre marginal a un effet semblable dans un court comme dans un long trajet. La transformation logarithmique du coût permet alors d'affecter les voyageurs selon le rapport de coût entre les parcours plutôt que selon l'écart.

Nous avons une certaine latitude à l'égard de ce sous-modèle puisqu'il a été démontré qu'il pouvait se détacher du grand programme d'optimisation de la demande (section 3.7.3). La seule relation qui est entretenue con-

cernait l'alimentation du problème principal par le coût composé par itinéraire. Il suffisait que la mesure issue soit calculée selon une unité comparable avec le coût d'équilibre établi par le processus d'affectation routière. Ainsi, la mesure composée fournie par l'équation 3.156 (laquelle tient compte du nombre variable de chemins parmi les trajets O-D) subit ensuite une transformation inverse, $c = \exp(\tilde{c})$, afin de retrouver l'unité initiale. Cette procédure ne fait pas d'entorse à notre cadre de modélisation puisque la transformation est temporaire, le temps d'évaluer une mesure composée appropriée.

Le processus de repartition modale a apporté davantage de complications pour diverses raisons. D'une part, la seule confrontation des coûts modaux est apparue insuffisante dans l'explication des partages observés. Il fallait en effet tenir compte de particularités existant au sein de chaque zone. La proportion de ménages ne possédant pas d'automobile à l'origine et une mesure synthétique de la difficulté de stationner (incluant le prix) à la destination complétaient le coût de transport. Théoriquement, il n'y a cependant pas de difficulté à intégrer des variables complémentaires car ils n'impliquent que des termes linéaires additionnels dans le problème d'optimisation.

La principale difficulté provient ensuite de la calibration des paramètres. Nous avons décidé de procéder à la calibration dans une phase préliminaire en utilisant des les partages modaux observés en 1981. Il s'agit d'estimer un modèle logit binomial doté de trois attributs. Les variables complémentaires n'ont pas engendré de problème spécifique, les relations observées se conformant assez bien avec leurs variations. Toutefois, la mesure du coût modal de transport, dans sa forme originale, ne pouvait convenir à cause de l'incomparabilité des effets marginaux parmi les paires O-D. En effet, la fonction logit munie d'arguments linéaires effectue le partage selon le strict écart entre les deux alternatives sans distinguer l'ordre de grandeur du trajet. Intuitivement, la solution réside (encore une fois) dans la transformation logarithmique de la variable de coût.

Pris séparément, le modèle logit s'accommode bien de cette modification d'échelle. Mais celle-ci ne peut se concilier avec la formulation générale

du programme d'optimisation. En effet, il n'y a aucune formulation alternative du problème qui puisse justifier l'emploi d'une transformation outre le fait d'écrire initialement le coût en log. Un tel procédé demeure impossible du fait que le coût par auto provient d'un trafic d'équilibre dans lequel toute mesure ne peut qu'être linéaire.

Dans la procédure itérative de solution du programme, nous aurions pu forcer un partage modal avec des coûts-log mais, dès lors, les propriétés de convergence se sont plus assurées. L'usage d'un coût linéaire reste nécessaire bien qu'il s'accorde mal avec les répartitions modales observées.

Pour résoudre ce dilemme, nous avons dû exploiter la flexibilité du problème d'optimisation qui n'impose aucune restriction quant à la désagrégation des paramètres. Si dans la fonction objective (3.144), le coefficient θ était remplacé par θ_{ij} , les conditions d'optimalité donnerait un modèle partage logit doté de paramètres multiples, permettant ainsi de moduler les relations modales paire par paire. Cette désagrégation rend pourtant impossible l'estimation des paramètres à cause d'une évidente insuffisance d'observation (3600 paramètres pour 3600 paires). Toutefois, nous savions que le problème ne provenait pas d'un comportement variable entre les paires mais d'une variabilité de magnitude. Ainsi, en établissant un dénominateur propre à chaque paire O-D, il était possible de standardiser leur ordre de grandeur et d'appliquer un paramètre de système. Ce dénominateur correspondait à la moyenne arithmétique des deux coûts modaux du trajet O-D en 1981. Du fait, nous imitions la transformation logarithmique avec un système linéaire muni d'un scalaire.

Les coûts composés par trajet peuvent être calculés selon une version de formule conventionnelle (3.158), i.e. l'équation (7.11). Il faut toutefois se rappeler qu'à cause de l'utilisation de facteurs de correction, la contribution des deux variables complémentaires s'en trouve modifiée. En effet, en réécrivant (7.11), on voit:

$$\gamma_{1,j} = \frac{1}{-\theta_{1,j}} \ln \sum_m \exp(-\theta_{1,j} c_{T,j} - \theta' g_{m,j} - \theta' h_{m,j})$$

$$\text{où } \theta_{1,j} = \theta / \bar{c}_{1,j}$$

Du fait, le dénominateur $-\theta_{1,j}$ varie selon le trajet O-D en question. Les variables $g_{m,j}$ et $h_{m,j}$ ne font plus une contribution linéaire dans le calcul. Le rôle qu'elles exercent augmentent avec la valeur de $\bar{c}_{1,j}$, la moyenne arithmétique des coûts modaux.

Enfin, aucune transformation n'a été requise pour ce qui est de la distribution spatiale. Le paramètre associé a été estimé en simultanéité avec ceux de la répartition modale.

L'intégration des phases de la demande de transport reste certes la propriété la plus intéressante du modèle d'optimisation (3.144-3.151). Cela permet d'effectuer une calibration simultanée et consistante. Néanmoins, lorsque des données désagrégées sont disponibles et qu'on procède à une calibration préliminaire des sous-modèles, on s'aperçoit que des transformations des mesures de coût seraient préférables. Or, dans le programme mathématique, la minimisation s'exerce sur deux termes de coût, l'un étant la sommation des coûts des usagers du transport en commun, l'autre, l'intégral de la fonction de congestion. A partir de ces termes, toutes les mesures de coût intermédiaires sont calculées. Comme il a été démontré précédemment, il est possible de transformer initialement les coûts, du moins à l'égard du transport en commun. Mais à cause de la formulation du trafic d'équilibre, il n'est pas question de modifier la linéarité des coûts; de fait, il faut revenir à l'échelle originale. Cette possibilité de transformer temporairement l'échelle ne vaut cependant que la phase d'affectation du mode collectif car il peut être isolé en sous-problème. Dans le reste des cas, il importe de conserver la linéarité du coût puisqu'aucune formulation théoriquement acceptable du problème d'optimisation ne génère des conditions d'optimalité impliquant une transformation de cette variable-clé.

Trafic d'équilibre

Le programme mathématique de demande de transport a été solutionné par l'algorithme de Evans. La procédure itérative est arrivée à terme lorsque toutes les conditions d'optimalité sont satisfaites. Bien que la distribution des voyageurs ne soit pas unique, les flux affectés sur le réseau correspondante à une solution unique [Erlander, Nguyen, Stewart, 1979].

Nous avons remarqué qu'en dépit de son unicité, la solution d'équilibre peut être non-convergente à l'approche du point d'équilibre. Autrement dit, de modestes variations de la demande et/ou de l'offre peuvent occasionner des changements discontinus d'équilibre. Ces changements restent néanmoins non-paradoxaux au sens qu'ils n'impliquent pas des hausses de coût de transport lorsque l'offre est bonifiée. Plutôt, il s'agit de modifications soudaines et peu explicables du trafic d'équilibre. Par exemple, une augmentation de la capacité du réseau routier dans l'ouest de la région pourrait entraîner une hausse de la circulation sur le pont Jacques-Cartier.

Ce comportement contre-intuitif de l'affectation routière peut cependant être justifié. D'une part, l'affectation procède selon un principe de minimisation dépourvu d'un terme de dispersion. L'amélioration d'un lien ou l'ajout d'un autre peuvent déclencher un transfert brusque de voyageurs entre les itinéraires, allégeant ou alourdissant la charge d'autres liens selon le cas. De ce fait, tous les trajets O-D pourrait être touchés par le changement initial. Fisk [1980] proposa un problème d'affectation de flux d'équilibre doté d'un terme d'entropie qui pourrait vraisemblablement enrayer les réaffectations discontinues.

D'autre part, certains changements du trafic d'équilibre sont difficiles à expliquer en raison de l'intégration de la distribution spatiale et de la répartition modale. Nous sommes en présence d'un problème avec demande variable. En l'occurrence, la demande pour certains trajets O-D devrait s'accroître si un ou plusieurs liens inscrits dans l'un de leurs itinéraires admissibles sont améliorés ou ajoutés. Les contraintes d'origines et de destinations font que la demande des autres trajets en dehors du corridor bonifié diminue d'autant. Cette réorganisation de la

circulation suscite souvent des hausses et des baisses sur des liens apparemment éloignés des liens modifiés.

Conception du réseau

Le problème de conception de réseau est, comme on sait, un problème central dans le schéma de modélisation proposé. Sa solution génère des impulsions importantes dans le système car il modifie l'offre de transport d'une façon non-continue.

L'obtention du sous-ensemble de liens optimaux pose d'abord certaines difficultés de calcul. L'algorithme de résolution de type arborescent apparaît très lourd; la recherche des bornes inférieures et des valeurs intermédiaires de la fonction objective exige chaque fois le solutionnement du problème de la demande. La procédure peut par conséquent devenir très lente si de nombreuses combinaisons de réseau sont possibles. Heureusement, dans l'application dans la région de Montréal, les réseaux étaient considérablement simplifiés, prévenant ainsi le déploiement de l'arbre de recherche.

Par ailleurs, l'utilité du modèle de conception de réseau devient limitée si l'ensemble des liens projetés est peu nombreux. En effet, les solutions paraissent quasiment triviales en présence d'un nombre restreint de projet. Si bien qu'un analyste avisé serait capable de sélectionner de visu le réseau optimal. Dans notre cas, bien que le modèle de conception a été opportun, nous remarquons que les liens (routiers comme de transport collectif) élus en premier correspondaient à ceux qui parcouraient les secteurs les plus denses. Les surprises étaient rares.

Nous pouvons adresser un autre reproche au modèle. Son application est statique. Il cherche en effet à optimiser un réseau selon une demande actuelle, un ensemble de projets plausibles et un budget établi sur une période fixe. Dans la réalité, le développement d'un réseau doit avoir une perspective de long terme. D'une part, les investissements consacrés sont échelonnés sur de plus longues périodes. D'autre part, la séquence de réalisation des projets la plus favorable ne coïncident pas nécessairement à un ordre privilégiant les projets optimaux sur une base quinquennale.

Parfois, il est préférable de développer des liens peu sollicités dont l'utilité ne sera mise en évidence que lors du parachèvement du réseau. Steenbrink [1974] presenta un cadre de modélisation permettant d'évaluer les coûts et bénéfices d'un projet de transport sur une plus longue période; les mesures y sont cumulatives dans le temps et actualisées. Depuis, il y eut peu de développement de cette approche.

8.3.5 Les sous-modèles de localisation

Structure fonctionnelle

Les sous-modèles de localisation des activités possèdent une structure linéaire autorégressive. Leurs composantes spatiales sont par contre non-linéaires et s'inspirent du modèle de Lowry. Dans notre application, nous avons considéré ces deux choses comme séparées. La procédure d'estimation a été conséquemment entreprise en deux étapes.

La première étape consiste à évaluer les paramètres de friction associées aux matrices d'interaction spatiale. Comme l'interaction entre activités peut être assimilée à des déplacements, la calibration s'est faite à partir de matrices de flux appropriées i.e. désagrégées par motifs. Jusqu'à ce point, il n'y a aucune difficulté opérationnelle ni même conceptuelle dans cette calibration préliminaire outre la présomption de paramètre fixe à la période de simulation. La fixité paramétrique reste une hypothèse courante dans ce type de modélisation à défaut d'information sur l'évolution des comportements.

Dans la deuxième étape, il s'agit d'estimer les paramètres traduisant les relations temporelles et fonctionnelles. Pour les besoins de la simulation, ces relations ont été passablement simplifiées (voir section 7.3.6) en comparaison du modèle complet (6.132, 6.134) qui prévoit tout les genres d'interaction. Parmi celles qui ont été omises, notons les matrices autorégressives spatiales qui synthétisent les relations inter-zonales au sein d'une même activité. Bien que ces relations doivent s'exercer en réalité (effets d'agglomération et de compétition), il n'y a pas de mesure permettant de les appréhender. En fait, il aurait été envisageable d'y présumer une structure spatiale et d'estimer un scalaire à partir des dis-

tributions observées en 1981. Cette démarche peut donner des résultats trompeurs car il n'est pas sûr que les associations statistiques détectées soient attribuables à de véritables relations spatiales au sein de la même activité.

En pratique, les effets d'agglomération se confondent facilement avec ceux résultant des interactions entre activités dont la répartition spatiale montre déjà des formes d'agglomération.

De même, les relations fonctionnelles aspatiales ont été omises. Celles-ci peuvent traduire des effets de compétition pour le sol entre activités à l'intérieur d'une même zone. Dans le cas où ces relations sont positives, il s'agit d'effets de symbiose ou de concentration. L'absence d'un tel terme dans le sous-modèle de localisation signifie que ces relations sont indirectement appréhendées à travers les matrices d'interaction spatiales dont les éléments diagonaux symbolisent des relations intra-zonales. Malheureusement, tous les éléments d'une même matrice expriment une fonction gravitaire de l'attraction et du coût de transport; les relations intra-zonales sont sur un même pied que celles entre zones. Il y a donc risque que la spécificité des relations internes soit diluée et confondue avec des interactions liées à la proximité des activités.

Rationalité dominante

En fait, le processus de localisation tel que représenté dans les sous-modèles est dominé par des interactions gravitaires et décrit ainsi surtout des relations de taille entre les zones. Sa capacité explicative reste limitée. Le développement urbain semble donc animé selon une rationalité de "marché" i.e. de proximité des autres activités.

Toute autre forme d'interaction spatiale, notamment les associations autoregressives, paraissent difficiles à concilier dans notre cadre. Leur estimation exige des valeurs normalisées (taux, densité) afin d'y soustraire les effets de corrélation entre tailles. Cette exigence apparaît conflictuelle avec les interactions de type gravitaire dont l'intensité dépend justement de l'importance respective des zones.

On ne peut ainsi prétendre que les sous-modèles interprètent rigoureusement les phénomènes sous-jacents à la localisation des activités. Il s'agit plutôt d'une formulation permettant de saisir les interactions entre activités et entre lieux sans nécessairement les discerner correctement. Les fonctions du coût de transport (ou de distance) sont souvent considérées comme des fonctions "proxy" ou déléguées. Elles décrivent toute sorte de relations qui se manifestent en association avec la distance des lieux.

Le fait de dériver les modèles selon des programmes optimisant les bénéfices de localisation reste déterminant. On sait que la satisfaction des contraintes marginales (autre qu'unitaires) conduit inévitablement à une structure fonctionnelle gravitaire. La version dynamique des sous-modèles n'y change rien mise à part l'introduction de délai de réaction et de permanence des activités.

Paramétrisation

Nos modèles "dynamiques" de localisation sont dotés de quatre types de paramètres. La complexité provient du fait qu'ils se situent tous à des niveaux différents mais qu'ils sont interdépendants. Au niveau le plus "primitif", se trouve le paramètre de la fonction du coût de transport. Son estimation ne peut être qu'exogène puisque la distribution spatiale des variables en question (les activités) ne présente aucune interaction explicite bien qu'elle en soit un produit. Le recours aux déplacements stratifiés par motif permet d'associer un processus d'interaction approprié.

Une fois la fonction de transport spécifiée, reste l'identification d'un facteur d'attraction. En fait, celui-ci n'est pas nécessairement un paramètre. Une variable d'attraction connue pourrait combler ce poste. Il paraît très difficile cependant de trouver une telle variable quand on cherche un très bon ajustement entre les distributions spatiales des activités localisantes et localisées. Il importe que la variable choisie entretienne une certaine proportionnalité avec la variable localisée. De plus, puisque ce facteur d'attraction est multiplié par la fonction de transport et de la variable localisante, il convient qu'il ne répète pas leur distribution. Dans notre application, nous avons fixé pour chaque

zone le terme d'attraction égal au rapport de l'activité localisée sur l'indice d'accessibilité formé par le produit de la fonction de transport et l'activité localisante.

L'autre type de paramètre ne concerne que les sous-modèles dotés d'au moins deux variables localisantes. Il s'agit de la pondération respective de chacune d'elles. Encore une fois, la distribution spatiale de l'activité localisée ne permet pas d'estimer leur contribution puisque les méthodes régressives ne saisissent que les associations statistiques. En effet, les coefficients de régression ne traduisent que l'apport original des variables sur la variance de la variable dépendante; souvent dans un système à deux régresseurs, une contribution peut être négative si une variable contient déjà la variance commune de l'autre. Ce problème survient constamment étant donné la forte collinéarité (causée par la corrélation des tailles) des variables localisantes. Cette approche contrevient à celle des modèles de localisation dont les "régresseurs" symbolisent des sommes d'interaction contribuant positivement à la localisation de l'activité. Pour les fins de l'application, nous avons fixé arbitrairement ces valeurs.

Les paramètres temporels mesurent d'une part la proportion d'une activité qui est permanente et d'autre part celle qui est sujette à relocalisation (mais qui peut rester dans la même zone). Puisque les variables distribuées spatialement doivent satisfaire une contrainte de total, le deuxième paramètre est complémentaire au premier ($2^{\circ} = 1 - 1^{\circ}$).

A la suite de plusieurs essais, nous avons noté l'extrême interdépendance des paramètres. Des valeurs différentes de paramètre de transport donnaient lieu par exemple à une autre distribution des termes d'attraction (qui contiennent implicitement une fonction d'accessibilité). Les paramètres temporels changeaient avec beaucoup de sensibilité. C'est comme si la structure paramétrique était très flottante.

Le pouvoir déterminant du paramètre de transport s'explique aisément. Lorsque sa valeur est élevée, la contribution localisante des zones voisines devient faible tandis que domine l'effet interne. A l'extrême, seule l'interaction intra-zonale est considérée, ce qui devrait réduire le terme

en question à la variable localisante. Toutefois, par sa morphologie, le facteur d'attraction tend à égaler le ratio des variables localisée/ localisante. En fin de compte, le terme d'interaction se réduit à l'activité localisée (puisque le facteur d'attraction multiplie la variable localisante) si bien que sa part explicative devrait se confondre avec la part autorégressive.

A l'opposé, lorsque le paramètre de transport tend vers zéro, toutes les zones de la région contribuent également à la localisation de l'activité sans égard à la distance. Du fait, l'accessibilité en s'uniformisant, le facteur d'attraction tend à se distribuer comme l'activité localisée (qui reste le numérateur de la fonction). Encore une fois, le terme d'interaction spatiale adopte la distribution de la variable localisée à cause de la constance de tous les autres termes.

A cet égard, dans un exercice de simulation, il importe donc de fixer les facteurs d'attraction de manière à ce qu'ils ne puissent varier au gré des fonctions de transport. Leur variation tend souvent à annuler la contribution des zones voisines en compensant pour les changements d'accessibilité.

8.3.6 Intégration du système de transport et de la localisation d'activité

L'intégration de la demande et de l'offre de transport n'a pas posé de problème particulier en terme d'application. Dans ce cas-ci, l'intégration reste complète i.e. que les sous-modèles sont tous issus d'un même programme d'optimisation. L'offre était représentée par un problème de conception à demande variable et avec réseau à capacité limitée; la demande devient un sous-problème non-linéaire de minimisation intégré dans l'autre mais peut être solutionnée à l'écart.

Le modèle de conception de réseau ne possède aucun paramètre et, du fait, n'exige aucune calibration. Néanmoins, il dépend des résultats du sous-problème de demande pour les calculs des bornes inférieures et des valeurs de la fonction objective. Le problème de demande contient quant à lui un certain nombre de paramètres. Lorsque ceux-ci sont connus préala-

blement, il suffit d'appliquer un algorithme comme celui d'Evans [1976] ou de Florian et Nguyen [1978] pour parvenir à la solution. Selon l'aporoche de Boyce et al. [1983], il semble toutefois possible de calibrer les paramètres en satisfaisant des contraintes d'entropie simultanément avec la résolution du problème d'optimisation. Nous avons écarté cette alternative (bien qu'elle soit la plus intégrée) à cause de la lourdeur excessive de la procédure conjointe de solution/calibration. En effet, il faut qu'à chaque itération de l'algorithme de solution, il faut amorcer une procédure itérative de type Newton-Raphson. On sait déjà que la solution du problème de demande constitue seulement une étape au sein d'une itération de l'algorithme "branche-et-borne" du problème de conception de réseau.

L'exogénéité des paramètres de transport permet d'appliquer rigoureusement le programme intègre d'offre et demande de transport. On reproche souvent à ces paramètres d'être dépendants de la structure spatiale [Erlander, Stewart 1978]. Ceci devient embêtant lorsque le réseau se développent en même temps que les origines et les destinations. Malgré tout, il n'a pas de véritable réponse à cela puisque le coût total et le niveau d'entropie, les constantes servant à calibrer les paramètres, sont eux-mêmes affectés par les changements de structure spatiale.

Comme il a déjà été dit antérieurement, le problème de localisation des activités n'est pas complètement intègre avec le transport. Il s'y concilie cependant au niveau des intrants et des extrants, notamment en s'alimentant des les matrices de coût compose issus du problème de demande de transport et en fournissant les données permettant de générer les origines et les destinations. Du point de vue théorique, les problèmes paraissent isomorphes. La maximisation des bénéfices de localisation s'assimile aisément avec une minimisation des coûts (ou maximisation de l'utilité).

Cette conciliation n'a causé aucun problème sérieux d'application tant au niveau numérique que de l'interprétation. Elle facilite même le travail en décomposant un phénomène fort complexe qui est celui des relations "transport/ structure urbaine". D'ailleurs, l'intégration incomplète était voulue de manière à distinguer la staticité du transport du caractère dynamique du développement urbain.

PAGINATION ERROR.

ERREUR DE PAGINATION.

TEXT COMPLETE.

LE TEXTE EST COMPLET.

NATIONAL LIBRARY OF CANADA.

BIBLIOTHEQUE NATIONALE DU CANADA.

CANADIAN THESES SERVICE.

SERVICE DES THESES CANADIENNES.

8.4 REFLEXIONS SUR L'UTILITE EMPIRIQUE DE LA MODELISATION

Le modèle spatio-temporel général devrait s'avérer utile dans un contexte de planification. L'usage principal est d'évaluer l'impact des projets d'expansion du transport sur le développement de la structure urbaine. Cependant, l'utilisateur ou l'analyste devra tenir compte des facteurs limitatifs qui entourent le modèle. Il bénéficiera aussi de certains avantages relatifs aux propriétés particulières du modèle.

8.4.1 Les limites

De par la complexité du phénomène, l'analyste ne devrait préconiser qu'une approche de simulation i.e. d'appliquer le modèle selon diverses conditions et présomptions préétablies. S'aventurer dans les prévisions comporte trop de risques car, d'une part, trop de valeurs paramétriques restent méconnues et, d'autre part, leur comportement évolutif est fixé arbitrairement.

Notre méconnaissance des caractéristiques distributionnelles des valeurs projetées empêche également de procéder à de véritables prévisions. Il demeure donc impossible d'y évaluer la validité statistique des estimés. Cette méconnaissance tient à la trop grande complexité mathématique du modèle général.

La disponibilité des données et la connaissance du phénomène fixent les limites de l'application. Le modèle présenté, comme tous les autres modèles, ne peut fournir des résultats contenant plus d'informations que celles dont il est pourvu initialement. Cependant, il saura les organiser afin d'offrir un éclairage essentiel du phénomène complexe que sont les relations entre transport et structure urbaine.

L'étendue, la richesse et la pertinence de la base de données sont les facteurs limitants. Nos besoins semblent par ailleurs relativement grands; des données de demande de transport ventilées par mode et origine-destination, des réseaux de transport topologiquement et géographiquement codés avec des caractéristiques détaillées sur la capacité et le temps de voyage sur leurs liens et, enfin, des données d'activités urbaines. Le tout

selon une base spatiale invariable sinon compatible et une base retrospective régulière! De telles exigences appellent donc au compromis.

Dans un espace métropolitain, les données de demande de transport, quand elles existent, proviennent habituellement des enquêtes origines-destinations. Souvent, ces enquêtes ont débuté que très récemment et sont exécutées selon une périodicité irrégulière. La construction d'une base rétrospective apparaît donc limitée. Aussi, il n'est pas sûr que le système spatial utilisé soit compatible avec ceux des autres types de données.

Dans bien des occasions, l'analyste devra lui-même composer les réseaux de transport nécessaires à son analyse. Il aura le loisir de choisir le niveau d'agrégation et d'éliminer les arcs jugés facultatifs. Par contre, il devra s'en remettre à des sources extérieures pour assigner les caractéristiques associées aux liens des réseaux. Il remarquera par le fait même la pauvreté des données disponibles de réseaux, limitées à quelques valeurs de capacité théorique et de vitesse maximale.

Enfin, l'utilisateur obtiendra des résultats dont la qualité équivaldra à celle des données d'activité urbaine. Il y a encore peu de régions métropolitaines où l'on trouve une authentique base de données consistantes entre variables, dans l'espace et dans le temps. Au Canada, les données relatives à la population sont très détaillées grâce aux recensements quinquennaux. En terme de transport, la population est la variable-clé dans la "production" des déplacements. L'aspect "attraction" est par contre décrit par un ensemble de variables dont les plus pertinentes restent difficiles à obtenir.

En fait, les responsables de la planification intéressés par l'analyse de l'évolution conjointe du système de transport et de l'organisation urbaine devront avant tout s'assurer d'une bonne base de données spatio-temporelles. D'ailleurs, il ne sert à rien de pouvoir simuler les développements futurs si les états actuel et antérieur demeurent méconnus.

Les paramètres paraissent un autre groupe de facteurs limitatifs. Outre le processus de localisation, notre modélisation semble relativement

peu paramétrisée, le sous-modèle de demande n'en demandant au moins trois (dépendant de la fonction de partage modal) tandis que le problème de conception de réseau étant entièrement non-paramétrisé.

Quand elles sont accomplies de façon externe, les phases de calibration et d'estimation sont généralement simples et directes. Elles utilisent des procédures éprouvées qu'on peut facilement retrouver dans des logiciels les plus complets et répandus. Cependant, pour distribuer la demande de transport, la démarche la plus cohérente consiste à effectuer les calibrations simultanément avec la recherche de la solution. Cette procédure conjointe nous est pourtant apparue extrêmement fastidieuse et lente car à chaque iteration de l'algorithme principal, il faut amorcer une nouvelle procédure de calibration. A cet égard, nous conseillons d'évaluer préalablement des valeurs paramétriques fiables afin de résoudre le problème d'optimisation sans contrainte d'entropie.

Le système d'équations synthétisant le processus de localisation des activités reste néanmoins fort paramétrisé. Dans le schème théorique le plus complet, la structure paramétrique peut être différenciée selon le type de relation fonctionnelle et d'interaction spatiale et adaptée à la structure de décalage. Une telle représentation ne se prête qu'à l'estimation (OLS, MOLS) que si l'on dispose d'une base chronologique de données longue et fine. Dans la plupart des cas, l'analyste devra se contenter d'une représentation plus simple et s'en remettre à des valeurs paramétriques connues a priori (comme les exposants des fonctions de transport) et à des procédures d'estimation basée sur peu de périodes mais compensée par un grand nombre d'unités spatiales.

8.4.2 Les avantages

L'intérêt de notre cadre de modélisation apparaît avant tout la consistance et la transparence des principes sous-jacents. Les trois grandes parties du modèle, la demande et l'offre de transport et la localisation d'activités, émanent toutes de problèmes d'optimisation apparentés et clairement définis. Les problèmes concernent l'optimisation d'un coût ou d'un bénéfice mais admettent une variabilité des comportements.

Plus que la consistance interne, la consistance avec les modèles reconnus en transport et en localisation devient un atout important. Elle permet de profiter des développements courants et, sans aucun doute, d'améliorer la formulation de notre modèle. Elle permet aussi de dévoiler certaines propriétés insoupçonnées.

Puisque toutes les applications ont leur propre spécificité, l'analyste appréciera également la souplesse d'un tel cadre de modélisation. Ici, nous n'avons présenté qu'une version possible. On pourrait raffiner par exemple la fonction de partage modal ou bien spécifier des fonctions de congestion propres à chaque lien. Les catégories d'activité urbaine pourrait davantage être décomposées.

Même si la formulation présentée s'accommode plutôt à des problèmes intra-urbains d'un niveau d'agrégation spatiale relativement grossier, il n'est pas impossible de l'adapter à un système géographique plus détaillé. Dans ce cas, il faut s'attendre à des besoins de données plus grands (notamment sur les caractéristiques zonales et la densité des réseaux de transports) et à des résultats dotés de plus grandes marges d'erreur.

On peut aussi penser à des applications interrégionales avec une approche théorique des places centrales plutôt que de l'utilisation du sol. Les relations fonctionnelles et spatiales y sont clairement différentes, la hiérarchie des villes important autant que la proximité. La modélisation du système de transport doit aussi être repensée: composante congestion moins apparente, processus de choix modal différent, friction de la distance moins sévère.

L'analyste n'est pas obligé d'appliquer le modèle complet. Selon ses connaissances ou ses intérêts, il peut rendre exogène une partie du modèle. Par exemple, plutôt qu'être modélisé, le développement des réseaux peut facilement être décrit par une politique de transport envisagée. Dans ce cas, on examinera son impact sur la demande de transport et l'organisation urbaine.

Chapitre 9

CONCLUSION

9.1 CONTEXTE DE LA MODELISATION

9.1.1 Cadre général

Les tentatives de modelisation conjointe de la structure urbaine et des systemes de transport se baserent souvent de la programmation mathematique comme principe d'integration [voir Los, 1979; Anas, 1984; Boyce 1984]. Dans tous les cas, il s'agit d'unifier des modeles de demande de transport avec ceux de localisation. La consideration d'un trafic d'équilibre fait introduire un élément de l'offre de transport.

Ces formulations frappent avant tout par leur elegance: il existe une forte consistance entre les composantes du systeme et on parvient, en résolvant le problème d'optimisation, à un équilibre general. Un tel équilibre correspond à une parfaite cohérence entre la demande de déplacement, les conditions de transport et la distribution des activités génératrices de déplacement.

Nous nous sommes certes inspirés de ces cadres intégrés de modelisation bien que nous doutions de l'existence d'un équilibre général dans la réalité. En d'autres mots, nous n'admettions pas le synchronisme de l'ajustement reciproque entre structure urbaine et systemes de transport.

Le synchronisme de la demande de transport et des conditions locales de circulation envers l'organisation spatiale des activités reste théorique-ment admissible. Les déplacements sont normalement générés sans délai selon une repartition donnée des activités quoiqu'une certaine inertie accompagne les modifications du comportement du voyageur (changement de trajet O-D, de mode, d'itinéraire).

Toutefois, l'atteinte d'un équilibre immédiat entre la structure urbaine et les conditions de transport semble peu concevable. Les composantes de la structure urbaine (activités, affectation du sol, bâtiments) ont un caractère trop permanent pour s'adapter correctement à un nouveau

patron d'accessibilité. Pourtant, un ajustement progressif existe bel et bien en réalité et c'est pour cette raison même qu'on s'intéresse à la modélisation de l'organisation urbaine.

Dans cette thèse, nous avons tenté d'introduire un concept de décalage à celui d'un équilibre général. Ultimement, en absence de modification des réseaux de transport, un tel équilibre paraît envisageable, du moins en théorie.

Par ailleurs, les temps de réaction sont suffisamment longs pour que des changements des réseaux surviennent. En effet, si on admet que des modifications de l'accessibilité peuvent influencer le développement urbain, on ne peut nier que les conditions actuelles de transport n'entraînent pas des améliorations des réseaux. Ce deuxième type d'ajustement se manifeste aussi avec certains délais mais prend une toute forme.

Bien que la programmation mathématique ait été choisie comme principe intégrateur, le modèle général ne se présente pas cependant comme un seul problème d'optimisation mais comme trois problèmes interreliés. Ce fractionnement des programmes s'explique par leurs échelles distinctes du temps et non pas par leurs principes d'optimisation lesquels apparaissent très apparentés et conciliables.

9.1.2 Modélisation de la demande de transport

Nous avons présenté les modèles classiques de demande de transport en abordant les quatre phases traditionnelles de répartition de la demande (génération et distribution spatiales, partage modal, affectation routière). Il a été trouvé que différentes approches d'optimisation conduisent à des formulations de modèle similaires sinon isomorphes. Toutefois, des principes plus universels s'harmonisent plus que d'autres à l'ensemble des phases. On peut alors éviter une démarche séquentielle, prévenant du même coup les problèmes d'ordre d'application et de non-convergence des résultats en présence d'effets rétroactifs (comme la congestion routière).

L'unification de la demande de transport en un seul problème mathématique équivaut à une approche "simultanée". La formulation du problème combine distribution/affectation de Beckmann et al. [1956] constitue le point de départ. Florian et al. [1975] ainsi qu'Evans [1976] présenteront des versions selon une notation plus commode et munies d'un algorithme de solution.

Nous avons proposé un problème d'optimisation combinant la distribution et l'affectation des flux avec le partage modal. L'intégration est accomplie grâce à la structuration hiérarchique du problème. L'approche de Boyce et al. [1983], de laquelle nous nous inspirons, repose sur la formulation des contraintes d'entropie. Dans notre cas, cette structure est acquise par l'incorporation, dans la fonction objective, de termes d'entropie traduisant le conditionnalité des phases de répartition. Un tel arrangement s'accommode de sous-ensembles d'alternatives de cardinalité inégale et n'empêche pas une affectation par trafic d'équilibre.

Dans un programme hiérarché, les liens entre les phases sont décrits par contraintes qui se présentent sous la forme de multiplicateurs de Lagrange dans les conditions d'optimalité. Lorsque des termes d'entropie sont insérés dans le programme, ces multiplicateurs coïncident à des coûts composés, mesures pratiques d'agrégation et consistantes avec le processus dérivé de distribution. Ces mesures sont isomorphes aux concepts "d'utilité composée" de Williams [1977] et de "valeur inclusive" de McFadden [1978].

Si on omettait le trafic d'équilibre, un tel programme se résoudrait simplement en ajustant les multiplicateurs lagrangiens de manière à satisfaire les contraintes marginales. Avec un réseau à capacité limitée et une affectation selon un principe des temps égaux, les conditions d'optimalité deviennent plus exigeantes car elles sollicitent un équilibre des temps de parcours avec leur fréquentation.

Pour cette raison, la solution du problème combiné nécessite l'emploi d'un algorithme conçu à cet effet. Nous avons essayé deux algorithmes, celui de Florian et Nguyen [1978] demandant la linéarisation de la fonction objective pour estimer la direction de descente, et celui d'Evans [1976]

dont la direction de descente est donnée par la solution même du problème non-linéaire. La formulation originale de la procédure d'Évans ne concerne pas le partage modal mais son incorporation ne pose aucun obstacle.

Seule la phase de génération des déplacements n'a pas été intégrée dans le problème de demande. La raison est qu'il s'agit d'une fonction linéaire qui ne nécessite aucune distribution. En fait, dans un problème comme celui-ci, on pourrait substituer, dans les contraintes marginales, les origines et destinations par des fonctions linéaires de variables d'activités. Cela ne change rien à l'exogénéité car les paramètres auraient dû être estimés préliminairement. La régression fut privilégiée pour l'estimation des modèles de génération.

9.1.3 Modélisation de l'offre de transport

Dans notre cadre, l'offre concerne exclusivement les réseaux de transport. Un problème de conception de réseau décrit le comportement d'offre. Il s'agit d'un programme consistant à sélectionner un sous-ensemble de liens parmi un ensemble de liens projetés en minimisant les coûts de transport des personnes et en respectant un budget de construction.

Dans une nomenclature des modèles de conception, il existe différentes présomptions concernant l'endogénéité de la demande et les restrictions de capacité des liens de réseau. Nous avons choisi une version dotée d'une demande variable (endogène) et de réseaux dont les liens sont assujettis par des fonctions de congestion.

Nous avons enrichi le cadre en y insérant deux réseaux, l'un routier pour les automobiles, l'autre étant les lignes de transport en commun. Pour le sous-problème de demande, cela implique une répartition des voyageurs entre les deux modes et une affectation spécifique pour chacun des réseaux. La seule distinction réside au niveau de l'affectation des usagers sur le réseau public qui n'a plus besoin d'être de type logit. Nous avons ici choisi d'affecter sur le plus court chemin de chaque paire O-D. De cette manière, le problème principal doit contenir un sous-problème identique à celui de la demande. Les problèmes d'offre et demande de transport sont donc parfaitement consistants.

Dans le problème principal, la variable décisionnelle est dichotomique, symbolisant l'alternative addition/omission d'un lien. Il s'agit donc d'un problème de "programmation entière". La solution de ces problèmes demande des algorithmes différents car l'espace faisable n'est plus continu mais plutôt ponctuel. Il n'y a donc pas de direction de descente proprement dite.

Nous avons privilégié une algorithmie de type arborescente. Il s'agit de construire un arbre de recherche dont on calcule pour chaque branche une borne intérieure identifiant la valeur la plus basse qu'on puisse obtenir avec une combinaison de liens. Si cette valeur dépasse la meilleure valeur de la fonction objective, alors on retourne au noeud précédent pour cheminer dans une autre branche. Si une solution intermédiaire ne peut accepter un lien supplémentaire sans excéder la contrainte budgétaire, alors cette solution est complète et on peut y calculer la fonction objective.

Les modèles d'offre ne sont pas systématiquement des problèmes entiers: le problème peut consister à augmenter la capacité des liens de façon tractionnaire [Abdulaal, Leblanc, 1979]. Certains trouvent cela irréaliste car on ne peut qu'accroître la capacité qu'en ajoutant des voies complètes.

Nous considérons que le modèle de conception avec les modifications de réseau qu'il propose, peut interpréter le développement des systèmes de transport. L'application dynamique proposée consiste à étendre les réseaux à chaque période de temps selon la demande courante et le budget de développement.

La volet dynamique se résume au délai nécessaire entre l'observation de la demande, la prise de décision et la mise en service des liens ajoutés. L'évolution de la demande et l'expansion des réseaux modifient sans aucun doute l'accessibilité des lieux et, indirectement, des activités. Selon les théories classiques de localisation, des ajustements de l'organisation urbaine se produisent en fonction du nouveau patron d'interaction.

9.1.4 Le processus de localisation

Dans le chapitre 5, nous avons analysé des systèmes spatio-temporels et leurs représentations mathématiques. Celles-ci conjuguent les trois dimensions caractérisant ces systèmes, les variables, le temps et l'espace. L'intérêt pour ces formulations provient de leur capacité à représenter des phénomènes dynamiques de localisation. En effet, leur structure autorégressive permet de représenter la permanence des activités urbaines à travers le temps.

L'approche retenue s'inspire principalement de celle de Bennett [1979] et de Cliff et Ord [1975; 1980]. La formulation des systèmes spatio-temporels est telle qu'on peut recourir aux méthodes classiques d'identification et d'estimation. En fait ces analyses bénéficient d'une littérature abondante mais qui malheureusement traite rarement de problèmes combinant l'espace, le temps et une structure multivariée.

Nous avons présenté des modèles classiques de localisation de type descriptif. Les formulations gravitaires consistant à allouer des consommateurs à des zones commerciales [Lakshmanan, Hansen, 1965] ou des travailleurs à des zones résidentielles [Wilson, 1969] suscitent de l'intérêt pour leur souplesse et leur possibilité de désagregation. Par contre, un modèle plus général comme celui de Lowry [1964] suggère un système d'équations parvenant à un état d'équilibre de plusieurs activités.

Quant aux problèmes de localisation de type normatif, il vise à une allocation optimale de ressources dans l'espace. Nous avons démontré que, les problèmes descriptifs et normatifs, écrits sous la forme de programme mathématique, diffèrent très peu. Le concept d'entropie permet d'échapper à la rationalité normative en appréhendant les interactions sous-tendues par le phénomène de localisation étudié.

Nous avons adopté le concept de "bénéfices de localisation" comme une mesure de l'utilité d'un lieu. Les coûts d'interaction et l'attraction du lieu sont synthétisés de façon consistante dans ce bénéfice. La maximisation des bénéfices de localisation permet de dériver tous les modèles classiques de localisation comme conditions d'optimalité. Cette approche a

l'avantage d'engendrer des mesures associées (coût composé, indice d'accessibilité) consistantes avec le principe d'allocation. De plus, elle s'apparente à la théorie de l'utilité aléatoire et aux modèles de "surplus collectif" [voir Neuberger, 1971] et de "surplus du consommateur" [Williams, 1976].

Un processus conditionnel de localisation résidentielle et commerciale de type Lowry peut être reformulé sous la forme d'un problème de maximisation des bénéfices. Les conditions d'optimalité fournissent un système d'équations gravitaires reliées par des relations fonctionnelles entre activités. Ce système n'est soluble qu'en parvenant à un état d'équilibre des distributions.

Nous avons démontré la convergence d'un tel état d'équilibre en adoptant le format matriciel. Cette représentation permet, outre de définir les équations d'équilibre linéaires, mais également des multiplicateurs spatiaux qui ont la faculté de convertir directement la distribution de l'activité exogène en celle de l'activité endogène. En fait ces matrices apparaissent des versions géographiques des multiplicateurs économiques de type «input-output» [voir Leontief, Strout, 1963]. Elles contiennent l'ensemble des relations spatiales directes et indirectes qui composent le processus de localisation.

9.1.5 La conversion vers un système dynamique

L'existence d'un état d'équilibre obtenu à l'aide d'un multiplicateur spatial dissimule le fait que l'analyse reste intemporelle. Il ne s'agit pas d'une convergence dans le temps mais une convergence strictement mathématique. Ce cadre statique ne convient que si l'on s'intéresse à l'organisation urbaine conséquente à une modification de facteurs exogènes. Il reste inapproprié lorsqu'il s'agit d'un problème dynamique de développement urbain.

Il faut admettre par ailleurs que l'état le plus commun d'un système urbain en est un de déséquilibre. Le retard de réaction, la permanence des établissements et l'irrégularité des perturbations ne permettent pas à l'organisation urbaine d'atteindre un véritable état d'équilibre. Malgré

tout, le principe de maximisation des bénéfices de localisation demeure effectif.

La modélisation proposée a exploité les propriétés de convergence des distributions spatiales en les assimilant à des dynamiques. On peut considérer que l'état d'équilibre (défini comme la solution du problème d'optimisation) devient un point d'attraction vers laquelle évoluent les composantes du système. En absence de perturbation, la dynamique s'identifie donc à la convergence mathématique de la solution optimale. Des modifications de l'environnement repositionnent le point d'équilibre si bien que la trajectoire dynamique se rectifie.

La considération d'un système «gradient» assure le passage d'un cadre statique à un cadre dynamique. La trajectoire temporelle est représentée par un système d'équations différentielles. Dans une application pratique, ce système reste de peu d'utilité puisque nos observations temporelles sont discrètes. Une représentation autoregressive ou à l'aide d'équations de différence suffit.

Nous avons démontré qu'une structure autoregressive adaptée à notre problème de localisation engendrait un véritable modèle de système spatio-temporel tel que défini au chapitre 5. Les distributions spatiales des activités sont représentées par des vecteurs indices du temps tandis que les interactions sont mises sous la forme de matrice de structure spatiales. Les coefficients sont des polynômes de décalage.

La formulation spatio-temporelle reste consistante avec ses origines statiques car elle conserve la structure du système d'équations solutionnant le problème de bénéfice de localisation. De même, l'état d'équilibre (optimum statique du problème) obtenu à l'aide de multiplicateurs spatiaux, définit le point «attracteur» du système dynamique. Il en représente l'aboutissement ultime.

La représentation d'un processus de localisation sous la forme d'un système dynamique (équations différentielles du temps) permet d'apercevoir divers propriétés évolutives, en particulier les catastrophes et les bifurcations. Nous avons d'abord examiné les propriétés de catastrophe de

la version dynamique du modèle de localisation commerciale. Les catastrophes correspondent à des changements subits de trajectoire associés à l'apparition ou la disparition d'un point d'équilibre. Normalement, elles sont causées par des modifications, même continues, de la valeur d'un ou des paramètres. Ces caractéristiques ont suscité une littérature abondante au début de la décennie quatre-vingt.

Les possibilités de catastrophe ressortent dès que la variable d'attraction évolue avec la taille de l'activité et que cette variable est dotée d'une fonction non-linéaire. Dans le cas d'une fonction puissance, une variation de l'exposant provoque des changements discontinus des états d'équilibre, passant de l'unicité à la multiplicité de ces états.

Chaque état correspond à une organisation urbaine stable, allant de la dispersion la plus équitable jusqu'à la concentration en un seul point. Seule la fonction d'attraction permet d'orienter la dynamique vers ces structures alternatives. Une modulation paramétrique progressive occasionne des catastrophes additionnelles, des activités disparaissant ici tandis que d'autres sont amplifiées ailleurs. On ne trouve pas de telles singularités fonctionnelles associées au paramètre de transport. Cependant, en influençant la taille des aires de marches, il dicte le degré de concentration spatiale des activités. Ainsi, la combinaison paramètres de transport et d'économies d'échelle gouverne l'organisation spatiale résultant d'un processus dynamique.

Il existe aussi des véritables bifurcations dynamiques qui provoquent des changements qualitatifs de la trajectoire. L'analyse au voisinage d'un point d'équilibre d'un système dynamique représenté par des équations différentielles permet d'identifier plusieurs comportements évolutifs typiques. Le développement des activités urbaines peut, par exemple, être décrit par de tels systèmes. Selon leur structure paramétrique, ceux-ci manifestent des dynamiques inter-activités «écologiquement distinctes». La transition d'une écologie à l'autre ou d'une forme de trajectoire à l'autre trahit une bifurcation dynamique. Ces possibilités d'analyse et leur rapprochement avec les approches préconisées en écologie mathématique ont été peu étudiées outre Dendrinos et Mullaly [1985].

9.1.6 Conciliation de la structure urbaine et du système de transport

La définition du cadre conceptuel conciliant la structure urbaine et le système de transport s'inspire de la représentation schématique des relations demande/offre de transport de Florian et Gaudry [1980]. Leur schéma met en relief deux type d'équilibre, l'un instantané évoquant l'affectation du trafic selon la congestion actuelle du réseau (eq. demande - performance), l'autre représentant l'ajustement de l'offre à la performance évaluée du système et indirectement à la demande correspondance (équilibre DPO). Dans le dernier des cas, l'ajustement exige un certain temps et il n'est pas sûr qu'un équilibre stable puisse être réalisé.

Nous y rajoutons un nouveau type d'équilibre, beaucoup plus général, qui concerne l'interdépendance du système de transport (offre et demande) avec le processus de localisation des activités urbaines. Ici, les délais de réaction sont très importants et l'ajustement est rarement complet.

Ce schéma conceptuel «améliore» peut être reformulé par un cadre opérationnel i.e. une séquence de problèmes d'optimisation et de modèles tous joints les autres. Ainsi, une certaine distribution spatiale des activités devient les conditions initiales. Cela permet de générer les déplacements produits et attirés par zone à l'aide des modèles de génération. La solution du programme intégré de la demande permet de distribuer les voyageurs par trajet O-D, de les répartir par mode et de les affecter sur les réseaux modaux respectifs. Elle permet aussi de calculer une matrice de coût composé de transport utilisable dans une phase ultérieure du processus.

Une telle demande distribuée sert de base d'évaluation à des éventuels modifications des réseaux. Le modèle de conception sélectionne les liens qui seront rajoutés dans une période subéquente. L'équilibre DPO n'est donc pas systématiquement obtenu puisque le décalage de réalisation laisse mouvoir la configuration spatiale de la demande.

Un processus de relocalisation des activités est parallèlement animé. Il s'ajuste aux données d'accessibilité issues d'une ou des périodes antérieures tout en tenant compte des distributions spatiales antécédentes.

Ceci est effectuée par un système d'équations spatio-temporelles doté de termes autoregressifs (permanence des activités) et de matrices géographiques (décrivant les interactions avant des effets localisants).

9.1.7 L'application à la région de Montréal

Dans cette thèse axée sur le développement d'un cadre de modélisation permettant la reproduction des effets rétroactifs entre la structure urbaine et un système de transport, nous avons tenté une application. Le fait que les méthodes de calibration et de solution des modèles soient connus n'impliquent pas que l'application s'est faite sans difficulté.

Plusieurs éléments doivent par conséquent être examinés comme la disponibilité des données de base, l'adéquation des calibrations, le réalisme des résultats et, finalement, le contrôle qu'a l'utilisateur dans le comportement du modèle. Déjà, avant même de procéder à une application, il a fallu se résoudre à une approche simulatoire plutôt que prévisionnelle. La difficulté d'obtenir une base spatio-temporelle de données de même que la soumission de l'évolution des réseaux de transport à des politiques de transport empêche de prédire l'état futur de la structure urbaine et des réseaux. On peut cependant simuler des évolutions possibles selon des paramètres généraux pré-sélectionnés.

La région métropolitaine de Montréal a servi de terrain d'essai. Ses caractéristiques se prêtaient favorablement à une telle application. La région, étendue et peuplée, se subdivise aisément en un nombre suffisant de zones ayant chacune une population significative. Les réseaux routiers et de transport collectif restent par ailleurs très articulés tout en offrant plusieurs possibilités de développement. Notre curiosité est ravivée par les multiples enjeux qui entourent le destin de Montréal en matière de transport.

L'application à la région de Montréal a d'abord révélé maints problèmes associés à la définition et la spécification de la base de données. L'absence d'une véritable base rétrospective des données d'activité a imposé une contrainte majeure. Qu'à cela ne tienne, nous avons préconisé une approche de simulation basée sur une année de référence. Il s'agit

d'une première simplification du cadre de modélisation.

Le système spatial a été choisi en fonction de la disponibilité des données et en accord avec le niveau d'agréation des réseaux de transport. Afin de limiter l'envergure de l'application tout en satisfaisant ses objectifs, les réseaux ont été définis à un degré relativement grossier. D'ailleurs, notre modélisation s'accommodait à différentes échelles d'agréation.

Comme la plupart des analyses empiriques traitant du transport et du développement urbain, nous nous sommes butés à l'indisponibilité d'une véritable base de données spatialisées sur les activités. Plutôt que les observer sous l'angle de l'occupation du sol, nous optons de les mesurer en termes d'effectifs comme la population et l'emploi. Si la désagréation sectorielle de l'emploi était à souhait, on ne pouvait en dire autant sur la désagréation géographique qu'il a fallu pourtant raffiner à l'aide d'information complémentaire.

L'opérationnalisation du modèle général a également apporté sa part de problème. L'un des principaux concerne l'inadéquation des formes fonctionnelles des sous-modèles de demande. Pris séparément, les sous-modèles peuvent subir des transformations permettant un meilleur ajustement aux observations. Toutefois, dans leur formulation intégrée, ils sont liés les uns aux autres de sorte qu'il devient impossible d'en modifier individuellement la forme fonctionnelle.

L'autre problème majeur consiste à procéder rigoureusement à l'estimation du modèle spatio-temporel de localisation tout en s'accommodant à la chétive base d'observations des activités urbaines. Ce problème surgira, il faut bien l'admettre, dans la plupart des applications même si on dispose d'une base retrospective plus étendue que la nôtre. L'autorégression par moindres carrés ne convient que dans les séries chronologiques suffisamment longues et décomposées; il ne faut donc pas s'attendre à de telles caractéristiques des données intra-urbaines.

Notre cadre de modélisation s'est pourtant avéré un outil intéressant pour l'analyse du développement urbain de la région de Montréal en regard des réseaux de transport. Les trois simulations qui ont été effectuées ont démontré des incidences diverses selon le type de réseau privilégié. Sans surprise, l'extension du transport collectif tend à favoriser une consolidation du tissu urbain à l'intérieur du périmètre de la ville centrale et de ses proches voisines. À l'opposé, le déploiement de la grille autoroutière active le développement des zones périphériques au dépens des zones plus centrales. Dans cette simulation de l'organisation urbaine, l'activité résidentielle est apparue davantage sensible aux modifications d'accessibilité.

Les simulations ont aussi révélé la complexité des relations entre les éléments de transport et localisation. Par exemple, le fait de privilégier l'essor d'un mode n'entraîne pas automatiquement un gain de clientèle auquel on s'attendrait. L'étalement urbain, plus ou moins actif selon les scénarios mais toujours omniprésent, encourage inconditionnellement l'usage de l'automobile en dépit d'un soutien explicite au transport en commun.

Aussi, l'augmentation de la capacité globale du réseau routier ne produit pas toujours les effets escomptés, du moins à long terme. D'une part, elle incite à utiliser davantage l'auto et à prolonger les déplacements. Le niveau antérieur de congestion est alors souvent rétabli après coup. Mais la disponibilité d'un mode alternatif empêche toutefois une trop forte détérioration des conditions routières. Cependant, en stimulant les développements domiciliaires suburbains et périurbains là où le transport public n'est plus compétitif, on perd définitivement cette soupape de sûreté.

D'après nos simulations, l'expansion autoroutière ne résout pas les problèmes de congestion; elle peut même les rendre plus permanents. Cette conclusion corrobore avec ce qu'on a pu observer depuis trente ans dans les agglomérations nord-américaines.

9.2 PERSPECTIVES DE DEVELOPPEMENT

Encore aujourd'hui, la modélisation des phénomènes de transport et de localisation reste une importante préoccupation de la recherche. Les "nouveaux modèles" se font plus rares car l'accent semble surtout mis sur un raffinement des formulations déjà connues sinon de nouvelles interprétations et des améliorations des algorithmes de solution.

Nous avons identifié quatre orientations de recherche qui permettent d'étendre la portée de la modélisation proposée dans cette thèse. La première consiste à formuler des problèmes d'optimisation s'accommodant davantage avec des données disponibles. Dans la seconde, il s'agit des améliorations sensibles qui peuvent être faites aux sous-modèles afin qu'ils correspondent davantage à la réalité observée.

Une troisième orientation de recherche serait de meilleures compréhension et interprétation des phénomènes modélisés, notamment à propos de leurs dynamiques. Enfin, pour les types de programme mathématique rencontrés, divers méthodes algorithmiques récemment développées ou en développement affichent des performances prometteuses.

9.2.1 Vers des formulations plus indulgentes

Il est inutile de répéter que les observations concernant le transport et les activités urbaines ont des structures particulières quand elles ne sont carrément pas disponibles. Les modèles "utilisables" doivent se montrer moins gourmands en terme de données sans pour autant sacrifier leur contenu théorique.

Il y a pourtant des données recueillies périodiquement et facilement qui sont rarement employées parce qu'elles se prêtent pas à prime abord à la modélisation. Si elles ne suscitent souvent pas d'intérêt, directement du moins, elles peuvent pourtant mises à contribution pour l'estimation de variables plus sollicitées.

Les comptages routiers correspondent bien à cette description. Ils ont l'avantage de fournir des données de circulation régulières et de pouvoir

s'installer rapidement à un emplacement donné. Avec leur structure particulière, ces comptages sont apparemment difficiles à concilier avec d'autres données de demande et servent plutôt à des applications ponctuelles.

Ainsi, plusieurs chercheurs se sont intéressés à la reconstitution d'une matrice O-D à partir des flux observés [Van Zuylen, Willumsen, 1980; Nguyen, 1982; Fisk, 1989]. L'une des approches consiste à substituer la contrainte du coût total du modèle d'entropie maximale par des conditions basées sur la circulation observée et la connaissance des itinéraires empruntés de chacun des trajets O-D. Dans ce cas, le réseau ne subit aucune congestion. L'approche alternative est fondée sur l'existence d'un trafic d'équilibre. Il s'agit de reproduire la distribution spatiale de la plus grande entropie qui, une fois affectée sur le réseau, correspond au trafic d'équilibre observé.

Il va sans dire que ces formulations alternatives peuvent se concilier avec la modélisation proposée ici. Les principes d'optimisation, les variables et la notation sont par ailleurs fort apparentées. Selon ces méthodes, il est sans doute possible d'utiliser d'autres types de données de réseau comme les passages de lignes-écrans ou les entrées aux tourniquets des stations de métro.

Dans notre modèle de transport, le réalisme du trafic d'équilibre dépend de la finesse et de la pertinence du réseau codé. Pour avoir une représentation juste, un tel réseau doit comprendre tous les liens routiers, autoroutes, artères et collectrices, où une circulation significative peut être observée. Dans bien des applications, comme la nôtre, la codification d'un réseau complet reste difficile voire impossible.

Nous avons utilisé une méthode permettant d'affecter le trafic sur un réseau principal en considérant le réseau secondaire comme des liens auxiliaires dépourvus de limite de capacité, assigné d'un coût fixe et unique par trajet O-D. L'affectation entre les deux niveaux de réseau se faisait selon le principe des temps égaux de Wardrop comme si chaque lien synthétique pouvait se comparer aux liens réguliers.

Nous avons évoqué la possibilité de procéder à une affectation hiérarchique à deux niveaux. Dans l'un, le trafic est reparti sur le réseau selon un principe optimal pour l'utilisateur. Dans l'autre, on partage la circulation entre réseaux principal et secondaire en ayant égard pour chaque trajet O-D au coût d'équilibre de l'un et au coût fixe de l'autre. Le partage est de type logit afin de tenir compte, au sein du réseau secondaire, de la multitude d'itinéraires possibles et de notre ignorance partielle du comportement des voyageurs [Brice, 1989b]. Cette approche s'inspire de celle de Fisk [1980] dont le problème d'affectation par trafic d'équilibre contient un terme d'entropie pour introduire une variabilité des comportements.

D'autres avenues semblent possibles. Pour traiter des sous-réseaux, soit extraits localement ou grossier, Hearn [1984] envisagea la décomposition des problèmes associés selon la méthode de Benders. On peut également préférer la simplification du réseau en procédant à une aggrégation systématique des liens et des nœuds [Chan, 1976].

Il y a cependant des avenues de recherche qui méritent d'être approfondies, notamment au sujet du processus de localisation. Nous avons privilégié l'approche "Spatial Time Series" de Bennett [1979] qui est caractérisée par une grande souplesse mais aussi par une grande complexité paramétrique. En retenant le concept pratique de matrice de structure spatiale de Cliff et Ord [1973; 1980], on réduit considérablement le nombre de paramètres à estimer. Dans beaucoup de cas, l'identification et l'estimation d'une matrice appropriée peuvent poser des problèmes quoiqu'il en soit. Il s'agissait de reproduire les interactions spatiales actives dans le processus de localisation. Les matrices de déplacements stratifiées par motif permettaient d'estimer les structures spatiales apparentées aux interactions localisatrices.

Si dans le cas de la structure géographique, il a été possible de simplifier (ou plutôt de décomposer) la forme fonctionnelle du modèle spatio-temporel, il faudrait faire de même avec la structure de décalage. Son identification et son estimation requièrent d'abondantes observations périodiques lesquelles sont indisponibles dans un cadre intra-urbain. L'une des orientations possibles reste de celle de Batty [1978; Batty,

March, 1979] ont les modèles urbains multiplicatifs utilisent seulement l'information de la période antérieure. Il n'est pas nécessaire non plus de procéder à une estimation sur plusieurs périodes. Ce cadre recursif s'apparente fortement au modèle d'entropie conditionnelle de Kullback [1959] mais s'y distingue par le remplacement des informations a priori par des observations antérieures. Toutefois, de tels modèles ne sont formulables qu'au premier ordre et, par conséquent, ne peuvent reproduire des variations cycliques contrairement aux modèles additifs de séries temporelles [Batty, March, 1979]. Heureusement, les phénomènes de développement urbain, bien qu'ils peuvent se manifester par saccades, dessinent rarement des fluctuations cycliques.

9.2.2 Bonification des sous-modèles

Le modèle général qui a été présenté ici est bien plus un «cadre de modélisation» qu'une formulation définitive. En effet, il s'agit d'une structure qui autorise l'intégration sinon la conciliation de modèles de transport et de localisation. Par conséquent, il devrait s'accommoder à toute représentation parente des sous-modèles dérivés par les problèmes d'optimisation. Il suffit d'en déterminer le problème équivalent et d'y trouver les modalités d'intégration.

Mais attention, toutes les formulations empiriques des sous-modèles n'ont pas nécessairement un programme mathématique équivalent et, même le cas échéant, il n'est pas garanti que celui-ci puisse se concilier avec d'autres.

Dans le chapitre 3, nous avons vu que le sous-problème d'affectation sur le réseau paraissait relativement interchangeable. Toujours en minimisant le coût des voyageurs et en substituant les termes du problème, on peut passer tour à tour d'une répartition "tout-ou-rien", logistique, par trafic d'équilibre optimal pour l'utilisateur et pour le système. Des combinaisons demeurent même envisageables comme l'introduction de l'entropie dans un trafic d'équilibre [Fisk, 1980]. Quant à la représentation de la relation coût/trafic (dans le cas où elle est nécessaire), diverses formes fonctionnelles monotones restent possibles.

A propos du partage modal, les alternatives semblent plus restreintes en dépit de l'existence d'une variété de fonctions de répartition éprouvées. D'abord, certaines n'ont pas leurs équivalences en problème d'optimisation. Ensuite, l'intégration avec les autres phases de la demande de transport suppose maintes restrictions. La plus grande latitude se situe au niveau de la composition de la «fonction d'utilité» en autant qu'elle reste intrinsèquement linéaire. Une somme linéaire des attributs de trajet, d'origine, de destination et de mode paraît acceptable car les termes équivalents dans le programme mathématique sont linéaires. Les variables exogènes peuvent subir une transformation tandis que le coût modal de transport, facteur endogène en présence d'un trafic d'équilibre, ne peut être transformé. En effet, il n'y a aucune fonction de distribution connue permettant de justifier une transformation d'un coût d'équilibre dans le problème de demande.

Cette obligation nous est apparue très limitative dans notre application car elle implique que chaque unité de coût a un effet identique pour toute paire O-D. De plus, diverses fonctions standards de partage sont prohibées car plusieurs imposent des transformations des variables [Gaudry, 1984]. Restent des formulations comme le modèle «dogit» [Gaudry, Dagenais, 1979] dont on sait qu'elles peuvent être dérivées par un programme non-linéaire (voir sous-chapitre 3.5) et dont les transformations restent facultatives.

A l'égard de la distribution spatiale, il semble y avoir peu de possibilités de raffinement. Les formulations non-gravitaires sont rares et celles qui existent, comme les «opportunités intervenantes», s'assimilent difficilement à un principe d'optimisation. La distribution spatiale telle que présentée ici peut toutefois être bonifiée en pratiquant une différenciation zonale du paramètre de transport. En autorisant des valeurs propres à des catégories d'origines ou de destinations (ex: centre-ville, quartiers centraux, banlieues et zones péri-urbaines), on reconnaîtrait la spécificité des comportements des voyageurs et de la structure spatiale.

La phase de génération spatiale procède en dehors du problème d'optimisation de la demande. Elle joue le rôle d'une identité comptable liant la demande de transport et le processus de localisation des acti-

vités. En conséquence, toute fonction identifiée à une relation numérique entre les données d'activités et le volume de déplacement qu'elles génèrent est acceptable. Ici, nous avons adopté le modèle de régression à cause de sa simplicité et de l'évidente linéarité entre activités et flux. Il n'y a pas de restriction sur la sophistication potentielle de la formulation de base comme la modulation du taux de mobilité selon les caractéristiques de la population.

Au contraire du modèle de demande, le modèle d'offre se prête moins bien à des raffinements. D'une part, la nature du problème d'optimisation principal reste simple. Il s'agit de concevoir un réseau permettant de minimiser les coûts de déplacements des voyageurs tout en respectant un budget d'amélioration. D'autre part, parmi les formulations possibles du problème de conception de réseau, nous avons présenté l'une des plus complètes i.e. avec une demande variable et un réseau à capacité limitée.

L'un des développements envisageables serait d'introduire plus explicitement la notion du temps dans le programme. Il peut s'agir de supputer les bénéfices et les coûts sur une longue période en les actualisant [Steenbrink, 1974].

La propriété dichotomique de la variable décisionnelle, les liens, pourrait être relâchée. De cette manière, le programme s'accommoderait de toutes sortes d'améliorations de réseau: élargissement de voies, vitesses praticables, fréquences d'autobus [Schéele, 1980], etc... Dans certains cas, le programme tendrait vers le problème d'augmentation de capacité d'Abdulaal et LeBlanc [1979].

Le système spatio-temporel modélisant le processus de localisation offre plusieurs alternatives de développement. Au chapitre 6, nous avons présenté la formulation complète comprenant, délai de réactions, interactions et autocorrélations spatiales. Toutes ces composantes se prêtent à des adaptations spécifiques. Les éléments des matrices d'interaction spatiale s'assimilent à tout type de structure géographique, fonction décroissante de la distance, contiguïté, hiérarchie. Ils peuvent inclure un terme d'attraction, soit en tant que coefficient, soit en tant qu'argument d'une fonction [Clarke, 1985].

L'autocorrélation spatiale n'a pas été exploitée dans notre application à la région de Montréal. Des problèmes d'identification de structure et de normalisation des observations surgissent dans le type d'application visé. Néanmoins, dans le cas où ces obstacles sont contournés, les matrices spatiales peuvent prendre diverses formes représentant l'interdépendance d'une activité à travers les zones.

Nous avons exploité la diagonale des matrices d'interaction que pour décrire les relations intrazonales de "marché" entre activités au même titre que les relations interzonales. D'autres types de relation spatiales comme la compétition pour le sol et les effets d'agglomération ont été omis. Leur incorporation permettrait de simuler des évolutions urbaines plus réalistes.

9.2.3 Vers de nouvelles interprétations

Les aptitudes du modèle général ont été surtout mises à profit pour décrire et simuler une réalité. Le potentiel analytique n'a été exploré que pour certaines propriétés dynamiques des sous-modèles de localisation. Nous avons examiné successivement l'existence d'un point d'équilibre dans un système à deux activités endogènes, les possibilités de catastrophe d'un modèle de localisation en regard avec des variations paramétriques et les bifurcations de trajectoire dans un processus co-evolutif.

Tout en restant dans le domaine des modèles de localisation, de nombreuses avenues de recherche s'offrent à nous. L'incidence des catastrophes et les formes urbaines conséquentes n'ont pas été analysées en regard de la fonction d'attraction choisie [Clarke, 1985], de la fonction de transport et des effets combinés de compétition et d'agglomération [Fotheringham, 1985]. Par ailleurs, seule la «catastrophe-pli», singularité du premier degré, a reçu notre attention. Pourtant, Kaashoek et Vorst [1984] révèlent l'existence d'une «catastrophe de rebroussement» ("cusp") dans le modèle de localisation commerciale.

Malgré sa complexité, le modèle général combinant demande et offre de transport et localisation d'activités mériterait des analyses de ses propriétés dynamiques. Pour l'instant, il ne pourrait faire l'objet que d'un

examen empirique. Par exemple, la dichotomie ajout/omission d'un lien de transport serait susceptible de promouvoir des organisations urbaines distinctives: l'intérêt serait d'examiner la discontinuité de l'équilibre en rapport avec le rajout considéré. Afin de dissiper le trop grand nombre de facteurs impliqués, il faudrait simplifier la formulation du modèle général comme fixer la demande et rendre le processus de localisation statique. De cette façon, on pourrait parvenir à une meilleure compréhension de l'impact de l'offre de transport.

Il y a bien sûr l'analyse de la modulation des paramètres de demande de transport et son impact sur la distribution d'équilibre. Comment, par exemple, les paramètres de répartition modale et de distribution O-D affectent-ils la forme urbaine? Nous savons qu'en effet, ils donnent lieu à des organisations différentes selon leurs valeurs. Nous ignorons cependant s'il existe des seuils au-delà desquels apparaissent des formes qualitativement distinctives.

Dans le système étudié, le plus grand facteur de discontinuité reste sans aucun doute le développement des réseaux de transport. D'abord, les liens sont des éléments entiers: leur ajout ou leur retranchement provoque nécessairement des changements brutaux de l'accessibilité. Ensuite, la grande complexité combinatoire des itinéraires, les effets de congestion et la rationalité individualiste des automobilistes rendent difficile la prévision de la circulation selon les modifications de réseaux. Certains résultats apparaissent même contre-intuitifs.

La solution du modèle de conception est elle-même discontinue. En effet, un lien appartenant au réseau optimal peut être omis si le budget de construction est augmenté ou diminué. La préselection du sous-ensemble des liens projetés influence également la configuration du réseau optimal.

Pour ces raisons, il importe de définir un cadre analytique permettant d'identifier les facteurs de discontinuité de l'organisation urbaine d'équilibre et de reconnaître la nature des relations impliquées. Jusqu'ici, nous prétendions reproduire le développement de la structure urbaine et soupçonnions des évolutions qualitatives mais il manque une théorie suffisamment élaborée permettant une interprétation claire des

phénomènes pressentis. Seule la théorie de catastrophe et de bifurcation [Hazewinkel et al., 1985] et la théorie de l'auto-organisation [Schieve, Allen, 1982] offrent un cadre d'analyse satisfaisant des propriétés structurelles et dynamiques de la forme urbaine. Ces théories méritent qu'on les étende davantage sur les systèmes de transport/localisation [Allen et al., 1980]. Mais cette tâche, on en conviendra, sera difficilement comblée étant donnée la grande complexité de ces systèmes et nos lacunes au niveau de leur compréhension.

9.2.4 Algorithmes de solution et procédures d'estimation

Les développements envisageables de notre cadre de modélisation sont certes assujettis par les méthodes disponibles de solution et d'estimation. En effet, les problèmes d'optimisation n'ont d'intérêt que s'ils sont rapidement et efficacement solubles. Aussi, les modèles descriptifs paramétrisés sont confrontés à des difficultés d'identification et d'estimation.

Dans le cas de modèle de demande de transport intégrant distribution spatiale, partage modal et affectation routière ou seulement deux des trois, il semble que l'algorithme d'Evans [1976] donne déjà une performance très appréciée [LeBlanc, Fahrangian, 1981; Boyce 1984]. Bien que le calcul de la direction de descente y soit plus exigeant que celui de l'algorithme de Frank-Wolfe [Florian et al., 1975; Florian, Nguyen, 1978], la convergence apparaît plus rapide.

Les deux algorithmes comportent un nombre égal d'étapes qui sont à peu près équivalentes. Selon notre expérience empirique, l'une d'elles mériteraient des améliorations. Il s'agit de la détermination de la longueur optimale de l'incrément qui se présente sous la forme d'un problème de minimisation unidimensionnel. Pour résoudre ce problème apparemment simple, nous avons opté pour la procédure de recherche de la «section dorée» qui reste facile à formuler et à programmer. L'inefficacité de la procédure réside dans le calcul de la fonction objective à chaque itération (de plus, à la première itération, quatre estimés sont nécessaires). Ce calcul consomme beaucoup de temps-ordinateur bien que répétitif. Pour cette raison, une procédure plus appropriée serait très appréciée.

Il y a sûrement d'autres algorithmes alternatifs pour résoudre le problème de conception de réseau. L'algorithme «branche-et-borne» que nous avons utilisé semble le moyen le plus simple de négocier avec une demande variable et des liens à capacité limitée. Sa structure arborescente s'accommode relativement bien avec l'aspect combinatoire du problème principal, sa non-convexité et de l'existence d'un sous-problème soluble par un autre algorithme.

Certains progrès ont récemment été fait concernant les problèmes de programmation à deux niveaux. Une méthode des «points extrêmes» [Bard, 1983] reste applicable lorsque le problème principal n'a comme contrainte que la non-négativité des flux et que la relation coût-traffic puisse être linéarisée [LeBlanc, Boyce, 1986]. Ceci implique que les coûts d'amélioration doivent d'insérer dans la fonction objective plutôt que dans une contrainte budgétaire. La fonction de congestion peut être linéarisée en approximant la courbe de coût par une série de lignes; chacune des "pièces" peut être interprétée comme des «unités de capacité».

L'algorithme de Bard ne convient pas toutefois à des applications de grands réseaux (plus de cent liens). LeBlanc et Boyce proposent donc de reprendre cette même structure algorithmique en remplaçant une des étapes par un programme d'optimisation soluble par la technique de Frank-Wolfe. Celle-ci peut résoudre le problème d'affectation selon le principe d'optimum pour l'utilisateur pour de très grands réseaux.

Cette modification permet de conserver le terme classique d'intégration de la fonction de congestion non-linéaire pour tous les liens. Néanmoins, cette procédure n'est pas encore définie avec une demande variable. Reste à savoir comment une telle procédure s'accommodera d'un algorithme destiné à résoudre le sous-problème de demande; selon Boyce [1986] cela dépend beaucoup de la convexité de la fonction objective du problème principal.

Concernant les procédures d'estimation des paramètres du problème de localisation, il ne semble pas y avoir des développements tangibles permettant d'éviter les obstacles que nous avons rencontrés. Les méthodes d'estimation qui ont été présentées au chapitre 5, les moindres carrés (OLS), le maximum de vraisemblance (ML), les moindres carrés à deux étapes

(TSLS) et les moindres carrés multivariés (MULS), sont bien connues mais ne conviennent qu'à une partie du problème. De plus, elles ne résolvent les difficultés associées au manque d'observations rétroactives.

Dans le type d'application rencontré, il convient de développer une méthode ou un ensemble de méthodes permettant d'estimer des paramètres temporels en se basant sur très peu de périodes. Pour la validité statistique des estimés, le grand nombre d'observations géographiques pourrait compenser. Quant à l'absence de comportement temporel observé, la connaissance d'un état anticipé, l'état d'équilibre par exemple, permettrait de déterminer une trajectoire et la structure paramétrique qui l'accompagne. A cet égard, beaucoup de recherches restent à s'effectuer pour qu'un tel schéma méthodologique puisse devenir opérationnel.

PAGINATION ERROR.

ERREUR DE PAGINATION.

TEXT COMPLETE.

LE TEXTE EST COMPLET.

NATIONAL LIBRARY OF CANADA.

BIBLIOTHEQUE NATIONALE DU CANADA.

CANADIAN THESES SERVICE.

SERVICE DES THESES CANADIENNES.

BIBLIOGRAPHIE

ABDULAAL M, LEBLANC L J [1979]

"Continuous equilibrium network design models", Transportation Research, (13B), pp 65-80.

AITKEN A C [1934]

"On least-squares and linear combinations of observations", Proceedings, Royal Society, (55), pp 42-48.

ALLEN P M [1978]

"Dynamique des centres urbains", Sciences et Techniques, (50) pp 15-19.

ALLEN P M [1982]

"Self-organization in the urban system", in Schieve, Allen (eds) Self-Organization and Dissipative Structures, University of Texas Press: Austin TX, pp 132-158.

ALLEN P M, BOON F, SANGLIER M [1980]

"A Dynamic Model of Intraurban Evolution", report II, Departement of Transportation U.S.A., Cambridge MA.

ALLEN P M, SANGLIER M [1979]

"A dynamic model of growth in a central place system", Geographical Analysis, (11), pp 256-272.

ALLEN P M, SANGLIER M [1981]

"Urban evolution, self-organization, and decisionmaking", Environment & Planning A, (13), pp 167-183.

ALONSO W [1964]

"Location and Land Use", Harvard University Press: Cambridge MA.

ALTSCHULER A A [1965]

"The City Planning Process: a Political Analysis", Cornell University Press: Ithaca NY.

ANAS A [1984]

"Discrete choice theory and the general equilibrium of employment, housing, and travel networks in a Lowry-type model of urban economy", Environment & Planning A, (16), pp 1489-1502.

ANDERSON T W [1958]

"Introduction to Multivariate Statistical Analysis", John Wiley: New York.

Association Canadienne des Hôpitaux (A.C.H.) [1981]

"Répertoire des hôpitaux canadiens 1981", Association canadienne des hôpitaux: Toronto.

AYENI B [1979]

"Concepts and Techniques in Urban Analysis", St.Martin Press: New York.

BABIN A, FLORIAN M, JAMES-LEFEBVRE L, SPIESS H [1982]

"EMME/2. An Interactive Graphic Method for Road and Transit Planning", CRT-204, Centre de recherche sur les transports, Université de Montreal.

BARCELO M, CHARBONNEAU F, HAMEL P [1989]

"Option préférable d'aménagement et étalement urbain 1978-1988, dans la région de Montréal - Examen et recommandations", Service de l'habitation et du développement urbain et Service de planification et de concertation, Ville de Montréal.

BARD J F [1983]

"An algorithm for solving the general bilevel programming problem". Mathematics of Operations Research, (8), pp 260-272.

BATTY M [1972]

"Dynamic simulation of an urban system" in Wilson (ed) Pattern & Processes in Urban & Regional Systems, Pion: London, pp 44-82.

BATTY M [1976]

"Urban Modelling: Algorithms, Calibrations, Predictions", Cambridge University Press: Cambridge.

BATTY M [1978]

"Operational urban models incorporating statics in a dynamic framework" in Karlqvist, Lundqvist, Snickars, Weibull (eds) Spatial Interaction Theory and Planning Models, North-Holland: Amsterdam, pp 227-252.

BATTY M, MARCH L [1979]

"Dynamic urban models based on information-minimising" in Bennett, Martin, Thrift (eds) towards Dynamic Analysis of Spatial Systems, Pion: London, pp 127-155.

BECKMANN M J, GOLOB T [1971]

"On the metaphysical foundation of traffic theory; entropy revisited" in Proceedings of the Fifth International Symposium on the Theory of Traffic Flow and Transportation, Elsevier: New York.

BECKMANN M J, MCGUIRE C B, WINSTEN C B [1956]

"Studies in the Economics of Transportation", Cowles Commission Monograph, Yale University Press: New Haven CT.

BEN-AKIVA M [1974]

"Structure of passenger travel demand models", Transportation Research Record, (526), pp 26-42.

BEN-AKIVA M [1977]

"Choice models with simple choice set generating processes", Department of Civil Engineering, MIT: Cambridge MA.

BEN-AKIVA M, LERMAN S R [1985]

"Discrete Choice Analysis: Theory and Application to Travel Demand", MIT Press: Cambridge MA.

BENNETT R J [1979]

"Spatial Time Series: Analysis, Forecasting & Control", Pion: London.

BENNETT R J [1981]

"Spatial and temporal analysis: Spatial time series" in Bennett, Wrigley (eds) Quantitative Geography: A British Experience, Routledge & Kegan Paul: London & Boston, pp 97-103.

BENNETT R J, CHORLEY R J [1978]

"Environmental Systems: Philosophy, Analysis & Control", Methuen: London.

BENNETT R J, HAINING R P, WILSON A G [1985]

"Spatial structure, spatial interaction, and their integration: A review of alternative models", Environment & Planning A, (17), pp 625-645.

BERKSON J [1944]

"Application of the logistic function to bio-assay", Journal of the American Statistical Association, (XII), pp 357-365.

BERLYANT A M [1970]

"Trend-surface and residual surface mapping and its application in geographical research", Soviet Geography, pp 839-849.

BERRY B J L, SCHWIND P J [1969]

"Information and entropy in migrant flows", Geographical Analysis, (1), pp 5-14.

BOX G E P, COX D R [1964]

"An analysis of transformations", Journal of the Royal Statistical Society, Series B, (26), pp 221-243.

BOX G E P, JENKINS G M [1970]

"Time Series Analysis, Forecasting & Control", Holden-Day: San Francisco CA.

BOYCE D E [1984]

"Urban transportation network-equilibrium and design models: recent achievements and future prospects", Environment & Planning A (16), pp 1445-1474.

BOYCE D E [1986]

"Integration of supply and demand models in transportation and location: problems formulation and research questions", Environment & Planning A, (18), pp 485-489.

BOYCE D E, CHON K S, LEE Y J, LIN K T, LEBLANC L J [1983]

"Implementation and computational issues for combined models of location, destination, mode, and route choice", Environment & Planning A, (15), pp 1219-1230.

BOYCE D E, FARHI A, WEISCHEDEL R [1973]

"Optimal network problem: a branch-and-bound algorithm", Environment & Planning, (5), pp 519-533.

BOYCE D E, JANSON B N [1980]

"A discrete transportation network design problem with combined trip distribution and assignment", Transportation Research, (14B) pp 147-154.

BOYCE D E, SOBERANES J [1979]

"Solutions to the optimal network design problem with shipments related to transportation cost", Transportation Research B, (13), pp 65-80.

BRAESS D. [1968]

"Über ein Paradox der Verkehrsplanung", Unternehmensforschung, (12), pp 258-268.

BRANSTON D [1976]

"Link capacity function: a review", Transportation Research, (10), pp 223-236.

BRICE S [1989a]

"Derivation of nested transport models within a mathematical programming framework", Transportation research, 23B, pp 19-28.

BRICE S [1989b]

"A traffic equilibrium assignment model adapted for an extracted coarse subnetwork", article soumis à Transportation Research B.

BRUTCHIE J F [1969]

"A general planning model", Management Science, (16), p 265.

BRUTCHIE J. F., TOAKLEY A. R., SHARPE R. [1971]

"A model for national development", Management Science, (18), pp B14-B18.

BRUYNOOGHE M. [1969]

"Un modèle intégré de distribution et d'affectation du trafic sur un réseau", Rapport de l'Institut de recherche du transport, Département de recherche opérationnelle et d'informatique: Paris.

BUREAU DU TRANSPORT METROPOLITAIN (B.T.M.) [1978]

"Prolongement de la ligne no 1 ouest du metro", Communauté Urbaine de Montréal.

BUREAU DU TRANSPORT METROPOLITAIN (B.T.M.) [1984]

"Rapport du Comité Ad Hoc sur le transport des personnes dans l'Est de la Communauté Urbaine de Montréal", version finale, Communauté Urbaine de Montréal.

BUREAU DU TRANSPORT METROPOLITAIN (B.T.M.) [1983]

"Rapport du Comité Ad Hoc sur le transport des personnes dans l'Est de la Communauté Urbaine de Montréal", Communauté Urbaine de Montréal.

BUSSIERE Y, MARCOUX R, TESSIER M [1988]

"Analyse prospective de la demande de transport des personnes dans la région métropolitaine de Montréal, 1981-1996", Etudes et documents no.56, INRS-Urbanisation: Montréal.

C.O.T.R.E.M. [1980]

"Recodage de l'Enquête O-D, CTCUM -1978 - système 200 zones", par J P Primeau, COTREM: Montréal.

C.O.T.R.E.M. [1981]

"Le metro regional - rapport d'étape sur la planification", COTREM: Montréal.

CANCELA DA FONCA J. P., DRACH A. [1985]

"Un modèle spatio-dynamique du paysage" in V. Berdoulay, M. Phipps (eds) Paysage et système, Editions de l'Université d'Ottawa: Ottawa, pp 103-110.

CARROTHERS G A P [1956]

"An historical review of the gravity and potential concepts of human interaction", Journal of the American Institute of Planners, (22), pp 94-102.

CHAN Y [1976]

"A method to simplify network representation in transportation planning", Transportation Research, (10), pp 179-191.

CHUNG J H, ACHOUR D, LAPOINTE A [1981]

"Economie Urbaine", Gaëtan Morin: Chicoutimi QC.

CLARKE M [1981]

"A note on the stability of equilibrium solutions of production-constrained spatial interaction models", Environment & Planning A, (13), pp 601-605.

CLARKE M [1985]

"The role of attractiveness functions in the determination of equilibrium solutions to production-constrained spatial interaction models", Environment & Planning A, (17), pp 175-183.

CLIFF A D, HAGGETT P, ORD J K, BASSETT K A, DAVIES R B [1975]

"Elements of Spatial Structure: a Quantitative Approach", Cambridge University Press: Cambridge UK.

CLIFF A D, ORD J K [1970]

"Spatial autocorrelation: a review of existing and new measures with applications", Economic Geography, (46), pp 269-292.

CLIFF A D, ORD J K [1973]

"Spatial Autocorrelation", Pion: London.

CLIFF A D, ORD J K [1975]

"Model building and the analysis of spatial pattern in human geography", Journal of the Royal Statistical Society A, (37), 297-384.

CLIFF A D, ORD J K [1980]

"Spatial Processes: Models and Applications", Pion: London.

COCHRANE D, ORCUTT G H [1949]

"Application of least squares regression to relationships containing auto-correlated error terms", Journal of the American Statistical Association, (44), pp 32-61.

COCHRANE R A [1975]

"A possible economic basis for the gravity model", Journal of Transport Economics and Policy, (9), pp 34-49.

COELHO J D, WILLIAMS H C W L, WILSON A G [1978]

"Entropy maximising submodels within overall mathematical programming frameworks: a correction", Geographical Analysis, (10), pp 195-201.

COELHO J D, WILSON A G [1976]

"The optimum location and size of shopping centres", Regional Studies, (10), pp 413-421.

COELHO J D, WILSON A G [1977]

"An equivalence theorem to integrate entropy maximising models within overall mathematical programming frameworks", Geographical Analysis, (9) pp 160-172.

COMMISSION DE TRANSPORT DE LA COMMUNAUTE URBAINE DE MONTREAL

(C.T.C.U.M.) [1983]

"Mobilité des personnes dans la région de Montréal", Commission de Transport de la Communauté Urbaine de Montréal, Service de l'Aménagement du Réseau.

COMMUNAUTE URBAINE DE MONTREAL (C.U.M.) [1970]

"Le transport public... un bond en avant". Service de la Planification, C.U.M.: Montréal.

COMMUNAUTE URBAINE DE MONTREAL (C.U.M.) [1974]

"Le transport public: les prolongements". Comité d'orientation générale du métro, Service de la Planification, C.U.M.: Montreal.

COMMUNAUTE URBAINE DE MONTREAL (C.U.M.) [1989]

"Un plan de transport en fonction des besoins réels de la population". Mémoire de la Communauté Urbaine de Montréal présenté à la Commission de l'aménagement et des équipements: Montréal.

CRECINE J P [1968]

"A Dynamic Model of Urban Structures". Rand Corporation: Santa Monica CA.

CURRY L [1972]

"A spatial analysis of gravity flows", Regional Studies, (6), pp 131-147.

CZAMANSKI S [1965]

"A method of forecasting metropolitan growth by means of distributed lag activity", Journal of Regional Science, (6), pp 35-49.

DAGANZO C F [1980]

"Multinomial Probit: the Theory and its Application to Demand Forecasting", Academic Press: New York.

DAGANZO C F, SHEFFI Y [1977]

"On stochastic models of traffic assignment", Transportation Science, (11), pp 253-274.

DALY A [1987]

"Estimating «tree» logit models", Transportation Research, (21B), pp 251-267.

DANTZIG G B [1960]

"On the shortest route through a network", Management Science, (6), pp 187-190.

DANTZIG G B [1963]

"Linear Programming and Extensions", Princeton University Press: Princeton NJ.

DENDRINOS D S, MULLALY H [1985]

"Urban Evolution: Studies in the Mathematical Ecology of Cities", Oxford University Press: Oxford UK.

DIAL R B [1971]

"A probabilistic multipath assignment model which obviates path enumeration", Transportation Research, (8), pp 85-96.

DICKEY J W, NAJAFI F T [1973]

"Regional land use schemes generated by TOPAZ", Regional Studies, (7), pp 373-386.

DIONNE R, FLORIAN M [1979]

"Exact and approximate algorithms for optimal network design", Networks, (9), 37-59.

DOMENCICH T, McFADDEN D [1975]

"Urban Travel Demand - A Behavioural Analysis", North-Holland: Amsterdam.

DOWNES J D, MORRELL D [1981]

"Variation of travel time budgets and trip rates in Reading", Transportation Research, (15A), pp 47-53.

DRAPER N R, SMITH H [1966]

"Applied Regression Analysis", Wiley: New York.

DURBIN J, WATSON G S [1950]

"Testing for serial correlation in least squares regression", I, Biometrika, (37), pp 409-428.

ERLANDER S [1977]

"Accessibility, entropy and the distribution and assignment of traffic", Transportation Research, (11), pp 149-153.

ERLANDER S, NGUYEN S, STEWART N F [1979]

"On the calibration of the combined distribution-assignment model", Transportation Research, (13B), pp 259-267.

ERLANDER S, STEWART N F [1978]

"Interactivity, accessibility, and cost in trip distribution", Transportation Research, (12), pp 237-240.

EVANS S P [1973]

"A relationship between the gravity model for trip distribution and the transportation problem in linear programming", Transportation Research, (7), pp 39-61.

EVANS S P [1976]

"Derivation and analysis of some models for combining trip distribution and assignment", Transportation Research, (10), pp 37-57.

FISK C S [1980]

"Some developments in equilibrium traffic assignment", Transportation Research, (14B), pp 243-255.

FISK C S [1989]

"Trip matrix estimation from link traffic counts: the congested network case", Transportation Research, (23B), pp 331-336.

FLORIAN M [1976]

"Traffic Equilibrium Methods". Lecture Notes in Economics and Mathematical Systems, (118), Springer Verlag: Berlin.

FLORIAN M [1977]

"A traffic equilibrium model of travel by car and public transit modes", Transportation Science, (11), pp 166-179.

FLORIAN M [1981]

"Utility, entropy and a «paradox» of traffic flow", Transportation Research A, (15), pp 327-330.

FLORIAN M [1984]

"An introduction to network models used in transportation planning", in M. Florian (ed) Transportation Planning Models, North-Holland: Amsterdam, pp 137-152.

FLORIAN M, FOX B [1976]

"On the probabilistic origin of Dial's multipath traffic assignment models", Transportation Research, (10), pp 339-341.

FLORIAN M, GAUDRY M J I [1980]

"A conceptual framework for the supply side in transportation systems", Transportation Research B, (14), pp 1-8.

FLORIAN M, NGUYEN S [1976]

"An application and validation of equilibrium trip assignment methods". Transportation Science, (10), pp 374-389.

FLORIAN M, NGUYEN S [1978]

"A combined trip distribution modal split and trip assignment model". Transportation Research, (12), pp 241-246.

FLORIAN M, NGUYEN S, FERLAND J [1975]

"On the combined distribution-assignment of traffic", Transportation Science, (9), pp 43-53.

FOTHERINGHAM A S [1985]

"Spatial competition and agglomeration in urban modelling". Environment & Planning A, (17), pp 213-230.

FRANK M, WOLFE P [1956]

"An algorithm for quadratic programming", Naval Research Logistics Quarterly, (3), pp 95-110.

GALLO G, PALLOTINO S [1984]

"Shortest path methods in transportation models", in M. Florian (ed) Transportation Planning Models, North-Holland: Amsterdam, pp 227-256.

GARIN R A [1966]

"A matrix formulation of the Lowry Model for intra-metropolitan activity location", Journal of the American Institute of Planners, (32), pp 361-364.

GAUDRY M J I [1975]

"An aggregate time-series analysis of the urban transit demand: the Montreal case", Transportation Research, (9), pp 249-258.

GAUDRY M J I [1980]

"Logit and logit models of travel mode choice in Montreal", Canadian Journal of Economics/ Revue canadienne d'Economique, (8), pp 268-279.

GAUDRY M J I [1984]

"Functional form and stochastic specification: current applications to aggregate time-series and cross-sectional travel demand models", in Florian (ed) Transportation Planning Models, Elsevier: Amsterdam, pp 61-78.

GAUDRY M J I, DAGENAIS M G [1979]

"The Logit Model", Transportation Research, (13B), pp 105-112.

GAUDRY M J I, WILLS M J [1978]

"Estimating the functional form of travel demand models", Transportation Research, (12), pp 257-289.

GAUDRY M J I, WILLS M J [1979]

"Testing the Logit model with aggregate time-series and cross-sectional travel data", Transportation Research, (13B), pp 155-166.

GOLDEN B L [1976]

"Shortest path algorithms: a comparison", Operations Research, (24), pp 1164-1168.

GOLDNER W [1968]

"Projective Land Use Model (PLUM): a Model for the Spatial Allocation of Activities and Land Uses in a Metropolitan Region", TR-219, Bay Area Transportation Study Commission: Berkeley CA.

GOLDNER W [1971]

"The Lowry model heritage", Journal of the American Institute of Planners, (37), pp 100-110.

GOULD P [1972]

"Pedagogic review", Annals of the Association of American Geographers, (62), pp 689-700.

GRANGER C W J [1969]

"Spatial data and time series analysis" in Scott (ed) London Papers of Regional Science vol. 1: Studies in Regional Science, Pion: London, pp 1-24.

GRIFFITH D A [1978]

"The Impact of Configuration and Spatial Autocorrelation on the Specification and Interpretation of Geographical Models", Ph.D. Thesis, University of Toronto.

GUMBEL E J [1958]

"Statistics of Extremes", Columbia University Press: New York.

HAGERSTRAND T. [1968]

"Innovation Diffusion as a Spatial Process", University of Chicago Press: Chicago.

HAGGETT P [1972]

"Contagious processes in a planar graph: an epidemiological application", in N. D. McGlashan (ed) Medical Geography, Methuen: London.

HAINING R F [1978]

"The moving average model for spatial interaction", Transactions of the Institute of British Geographers, (NS 3), pp 202-225.

HANSEN W G [1959]

"How accessibility shapes land use", Journal of the American Institute of Planners, (25), pp 73-76.

HARRIS B [1970]

"Change and Equilibrium in the Urban System", Institute for Environmental Studies, University of Pennsylvania: Philadelphia.

HARRIS B, CHOUKROUN J M, WILSON A G [1982]

"Economies of scale and the existence of supply-side equilibria in a production-constrained spatial interaction model", Environment & Planning A, (14), pp 823-837.

HARRIS B, WILSON A G [1978]

"Equilibrium values and dynamics of attractiveness in production-constrained spatial interaction models", Environment & Planning A, (10), pp 371-388.

HAZEWINKEL M, JURKOWICH R, PAELINCK J H P [1985]

"Bifurcation Analysis: Principles, Applications and Synthesis", D. Reidel: Dordrecht Neth.

HEARNS D [1984]

"Practical and theoretical aspects of aggregation problems in transportation planning models" in Florian (ed) Transportation Planning Models, Elsevier: Amsterdam.

HERBERT D J, STEVENS B H [1960]

"A model for the distribution of residential activity in urban areas", Journal of Regional Science, (2), pp 21-36.

HILL D M [1965]

"A growth allocation model for the Boston region", Journal of the American Institute of Planners, (31), pp 111-120.

HIRSCH M W, SMALE S [1974]

"Differential Equations, Dynamical Systems and Linear Algebra", Academic Press: New York.

HITCHCOCK F L [1941]

"Distribution of a product from several sources to numerous localities", Journal of Mathematical Physics, (20), pp 224-230.

HOANG H H [1973]

"A computational approach to the selection of an optimal network", Management Science, (19), pp 488-498.

HOPKINS L [1979]

"Quadratic versus linear models for land-use plan design", Environment & Planning A, (11), pp 291-298.

HORDJIK L, NIJKAMP P [1977]

"Dynamic models of spatial autocorrelation", Environment & Planning A, (9), pp 505-519.

HORDJIK L, NIJKAMP P [1978]

"Estimation of spatio-temporal models: new direction via distributed lags and Markov scheme" in Karlqvist, Lundqvist, Snickars, Weibull (eds) Spatial Interaction Theory and Planning Models, North-Holland: Amsterdam, pp 177-199.

HOTELLING H. [1938]

"The general welfare in relation to taxation and of railway utility rates", Econometrica, (6), pp 242-269.

HUFF D L [1964]

"Defining and estimating a trading area", Journal of Marketing, (28), pp 4-38.

HUTCHINSON B G [1974]

"Principles of Urban Transport Systems Planning", Scribner Books, McGraw-Hill: New York.

INGRAM G K, KAIN J F, GINN J R [1972]

"The Detroit Prototype of the NBER Urban Simulation Model", Columbia University Press: New York.

ISARD W, LIOSSATOS P [1979]

"Spatial Dynamics and Optimal Space-Time Development", North-Holland: Amsterdam.

JAYNES E T [1957]

"Information Theory and Statistical Mechanics" Physical Review, (106), pp 620-630.

JENKINS G M, WATTS D G [1968]

"Spectral Analysis and its Applications", Holden-Day: San Francisco.

JOHNSTON J [1972]

"Econometric Methods", 2nd edition, McGraw-Hill: New York.

KAASHOEK J F, VORST A C F [1984]

"The cusp catastrophe in the urban retail model". Environment & Planning A, (16), pp 851-862.

KANAFANI A. [1983]

"Transportation Demand Analysis", McGraw-Hill: New York.

KOOPMANS T C, BECKMANN M J [1957]

"Assignment problem and the location of economic activities", Econometrica, (25), pp 53-76.

KOYCK L M [1954]

"Distributed Lags and Investment Analysis", North-Holland: Amsterdam.

KUHN H W [1982]

"Complementary problems: an eclectic view", in T. Larsson, P. Smeds (eds) Nordic Symposium on Linear Complementarity Problems and Related Areas, Linköping Institute of Technology: Linköping Sweden, pp 1-16.

KUHN H W, TUCKER A W [1951]

"Nonlinear programming" in Neyman (ed) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press: Berkeley CA.

KULLBACK S [1959]

"Information Theory and Statistics", John Wiley: New York.

LACAVA G J, WELLAR B S [1970]

"Linear Programs as Planning Models: Analysis of the Journey-to-Work in a Planning Context", Technical Report, Institute for Social and Environmental Studies, University of Kansas: Lawrence KS.

LAKSHMANAN T R, HANSEN W G [1965]

"A retail market potential model", Journal of the American Institute of Planners, (31), pp 134-143.

LAMONDE P, LEDENT J, POLESE M [1984]

"Perspective d'emplois et de population dans la région métropolitaine de Montréal, par zone d'analyse - Horizon 1996", étude effectuée pour le Service de planification du territoire de la C.U.M., INRS-Urbanisation: Montréal.

LAMONDE P, BUSSIERE Y, BRICE S, MORIN M [1989]

"Développement urbain et stratégie de transport pour Montréal - Horizon 2001", étude effectuée pour la Ville de Montréal, INRS-Urbanisation: Montréal.

LEBLANC L J [1975]

"An algorithm for the discrete network design problem", Transportation Science, (9), pp 183-199.

LEBLANC L J, BOYCE D E [1986]

"A bilevel programming algorithm for exact solution of the network design problem with user-optimal flows, Transportation Research, (20B), pp 259-265.

LEBLANC L J, FARHANGIAN K [1981]

"Efficient algorithms for solving elastic demand traffic assignment problems and mode split-assignment problems", Transportation Science, (15), pp 306-317.

LEBLANC L J, MORLOK E K [1976]

"An analysis and comparison of behavioural assumptions in traffic assignment" in Florian (ed) "Traffic Equilibrium Methods", Springer-Verlag: Heidelberg.

LEONARDI G [1978]

"Optimal facility location by accessibility maximisation", Environment & Planning A, (10), pp 1287-1305.

LEONTIEF W, STROUT A [1963]

"Multi-regional input-output analysis" in Barna (ed) "Structural interdependence and economic development", MacMillan: London.

LEVINSON H S [1976]

"Urban travel characteristics" in Baerweld (ed) "Transportation and Traffic Engineering", Institute of Transportation Engineers, Prentice-Hall, Englewood Cliffs NJ.

LOWRY I S [1964]

"A Model of Metropolis", RM-4035-RC, RAND Corporation: Santa Monica CA.

LOS M [1979]

"Combined residential-location and transportation models", Environment & Planning A, (11), pp 1241-1265.

LUCE R D [1959]

"Individual Choice Behaviour", John Wiley: New York.

MACKETT R L, MOUNTCASTLE G D [1977]

"Developments of the Lowry models" in A. G. Wilson, P. H. Rees, C. M. Leigh (eds) Models of Cities and Regions, John Wiley: Chichester UK.

MACINNON R D [1976]

"Optimization Models of Transportation Network Improvement: Review and Future Prospects", Working Paper, IIASA: Laxenburg, Austria.

MAGNANTI T L, WONG R T [1984]

"Network design and transportation planning: models and algorithms", Transportation Science, (18), pp 1-55.

MANHEIM M. L. [1980]

"Understanding «supply» in transportation systems", Transportation Research, (14A), pp 119-135.

MANSKI C F [1977]

"Structure of random utility models", Theory and Decision, (8), pp 229-254.

MARCHAND C [1981a]

"L'estimation des fluctuations économiques dans un contexte multi-régional", La Revue Canadienne des Sciences Régionales, (4), pp 187-206.

MARCHAND C [1981b]

"Maximum entropy spectra and the spatial and temporal dimensions of economic fluctuations in an urban systems", Geographical Analysis, (13), pp 95-116.

MARTIN R L, OEPPEN J E [1975]

"The identification of regional forecasting models using space-time correlation functions", Transactions of the Institute of British Geographers, (66) pp 95-118.

MASSEY D B [1971]

"The basic : service categorisation in planning", WP-63. Centre for Environmental Studies: London.

MAY R M [1981]

"Theoretical Ecology: Principles and Applications", 2nd edition, Sinauer Ass.: Sunderland MA.

MAY R M, OSTER G F [1976]

"Bifurcations and dynamic complexity in simple ecological models", American Naturalist, (110), p 573-599.

MAYBERRY J P [1970]

"Structural requirements for abstract-mode models of passenger transport" in Quandt (ed) "The Demand for Travel: Theory and Measurement", Heath Lexington Books: Lexington MA, pp 103-125.

McFADDEN D [1973]

"Conditional logit analysis and qualitative choice behavior", in Zarembka (ed) Frontiers in Econometrics, Academic Press: New York, pp 105-142.

McFADDEN D [1978]

"Modelling the choice of residential location", in Karlqvist, Lundqvist, Snickars, Weibull (eds) Spatial Interaction Theory and Planning Models, North Holland: Amsterdam, pp 75-96.

MILLER E J, CUBUKGIL A [1981]

"Land Use Trends and Transportation Demand Forecasting: And Empirical and Theoretical Investigation", Research Report no.75, University of Toronto/York University Joint Program in Transportation: Toronto.

MINISTERE DES TRANSPORTS DU QUEBEC [1982]

"Le transport en commun, un choix régional", Ministère des transports: Québec.

MINISTERE DES TRANSPORTS DU QUEBEC [1988]

"Le transport dans la région de Montréal - Plan d'action 1988-1998", Ministère des transports: Québec.

MITCHELL R B, RAPKIN C [1954]

"Urban Traffic: A Function of Land Use", Columbia University Press: New York.

MORAN P A P [1950]

"Notes on continuous stochastic phenomena", Biometrika, (37), pp 17-23.

MURATA Y [1977]

"Mathematics for Stability and Optimization of Economic Systems", Academic Press: New York.

MURCHLAND J D [1966]

"Some remarks on the gravity model of traffic distribution, and an equivalent maximization formulation", Technical Report LSE-TNT-30, London School of Economics: London.

MURCHLAND J D [1970]

"Braess Paradox of traffic flow", Transportation Research, (4), pp 391-394.

NEUBERGER H L I [1971]

"User benefit in the evaluation of transport and land use plans" Journal of Transport Economics and Policy, (5), pp 52-75.

NGUYEN S [1974]

"Une approche unifiée des méthodes d'équilibre pour l'affectation du trafic", Thèse de Ph.D., Département d'Informatique, Université de Montréal.

NGUYEN S [1982]

"Modèles de distribution spatiale tenant compte des itinéraires", Centre de recherche sur les transports, Université de Montréal, Publication no. 255.

NICOLIS G, PRIGOGINE I [1977]

"Self-Organization in Non-Equilibrium Systems", John Wiley, New York.

NIJKAMP P [1972]

"Planning of Industrial Complexes by means of Geometric Programming", Rotterdam University Press: Rotterdam.

NORCLIFFE G B [1969]

"On the use and limitations of trend surface models", Canadian Geographers, (8), pp 338-348.

O'SULLIVAN P, HOLTSCLOW G D, BARBER G [1979]

"Transport Network Planning", Croom Helm: London.

OFFICE DE PLANIFICATION ET DE DEVELOPPEMENT DU QUEBEC (O.P.D.Q.) [1987]

"Option préférable d'aménagement de Montréal - Tendances de l'évolution urbaine, projets de programmes gouvernementaux, 1981-1987", Direction régionale de Montréal, O.P.D.Q.: Montréal.

OLLSON G [1965]

"Distance and human interaction: a review and bibliography", Bibliography Series no.2, Regional Science Research Institute: Philadelphia.

OLLSON G [1970]

"Explanation, prediction and meaning variance: an assessment of distance interaction models", Economic Geography, (46), pp 223-233.

ORD J K [1975]

"Estimation methods for models of spatial interaction", Journal of the American Statistical Association, (70), pp 120-126.

OU F-L, YU J C [1982]

"Effect of urban character on transferability of travel demand models" in "Advances in Trip Generation and Quantitative Methods", Transportation Research Record 874, National Research Council: Washington, pp 5-12.

PHILIPS L, BLOMME R [1981]

"Analyse chronologique", 2^eéd., Cabay-Economica: Louvain Belgique.

PHIPPS M. [1987]

"Ontogénie des patrons paysagiques", article présenté au colloque sur «L'ontogénie des formes spatiales», 55^e Congrès de l'ACFAS, Université d'Ottawa.

POSTON T, WILSON A G [1977]

"Facility size vs distance travelled: urban services and the fold catastrophe", Environment & Planning A, (9), pp 681-686.

POTTS R B, OLIVER R M [1972]

"Flows in Transportation Networks", Academic Press: New York.

PRENDERGAST L S, WILLIAMS R D [1981]

"Individual travel time budget", Transportation Research, (15A), pp 39-46.

PRIGOGINE I, ALLEN P M [1982]

"The challenge of complexity", in Schieve, Allen (eds) Self-Organization and Dissipative Structures, University of Texas Press: Austin TX, pp 3-39.

PRIGOGINE I, ALLEN P M, HERMAN R [1977]

"Long term trends and the evolution of complexity" in Laszlo, Bierman (eds) Goals in a Global Community, (1), Pergamon Press: Oxford UK, pp 1-63.

PRIGOGINE I, STENGERS I [1979]

"La nouvelle alliance: métamorphose de la science", Gallimard: Paris.

PUTNAM S H [1975]

"Urban land use and transportation models: a state-of-the-art summary", Transportation Research, (9), pp 187-202.

PUTNAM S H [1983]

"Integrated Urban Models: Policy Analysis of Transportation and Land-Use", Pion: London.

REILLY W J [1931]

"The Law of Retail Gravitation", Knickerbocker Press: New York.

RIDLEY T M [1968]

"An investment policy to reduce the travel time in a transportation network", Transportation Research, (2), pp 409-424.

ROCKAFELLAR R T [1969]

"Convex analysis", Princeton University Press: Princeton NJ.

SAMUELSON P A [1943]

"A fundamental multiplier identity", Econometrica, (11), pp 221-226.

SAMUELSON P A [1976]

"Economics", 10th edition, McGraw-Hill: New York.

SCHEELE S [1980]

"A supply model for public transit services", Transportation Research, (14B), pp 133-146.

SCHIEVE W C, ALLEN P M [1982]

"Self-Organization and Dissipative Structures: Applications in the Physical and Social Sciences", University of Texas Press: Austin TX.

SCOTT A J [1969]

"The optimal network problem: some computational procedures", Transportation Research, (3), pp 201-210.

SENIOR M L, WILLIAMS H C W L [1977]

"Model based transport policy assessment 1: the use of alternative forecasting models", Traffic Engineering and Control, (18), pp 402-406.

SENIOR M L, WILSON A G [1974]

"Disaggregated residential location models: some tests and further theoretical developments". in Cripps (ed), Space-time Concepts in Urban and Regional Science, Pion: London.

SHANNON C E [1948]

"A mathematical theory of communication". Bell System Technical Journal, (27), pp 379-423 & 623-656.

SHEFFI Y [1979]

"Estimating choice probabilities among nested alternatives". Transportation Research, (13B), pp 189-205.

SHEFFI Y, DAGANZO C F [1978]

"Another «paradox» of traffic flow". Transportation Research, (12), pp 43-46.

SHEPPARD E S [1976]

"Interaction Feedback Modelling: Explorations into Configurations Flows, and the Dynamics of Spatial Systems". Ph.D. Thesis, University of Toronto.

SHEPPARD E S [1978]

"Theoretical underpinnings of the gravity hypothesis". Geographical Analysis, (10), pp 386-402.

SHEPPARD E S [1979a]

"Geographic Potentials" Annals of the Association of American Geographers, (69), pp 438-447.

SHEPPARD E S [1979b]

"Spatial interaction and geographic theory". in Gale, Olson (eds) Philosophy in Geography, Reidel: Dordrecht Neth, pp 361-378.

SIMMONS D M [1975]

"Nonlinear Programming for Operation Research", Prentice-Hall: Englewood Cliffs NJ.

SMITH M J [1979]

"Existence, uniqueness and stability of traffic equilibria". Transportation Research, (13B), pp 295-304.

SPIEGEL M R [1981]

"Probabilites et statistiques: cours et problemes". traduit par R. Jacoud. McGraw-Hill: Montreal et Paris.

SPIESS H [1987]

"A maximum likelihood model for estimating origin-destination matrices". Transportation Research, (21B), pp 395-412.

STAIRS S [1967]

"Selecting an optimal traffic network". Journal of Transport Economics and Policy, (2) pp 218-231.

STEENBRINK P A [1974]

"Optimization of Transport Networks", John Wiley: New York.

STEWART J Q [1941]

"An inverse distance variation for certain social influences". Science, (93), pp 89-90.

STEWART N F [1980]

"Equilibrium versus system optimal flow: Some examples". Transportation Research, (14A), pp 81-84.

STOPHER P R, MEYBURG A H [1975]

"Travel Demand Modeling", D.C.Heath: Lexington MA.

STOPHER P R, MEYBURG A H [1976]

"Behavioural travel-demand models" in Stopher, Meyburg (eds) Behavioural Travel-Demand Models, D C Heath: Lexington MA, pp 1-53.

TAHA H A [1976]

"Operations Research: an Introduction", MacMillan: New York.

THEIL H [1969]

"A multinomial extension of the linear logit model", International Economic Review, (10), pp 251-259.

THUM R [1972]

"Stabilité structurelle et morphogenèse", Ediscience: Paris.

TINKLER K J [1974]

"On summing power series expansions of accessibility matrices by the inverse method", Geographical Analysis, (6), pp 175-178.

TOMLIN J A [1971]

"A mathematical programming model for the combined distribution -assignment of traffic", Transportation Science, (5), pp 122-140.

VAN DE PAMME C [1975]

"Methods for Linear and Quadratic Programming", North-Holland: Amsterdam.

VAN LIEROP W F J [1986]

"Spatial Interaction Modelling and Residential Mobility", Gower: Aldershot UK.

VAN ZUYLEN J H. WILLUMSEN L G [1980]

"The most likely trip matrix estimated from traffic counts", Transportation Research, (14B), pp 281-293.

WARDROP J C [1952]

"Some theoretical aspects of road traffic research", Proceedings, Institute of Civil Engineering, Part II, (1), pp 352-378.

WEBBER M J [1977]

"Pedagogy again: what is entropy?", Annals of the Association of American Geographers (67), pp 254-266.

WEBBER M J [1984]

"Explanation, prediction and planning: the Lowry model", Pion: London.

WELLAR B S, LACAVA G J [1974]

"Parametrization of the gamma probability function for examining and modeling space use", in Yeates (ed) "Proceedings of the 1972 Meeting of the 16th Commission on Quantitative Geography", McGill-Queen's University Press: Montreal & London Ont.

WHITING P D, HILLIER J A [1960]

"A method for finding the shortest route through a road network", Operational Research Quarterly, (11), pp 27-40.

WHITTLE P [1954]

"On stationary processes in the plane", Biometrika, (41), pp 434-449.

WILLIAMS H C W L [1976]

"Travel demand models, duality relations and user benefit measures", Journal of Regional Science, (16), pp 147-166.

WILLIAMS H C W L [1977]

"On the formation of travel demand models and economic evaluation measures of user benefit", Environment & Planning A, (9) pp 285-344.

WILLIAMS H C W L, SENIOR M L [1978]

"Accessibility, spatial interaction and the spatial benefit analysis of land use - transportation plans", in Karlqvist, Lundqvist, Snickars, Weibull (eds) Spatial Interaction Theory and Planning Models, North Holland: Amsterdam, pp 253-287.

WILSON A G [1967]

"A statistical theory of spatial distribution models", Transportation Research, (1), pp 253-269.

WILSON A G [1969]

"Location models", Journal of Regional Science, (9), pp 377-385.

WILSON A G [1970]

"Entropy in Urban and Regional Modelling", Pion: London.

WILSON A G [1971a]

"A family of spatial interaction models, and associated developments",
Environment & Planning, (3), pp 1-32.

WILSON A G [1971b]

"Generalising the Lowry model", in Wilson (ed) Urban and Regional Planning,
Pion: London.

WILSON A G [1974]

"Urban and Regional Models in Geography and Planning", John Wiley: London.

WILSON A G [1976]

"Retailers' profits and consumers' welfare in a spatial interaction model",
in I. Masser (ed) Theory and Practice in Regional Science, Pion: London, pp
42-59.

WILSON A G [1978]

"Towards models of the evolution and genesis of urban structure" in Ben-
nett, Martin, Thrift (eds) Towards Dynamic Analysis of Spatial Systems,
Pion: London, pp 79-90.

WILSON A G [1981]

"Catastrophe Theory and Bifurcation: Applications to Urban and Regional
Systems", Croom Helm: London.

WILSON A G, COELHO J D, MacGILL S M, WILLIAMS H C W L [1981]

"Optimization in Locational and Transport Analysis", John Wiley & Sons:
Chichester UK.

WONG R T [1984]

"Introduction and recent advances in network design: models and algorithms"
in Florian (ed) Transportation Planning Models, North Holland: Amsterdam,
pp 187-225.

WOODTON H J, PICK G W [1967]

"A model for trip generated by household", Journal of Transport Economics and Policy, (1), pp 137-153.

ZANGWILL W I [1969]

"Nonlinear Programming: A Unified Approach", Prentice-Hall: Englewood Cliffs NJ.

ZEEMAN E C [1977]

"Catastrophe Theory", Addison-Wesley: Reading MA.

ZELLNER A [1962]

"An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias", Journal of the American Statistical Association, (57), pp 348-368.