A STUDY OF THE RELATIONSHIP BETWEEN FOREST DISTRIBUTION AND ENVIRONMENTAL VARIABLES USING INFORMATION THEORY - A Regional Scale Model for Predicting Forest Response to Global Warming

by

WEI JIANG

B.Sc.(PU) Geography; M.Sc.(PU) Geography

Submitted to the School of Graduate Studies
and Research in Partial Fulfilment of
the Requirements of the Degree of
Doctor of Philosophy

Department of Geography
University of Ottawa
Ottawa, Ontario

ISBN 0-612-15633-8

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGMENTS

I wish to express my sincere thanks to my supervisor, Dr. Michel Phipps, for his suggestion of this subject and consistent guidance through each phase of this research. His valuable advice, insightful thoughts, and careful correction of the thesis drafts have greatly contributed in shaping this thesis. This thesis would never come into being without his thoughtful guidance and patient help.

I would also thank Dr. D. Lagarec and Dr. K. Gajewski in Department of Geography, University of Ottawa for their valuable suggestions, guidance and important comments. In the early and the middle stages of the research, Dr. Lagarec provided many valuable suggestions, particularly for the analysis of the impact of spatial scale on the forest distribution entropy. Dr. Gajewski not only provided important comment and suggestion in every stage of this research, but also corrected the draft of this thesis both in academic aspects and in English grammar.

Great appreciation should also be given to those people who provided data for this research. Mr. S.L.Gray at the Petawawa National Forestry Institute provided the forest data. Mr. T. Webb at the Canadian Climate Centre, Atmospheric Environment Centre (Downsview, Ontario, Canada) provided the CCC GCM climatic data. O.J. (Phil) Gallien at the Centre for Land and Biological Resources

Research, National Soil Data Base Unit, Agriculture Canada provided the soil and soil texture data. Without their help, thesis would not be able to finish.

Wei Jiang

CONTENTS

LIST OF FIGURES, MAPS, AND TABLES

ABSTRACT

Many studies on forest or vegetation response to global warming have been done using the gap model or empirical models. Thus far, there is no good regional model allowing to predict forest change at an intermediate scale. In this study, we have developed a model of this type, called Knowledge Base Forest Model (KBFM), using an information analytical tool (PEGASE)based on information theory.

Using this model and data from the Canadian Climatic Centre genral circulation model, we could predict the future distribution of forest types in the research area : the Province of Manitoba. The study shows that the KBFM may well be used to predict the future regional distribution of forest types. Its main advantages are : (1) environmental variables used as predictors can be qualitatives (e.g. soil texture) as well as quantitative (e.g. temperature); (2) the KBFM provides the possibility to account for the role of soil factors in the forest response to global warming; (3) the KBFM can predict forest type distribution using various climatic scenarios; (4) the KBFM can predict forest type distribution with greater details than empirical models.

RÉSUMÉ

De nombreuses études sur la réponse de la forêt et de la végétation au réchauffement global ont été faites à l'aide du modèle dit de la « trouée » et de modèles empiriques. Jusqu'à présent, il n'existait pas de bon modèle régional permettant d'explorer les réactions de la forêt aux changements climatiques à échelle moyenne. Dans cette étude, nous avons développé un modèle de ce genre, nommé Système expert forestier (SEF), à partir d'une méthode d'analyse de l'information (PEGASE) fondée sur la théorie de l'information.

En utilisant ce modèle et les données des projections climatiques fournies par le Modèle de circulation générale (MCG) du Centre Canadien de Climatologie, on a pu prédire la distribution future des types forestiers dans la zone d'étude : la Province du Manitoba. L'étude montre que le SEF peut être très bien utilisé à échelle régionale. Ses principaux avantages sont : (1) les variables environnementales utilisées pour prédire la nouvelle distribution peuvent être aussi bien qualitatives (ex.: texture du sol) que quantitatives (ex.: température); (2) le SEF tient compte du rôle des facteurs édaphiques dans la réponse de la forêt au réchauffement global; (3) le SEF peut prédire la distribution future des types forestiers à partir de divers scénarios climatiques; (4) le SEF peut faire des prédictions plus précises et détaillées que les modèles empiriques.

Chapter 1. Introduction

The earth's energy balance is dominated by atmospheric heat trapping; about 70% of the incoming solar energy is absorbed by the earth. This absorbed energy is re-emitted at infrared wavelengths by the atmosphere and the earth's surface. Eighty five percent of the radiation emitted by the surface is trapped by water vapour and greenhouse gases and returned to the earth. This phenomenon is called the greenhouse effect. The result of this greenhouse effect is to induce an average temperature of the earth at the surface of $+15^{o}C$ rather than the $-18^{o}C$ that it would have without the heat trapping. This greenhouse effect is essential for the existence of life on earth. Through this process of radiative absorption, greenhouse gases -- of which carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and chlorofluorocarbons (CFCs), together with water vapour (H_2O) and ozone (O_3) are the most important -- play a major role in determining the earth's climate (*Rotmans 1990*).

Since the industrial revolution, mankind has caused a considerable increase in the atmospheric content of these gases. The primary cause of the increased concentrations is the large-scale and world wide use of fossil fuels to generate heat, power, and electricity for the growing world population and the rise of life quality. Changes in land use, especially agriculture, also contribute to the

concentration increase by accelerating the decomposition of soil organic matter of large areas of land. For more than a century the use of fossil fuels has been on the increase and this trend is still continuing. As a result, the atmospheric concentration of CO_2 has increased by about 25% since the beginning of this century. The concentration of some other greenhouse gases has increased by even larger factors (*Rotmans 1990*).

Although some scientists doubt the reality of global warming, most believe that increasing concentration of greenhouse gases will raise the earth's temperature. "*A survey of models reported in IPCC(1990) and reviewed in IPCC(1992) indicated with high confidence that an equilibrium climate change due to a doubling of CO_2 or its radiative equivalent due to all the greenhouse gases will warm the lower atmosphere (troposphere) and cool the stratosphere and that global average tropospheric warming will range between 1.5 and 4.5^{0}C ...*"(*Mendelsohn et al. 1994*). The controversy about global warming is mainly about the magnitude and the time frame of the temperature increase (*Rotmans 1990*). The future temperature curve will depend on fossil fuel use, the adoption of alternative energy sources and the rate of deforestation. Increasing concentrations of the greenhouse gases and rising temperature may trigger both positive and negative feedback mechanisms, such as the increased uptake of CO_2 by plants

and the oceans, the increased release of CO_2 by organic matter in soils, or increased CH_4 release from methane hydrates and methane locked up in peat-bogs and arctic permafrost. In spite of these uncertainties, the atmospheric CO_2 concentration is expected to have doubled by the middle of the next century (*Rotmans, 1990*).

Today, our prediction of future temperatures mainly comes from a class of model called General Circulation Models (GCMs). "*The aim of GCMs is the calculation of the full three-dimensional character of the climate. The solution of a series of equations (Table 1-1) that describe the movement of energy and momentum and the conservation of mass and water vapor is therefore required. Generally the equations are solved to give the wind field (i.e. mass movement) at next timestep. ... The first step in obtaining a solution is to specify the atmospheric conditions at a number of grid points obtained by dividing the Earth's surface into a series of ... grids. Conditions are specified at each grid point for the surface and several layers in the atmosphere. The equations in Table 1-1 are then solved at each grid point using numerical techniques.*" (*Dickinson 1983*).

Table 1-1. Fundamental equations solved in GCMs

1. Conservation of energy (the first law of thermodynamics)
   i.e. Input energy=increase in internal energy plus work done
2. Conservation of momentum (Newton's second law of motion)
   i.e. Force = mass * acceleration
3. Conservation of mass (the continuity equation)
   i.e. The sum of the gradients of the product of density and windspeed in the three orthogonal directions is zero
4. Ideal gas law (an approximation to the equation of state)
   i.e. Pressure * value = gas * absolute temperature

Source: A history of and Introduction to Climate Models (Dickinson 1983)

GCMs have limitations at the present time. For example, none of the GCMs in current use can produce acceptably accurate representations of even current climatic conditions. All GCMs have basically the same design; however, they gave different predictions of global mean climate change and regional distributions of change. The most important reason, perhaps, is that some climatic feedbacks are not well understood and simulated in GCMs (*Mendelsohn 1994*). Moreover, the predictive accuracy of GCMs is closely related to the grid size of GCMs, which is limited by the power of computer. In spite of the limitations, GCMs have the potential to approach the real atmospheric situation very closely (*Dickinson 1983*). Most GCMs predict that the global mean temperature will increase by 1.9 to 5.2^{0}C as the atmospheric concentration of CO_2 double in the next 100 years. *Table 1-2* gives the predictions of several GCMs (*Cohen 1991*). If we accept the validity of GCM projections, it means that global warming could be significant and happen at very rapid rate

(*Hansen et al. 1988*).

Facing such a potential and dramatic climate change, scientists started to investigate the potential impacts of this global warming. The question of how vegetation and forest communities will respond to global warming has also attracted many ecologists' and geographers' attention.

Table 1-2. The prediction of global average temperature increase after global warming*.

General Circulation Models	Global Average Temperature Increase
Oregon State University (OSU)	2.8°C
Canadian Climate Centre (CCC)	3.5°C
US National Centre for Atmospheric Research (NCAR)	4.0°C
National Climatic Data Centre(USA) NOAA's Geophysical Fluid Dynamics Laboratory (GFDL),	4.0°C
NASA(USA) Goddard Institute for Space Studies (GISS)	4.2°C
United Kingdom Meteorological Office (UKMO)	1.9-5.2°C

* **Source:** ***Application of the Canadian Climate Centre General Circulation Model Output for Regional Climate Impact Studies - Guideline for Users*** **by Canadian Climate Centre.**

Vegetation and forest communities exist in a dynamic equilibrium with the climatic conditions. Historical and paleoecological evidence shows that climate change has significant impacts on forests (*Webb 1981, Davis 1989, Davis et al. 1986, Davis et al. 1985*). For example, Webb (*1981*) found that forest species moved northward as the climate warmed and southward as climate cooled in the northern hemisphere. Davis (*1989*) did not agree with Webb's conclusion and she believed that with climate warming, intact forest ecosystems have not moved northward as a unit; instead, species have responded individualistically. Although peleoecologists have varying interpretations of forest response to climatic change in geological past, one thing is certain: global warming caused by the greenhouse effect will affect the forests in many aspects such as species extinction, forest productivity change, forest spatial distribution shift, etc. Many studies have been carried out; however, scientists do not have a very clear picture of what these impacts will be.

This thesis will focus on the impact of global warming on forest spatial distribution shifts. Understanding the impact of global warming on forest spatial distribution is important both from an academic point of view and a human application perspective. For example, an understanding of forest distribution shifts may help us to assess the potential risk of species extinction, and it can help

scientists to develop ecological strategies to minimize the impact. Forests are an important natural resource, particularly in Canada. The knowledge of the impacts of global warming on forest resources may help government to plan long-term strategies for more ecologically sound land use practices. It can also help the forestry industry to assess potential economic impact.

Chapter 2. Literature Review

Research focusing on vegetation response to climate change has followed two main approaches: (1) paleoecological research, and (2) modelling. Although paleoecological research does not relate directly to vegetation response to future global warming, it provides valuable information about the response to climate changes of the past. This constitutes a useful basis for understanding possible forest change.

Modelling is the most frequently used approach for predicting the potential forest response to global warming. "*The advantage of simulation models is that they permit experimentation and observations over long periods of time. An investigator can manipulate conditions ... and examine responses over decades or centuries.*" (*Brubaker 1986*). So far much research has been carried out to predict vegetation, especially forest community, response to global warming. We may conveniently classify the models that have been used into two types: gap models and empirical models. Using both kinds of models, studies have been conducted for two different objectives: a) predict the composition of the forest by studying the influence of climatic factors on forest succession at a local-scale using gap models; or b) predict the shift of vegetation range at a global or continental scale with empirical models.

2.1 The gap Models

The first version of gap model, called "*JABOWA*", was developed by Botkin and his co-workers (1972). Since this time other versions were developed, such as "*FORET*" (*Shugart et al. 1977, 1979*), and "*FORENA*" (*Shugart 1984*). These gap models are "*mixed-species*" and "*mixed-age*" stochastic stand simulators, implemented in FORTRAN and designed to mimic processes controlling tree establishment, growth, replacement and mortality within the forest (*Solomon 1986*). Because of the complexity of the models, simulations are carried out at a very small scale (*1/12 ha*). Competition among tree species, succession of the forest community, and changes in species composition as a function of time and climate are mimicked successfully by the models. For example, *Shugart et al. (1977)* simulated the dynamics of 33 tree species typical of southern Appalachian deciduous forest very successfully.

Solomon et al. (1981) were the first to relate forest dynamics and historic climatic change using this model. They claimed that they found a strong similarity between the vegetation history and the simulation outcomes produced by the *FORET* model. Their research showed the validity of their model in simulating long-term forest dynamics. *Solomon (1986)* used the *FORENA* model to simulate the temporal response of forests to CO_2-induced climate change in

eastern North America. Using a local scale model, he explored forest distribution shifts, changes of forest productivity and carbon storage change at a sub-continental scale by simulating forest at 21 locations, including tundra, north and south boreal forests, north, west and east deciduous forests. Solomon's research showed dramatic vegetation shifts as a result of climatic change. *Pastor et al. (1988)* included the variables of soil water and soil nitrogen availability in their gap model. They found that "*the **heterogeneity of landscape,** particularly the **distribution of various soils,** becomes an important factor determining forest responses to climatic change*" (*Pastor et al. 1988*). However, this study did not explore how the soil spatial distribution impact the behavior of forest response to global warming. *Overpeck et al. (1990)* considered that forest disturbances, such as forest fire, tornado, hurricane, etc., may be very important factors influencing the behavior of forest response to climatic change. They included in their model the probability of catastrophic disturbance causing the death of all trees on a given plot in their model. They found that "*climate-induced increase in disturbance could significantly alter the total biomass and compositional response of forest to future warming, and that an increase in disturbance frequency is also likely to increase the rate at which natural vegetation responds to climatic change*" (*Overpeck et al. 1990*). They also suggested that "*many other factors, like soil dynamics, seed*

sources, pollution, pests, pathogens, forest management and the accumulation of fire fuels must be considered in predicting forest respond to climate change" (*Overpeck et al. 1990*). *Bowes et al.*(1993) used the *FORENA* model to study the forest responses to climate change at a regional scale (in Missouri, Iowa, Nebraska, and Kansas) by choosing three representative simulation locations in Missouri. The three selected locations represent the three major land resource areas of that region that have significant forest cover. This research was similar to Solomon's but at a smaller scale. It included a variety of predictions: prediction of forest species composition change, prediction of forest spatial distribution change, change of the forest productivity, and possible change of future forestry industry in that area.

Basically, a gap model is a local scale model. Some researchers also used it at a continental and regional scales, such as the above mentioned Solomon's and Bowes's research.

As a tool to investigate forest response to CO_2-induced climate change, the gap model has been widely used to study the following questions:

(1) The impact of global warming on species composition of forest community;

(2) The impact of global warming on forest productivity;

(3) The impact of global warming on forest biomass;

(4) To suggest factors that may be important in influencing forest response to climate change;

Although this model cannot directly simulate the change of a vegetation boundary, it can investigate this question by extrapolation from several simulations. The primary advantage of gap models is their ability to predict transient states.

2.2. Limitations of gap models.

Gap models make a number of contributions to our understanding of forest response to climate change. However, they have some limitations.

First, these models were originally developed to simulate forest succession in a small area (*1/12 ha*). Dale *et al.* (*1994*) pointed out that "*such a gap size may not be adequate for the effective regeneration of some species*". Also, Urban *et al.* (*1993*) recognize the artificiality of having an upper temperature limit on species occurrence. These models are not sensitive to small errors in

climate parameter estimation (*Botkin et al. 1992*).

Another limitation of these models is that "*the growth-climate relationships of species are defined as parabolas with their end points at the present-day limits of the species in growing degree days*" (*Hanson et al. 1989*). Davis *et al.* (*1985*) noted that "*the growth equation assumes that the current vegetation is in a steady state with the average climate; that the climate is optimal at the midpoint and least favourable at the extremes; and the relationship will not change as the climate changes*". In addition, "*the parameterization is based on the realized niche of each species. It does not consider that the range of a species might be limited by competition with others, or that species might be able to expand their ranges in the absence of other species*" (*Botkin 1985*).

Thirdly, when simulating forest response to global warming at a regional or continental scale using gap models, some variables, such as soil properties, are not usually taken into account; however, they are very important in shaping the forest distribution at regional scales.

2.3 Empirical models

Another approach for predicting vegetation responses to climate

change is based on models using statistical relationships between vegetation and climate. *Hanson et al. (1989)* call this approach a "*transfer function*" model; other people call it an "*empirical model* "*(Gates 1990)*. This approach is based on the existing vegetation-climate relationships taking into account vegetation or forest types distributions and selected climate variables. The future vegetation is then projected onto a map based on these relationships but using new climate scenarios from GCMs. This class of models can deal with an area much larger than those of gap models - global or continental scales. Moreover, non-climatic variables are normally not included. For example, soil properties are not considered to be important in influencing future vegetation distribution at these scales.

Box *(1981)* developed a quantitative global scale terrestrial vegetation model using this approach. His model consists of : (1) a world classification of important terrestrial plant growth forms (life forms); (2) a set of predictive variables representing the main climatic correlates of these forms; and (3) empirically obtained hypothetical limiting values defining an ecoclimatic envelope for each plant form. Box classified world vegetation into 41 plant life forms and chose 8 predictive climatic variables including mean temperature of the warmest month, mean temperature of the coldest month, annual range of monthly mean temperatures,

average annual precipitation, annual moisture index, highest average monthly precipitation, lowest average monthly precipitation and average precipitation of the warmest month. The model relating plant forms to environmental conditions is based on empirically derived environmental envelopes composed of an upper and a lower tolerance limit for each life form with respect to each climate variable. This model was tested with 65 sites for which climatic and detailed vegetation date were available. Box found the predictions for the 65 sites were very successful. These results suggest that empirical models are able to predict the general features of world vegetation distributions and vegetation structure. The results also suggest that the basic features of plant form and vegetation structure are determined primarily by the general levels and mean seasonal patterns of temperature and climatic water balance. Box suggested this approach could be used for predicting the vegetation after climate change.

Emanuel et al. (1985) used the lifezone classification of Holdridge *(1947)* to explore global vegetation change following global warming. This system is a scheme for relating the characters of natural vegetation to climate indices. Three climatic indices are used in the Holdridge lifezone classification: average annual precipitation; mean annual biotemperature and potential evapotranspiration ratio. Two Holdridge lifezone maps were made:

one was derived from a worldwide network of approximately 8,000 meteorological stations; the other was based on the predicted future climate from the GISS GCM. Emanuel *et al.* predicted a large decrease in the area occupied by boreal forest (by 37%) and tundra (by 32%) and an expansion of grassland under a warmer climate. In this modelling exercise, precipitation was kept unchanged at current levels after global warming; therefore, the problems of this prediction are obvious since this assumption might not be true.

In a similar studies, Henderson-Sellers (*1991*) generalized the vegetation into desert, tundra, grass and shrub, woods and rainforest. Two scenarios of future climate from two versions of GCMs were used for estimating the Holdridge lifezone map after CO_2 doubling. She found that the highly generalized vegetation grouping shows relatively little sensitivity to the climate changes produced by doubling CO_2, which is in contradiction to Emanuel *et al's* (*1985*) predictions. The differences, however, were caused by Emanuel *et al's* failure to correctly compute the biotemperatures resulting from doubling CO_2 (*Henderson-Sellers 1991*).

Rizzo *et al.* (*1992*) used an empirical model to assess the sensitivity of Canada's ecosystems to climate change. Unlike the previous studies, they worked at a continental instead of a global

scale. The ecosystem classification system they used is the Ecoclimatic Regions of Canada produced by the Canada Committee On Ecological Land Classification (*Ecoregions Working Group 1989*). Eight ecoclimatic provinces - Arctic, Subarctic, Dry continental boreal, Moist continental boreal, Maritime boreal, Cool temperate, Moderate temperate, and Grassland were used in their model. The climate variables were selected in two steps. First, a group of experts identified the climate variables which were thought to be important in controlling each of the separate ecoclimatic provinces. Secondly, only climate variables that could be generated from the GCMs were considered. Nine variables satisfied these constraints -- mean annual temperature, mean July temperature, mean winter temperature, mean summer temperature and growing season temperatures, total annual precipitation, total winter precipitation, total summer and growing season precipitations. Discriminant analysis was used to derive a series of **linear functions**, one for each ecoclimatic province, based on the current climate variables. Climatic data corresponding to a doubled CO_2 produced by the GISS GCM were introduced into the linear functions to derive the Ecoclimatic Provinces under these new conditions. Their model suggests that doubled CO_2 would cause major changes to Canada's ecosystems: a reduction of the Subarctic, Arctic and Boreal Ecoclimatic provinces, an increase in the size of the Cool temperate, Moderate temperate and Grassland ecoclimatic provinces,

and the formation of Transitional Grasslands and Semi-Desert ecoclimatic provinces. Rizzo *et al.* (1992) suggested, however, that the changes would not be instantaneous because of the time-lag due to soil development processes and species migration.

Prentice *et al.* (*1992*) developed a global biome model to predict global patterns in vegetation physiognomy. A biome is a combination of dominant plant functional types. Functional plant types are groups of species that germinate and grow under similar sets of environmental conditions, for example dry-deciduous sclerophyll. This model, innovated many aspects of Box's (*1981*) model, was based on a small number of plant functional types applied with an environmental sieve and dominance hierarchy. Prentice *et al. (1992)* selected mean coldest-month temperature, annual accumulated temperature over 5°C, a drought index incorporating the seasonality of precipitation and the available water capacity of the soil as the environmental limiting variables. This model predicts which plant types can occur in a given environment and selects the potentially dominant types from among them. This model is an "*attempt to translate experimental and observational findings on the climatic controls of plant distribution into testable predictions of large-scale vegetation*". (*Prentice et al.* 1992). Prentice *et al.* found their model predicted present-day biomes more accurately than the Holdridge scheme and they believed that their

model is better than the Holdridge scheme both on empirical and theoretical aspects.

2.4 Limitations of Empirical Models

Empirical models usually give two kind of predictions, the change of vegetation boundaries; and the change of the area of vegetation types. These models usually deal with global or continental scales. The vegetation predicted by empirical models becomes more general as the research scales increase.

The use of empirical models depends on three assumptions. First, the vegetation or species are now, and will remain, in equilibrium with climate. Secondly, the relationship between vegetation and climate will be the same after global warming. Thirdly, the chosen climate variables are those which are believed to critically control the vegetation distribution, and the secondary effects or interactions among species are not considered as important.

Empirical models do not take into account the rate of climate change. In these models, no matter how fast the climate change, the simulation output will always be the same as long as the final climate data is the same. Most of these models do not consider non-

climatic variables, for example, soils are usually not considered. Prentice *et al.*'s biomes model (*1992*) is an exception. Moreover, vegetation and climatic variables are simplistically linked by linear relationships.

2.5 Problems

As mentioned above, scientists investigate the question of global warming impacts on vegetation mainly using two kinds of models, gap models and the empirical models. These two classes of models deal with different scales and answer different questions (*Table 2-1*)

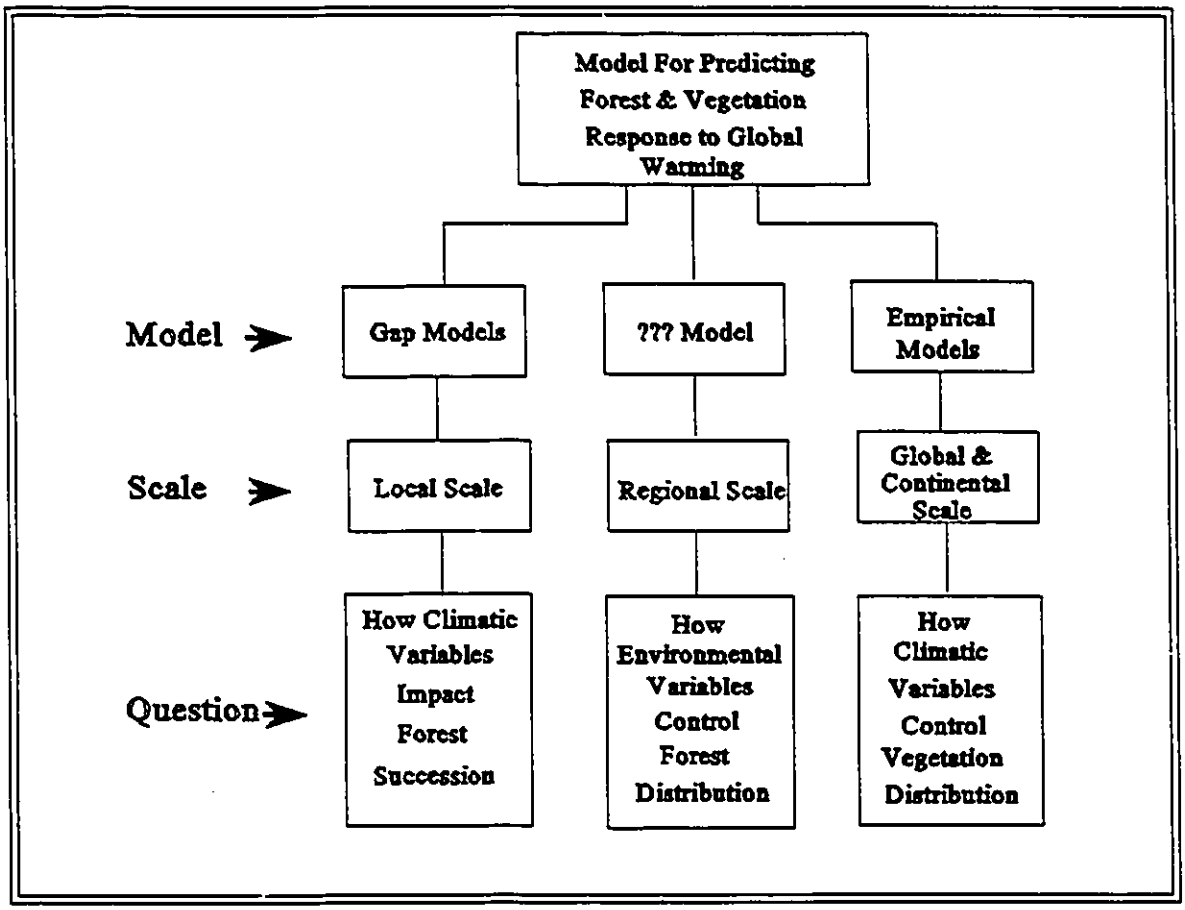
Table 2-1 The summary of the main characteristics of the present generation of gap models and empirical models

Items		Gap Models	Empirical Models
Scales	Local	Yes	No
	Regional	*	No
	Continental	No	Yes
	Global	No	Yes
Questions	Boundary change	No	Yes
	Forest Area Change	No	Yes
	Productivity Change	Yes	No
	Biomass Change	Yes	No
	Species Change	Yes	No
	Transient States	Yes	No

** Although Gap models have been used at a regional scale. It is not fundamentally a regional scale model.*

Both gap and empirical models have made significant contributions to our understanding of climate impacts on the vegetation. However, there are questions remaining (*Figure 2-1*).

Figure 2-1 Unresolved questions of climate impacts on vegetation.



The first problem is that there is an unresolved spatial scale between the local scale gap models and the global (or continental) scale empirical models. Because there are no models dealing with the regional scale, many important ecological processes (such as species migration, forest fire) and environmental factors (such as soil properties) cannot be included in both gap and empirical models. "*Comparative studies of the same phenomena at* ***different scales*** *would aid in the development of extrapolation rules*" (*Graham et al. 1990*). The study of the impact of global warming on vegetation at various scales, therefore, is important. Lenihan (*1993*) believed "*the development of spatially explicit models that predict the transient response of vegetation at regional to global scales will be a major task ...*". The development of a regional scale model will help us to understand forest response to global warming.

The second problem is that we do not have a model which enables us to investigate the potential importance of non-climatic factors (mainly soil properties) on forest response to global warming. The simulation of the impact of the interaction among climatic and non-climatic variables on forest has not been well explored. Many studies have mentioned the potential importance of the impact of soil properties on forest response to climatic change (*Bernabo 1981, Prentice 1986, Pennington 1986, Pastor et al. 1988*). However,

we do not presently have an appropriate model to investigate this question.

So far, some people (*Bowes et al. 1993*) use gap models as a substitute to study forest response to climatic change at a regional scale by integrating several simulations. This kind of research gives the compositional change of plant communities at each of several simulations, and then produces the regional forest communities change by interpolation. This kind of approach is not a really a regional scale study, because many of the regional scale properties, such as soils, are not included in the model. This model does not help us to obtain something new at a regional scale. Empirical models are also less appropriate at a regional scale. The relationships between forest distribution and climatic variables are obtained by using a spatial boundary matching methods and they are too general to be applicable with a sufficient accuracy at regional scale. Moreover, the inclusion of climatic and non-climatic variables in empirical models will make the spatial boundary matching impossible.

In summary, we need new models to study forest response to climatic change at a regional scale. In this study, we will explore one approach to this problem.

Chapter 3. Research Plan

The preceding section showed that our knowledge about the response of vegetation or forest to global warming is limited by the types of models in use. Although it is known that some of the important factors or ecological processes which need to be included in predictive models are not included, we are yet unable to do so because of the limitation of the models we have. Without the development of new types of models, our knowledge about global warming impact will not progress much, since we just repeatedly apply these two types of models in different areas. We need new models, based on different predictive techniques, so that we can explore the vegetation response to global warming at various scales, and include more environmental factors and ecological processes. The development of these new models will be an important task for the scientific community.

The development of regional scale models is particularly desirable since regional scale models have great potential for application to practical problems. Except for several large countries in the world, most national or provincial governments manage areas of regional size. For example, the provincial departments in Canada and most of the national governments in Europe could all benefit from the use of such models. For the management of natural

resources, a prediction about vegetation response to global warming will be very useful to enable appropriate planning. A regional scale model can provide such scenarios.

Regional scale studies are increasingly being attempted. For example, a multi-disciplinary project at a regional scale has been finished in the area of MINK region in United States (*Bowes 1993*). A subproject was concerned with the impact of global warming on the forest resources and the impact on forestry. A similar Canadian research project is also being carried out in the MacKenzie River Basin Area of Canada in a project led by Dr. Stuart Cohen.

Today, with the development of GIS technology, most provincial and national governments have established GIS systems which store a large number of climatic and other environmental variables. These data can be helpful in building such a regional-scale model.

This thesis will focus on the following two main aspects: (1) make full use of available data and develop a new model based on information theory to predict the forest response to global warming at a regional scale; and (2) apply this model in a specific area (province of Manitoba) and test its applicability. It will be shown that how this model enables us to explore the response of forest to global warming in a new way in Chapter 5.

3.1 Questions and Hypotheses

The research based on this regional scale model will concentrate on the following questions:

(1) How do climatic and non-climatic variables interactively control the present-day forest distribution at a regional scale?

(2) Which environmental variables (both climatic and non-climatic) are the critical factors controlling forest distribution at regional scales? A critical factor is either a climatic or a non-climatic variable which is a good predictor of forest distribution. The concept of critical factor will be discussed in detail in chapter 4.

(3) What could be the contribution of an analytical method based on information theory to construct the predictive model.

This research is carried out based on the following hypotheses:

(1) At a regional-scale, climatic variables are not the only factors which control forest spatial distribution. Other factors, like soil texture, parent materials of soil and geomorphological factors, may also be potentially important.

(2) The spatial distribution of any forest type is principally controlled by some critical variables in an interactive way. Different forest types are not necessarily sensitive to the same combination of variables.

In addition to the above hypotheses, we also assume:

(1) Vegetation is now, and will, remain in equilibrium with climate. The relationship between vegetation and environmental variables will be the same before and after global warming. This assumption is important because it would be impossible to predict future forest distribution without this assumption. This is due to the fact that the actual relationship between vegetation and environmental variables after global warming is impossible to obtain at this point.

(2) The present forest distribution is a reliable indicator of ecological conditions at a site although historical forest fires, clear-cut and other human disturbances exert an influence on forest pattern and weaken the association between forest and environmental variables.

3.2 Methodology

The methodology in this thesis includes two parts. One is an analytical method based on information theory, which is used for analyzing the relationship between forest distribution and environmental variables. The other is a knowledge-based forest model developed for predicting forest response to global warming.

3.2.1 Information Analysis and PEGASE

PEGASE (*Partition d'un Ensemble Geographique pour l'Analyse Spatiale Ecologique*) was developed by Dr. Michel Phipps (*1981 a and b*) "*as a means of determining ecological relationships among a set of overlaid choropleth maps*" (*Davis et al. 1990*). It is a method based on information theory. It has been also called "*Mutual information analysis* " by Davis *et al.* (*1990*). Since this method was developed in the late 1970's, it has been used in several studies. Phipps *et al.* (*1986*) used this method to study the ecological organization of spatial pattern of land use system pattern and showed that this method was effective. Davis et Dozier (*1990*) used this method to develop an ecological land classification in a 73 square kilometre region of northern Santa Barbara County, California. They explored a new way to divide a heterogenous region into relatively homogenous subregions based on

the analysis of the spatial correspondence between vegetation and environmental variables including geology, elevation, slope, seasonal insolation and hill slope position. They demonstrated that *PEGASE* is useful in recognizing the vegetation pattern as well as in land classification. Baudry (*1989*) studied the relationship between ecological systems and the combined effects of environment and agricultural activities. Dumanski *et al.*(*1987*) investigated how soil physical factors and economic factors interact at the farm level and how agricultural land use evolves as a consequence. These applications in a variety of situations have shown the relevance of this method for:

(1) investigating the relationship between a specific qualitative variable (which is called the dependent variable) and a set of other variables (which are called the independent variables);

(2) classifying or grouping observations according to the relation between attributes;

(3) finding the optimal combination of predictive variables to predict the dependent variable.

According to Phipps (*1981 a and b*), this method has two essential characters: (i) its multivariate nature and (ii) its divisive

classification strategy allowing the classification of sites into homogeneous groups with regard to their attributes. The multivariate nature means that this method deals with multiple variables. As mentioned previously, there are two kinds of variables: the dependent variable and the independent ones. Ideally, the variables should be qualitative ones. Since we can change quantitative variable into qualitative one by classifying quantitative variable into several groups according its value; therefore, this method can handle both qualitative and quantitative variables. The ability to deal with multiple qualitative variables is one of the main advantages of this method. The divisive classification character means that this method can finally find the relationships between the dependent and independent variables by a divisive, hierarchical processes of classification of sites. This relationship is the basis of the predictive ability of the methodology.

3.2.1.1 The PEGASE Algorithm:

The best way to present this algorithm is to understand how information theory is used to study the relationships between two variables.

Suppose a spatial grid with N cells, and a qualitative dependent

variable mapped onto this grid (*e.g. vegetation*) with m possible outcomes (or community types) and a frequency array f_j ($j = 1,..,\ m$) where f_j is the number of cells associated with the jth community type ($N,\ f_m$) define a macrostate of this system. Any particular spatial arrangement of the m outcomes mapped onto the N cells so as to satisfy the values f_j is a possible spatial configuration subject to the macrostate. The number of potential configurations is given by:

$$W(v) = \frac{N!}{\prod_{j=1}^{m} f_j!} \qquad \text{with} \quad \sum_{j=1}^{m} f_j = N \qquad (1)$$

The observed pattern is just one of these numerous arrangements. For a system with a finite number of elements, the entropy is given by Boltzmann's formula:

$$S(v) = \frac{1}{N} \ln(W(v)) \qquad (2)$$

A basic assumption underlying the entropy concept says that all configurations associated with the macrostate are assumed to be equally likely to occur. This implies that any configuration, including the observed pattern, has a probability of occurrence given by:

$$p(v) = \frac{1}{W(v)} \qquad (3)$$

Furthermore, the concept of entropy implies that the probability of any two cells bearing a particular community type is equal over the whole pattern. This spatial homogeneity assumption opposes the concept of an ecological organization according to which cells differ in such characteristics as soil, topography and mesoclimate, and provide community types with various living conditions. According to Phipps (*1981a*), "... *each community type tends to occupy particular types of sites and this is reflected by the probability of encountering this particular type at various location. This contradiction implies that the actual entropy of the system is less than the one calculated by equation (2), and the number of possible patterns is less than that in equation (1)*".

This means that the system contains a negentropy representing the ecological constraints limiting its freedom. This negentropy is provided by various independent variables, such as soil, temperature, precipitation, etc. For example, if the initial set of cells is broken down according to soil types and if these subsets have significantly distinct probabilities regarding the community types, they represent a differentiation of the pattern and an ecological organization of space. "*In terms of information theory, we may say that looking at the vegetation through the information channel of an ecological characteristic, introduces negentropy into the system, thus revealing some order in the pattern*" (*Phipps 1981a*).

Let us assume f_{ij} denotes the number of cells with the jth type of community and the ith type soil. We have:

$$\sum_{j=1}^{m} f_{ij}=f_{i}. \qquad (4)$$

and

$$\sum_{i=1}^{n} f_{ij}=f._{j} \qquad (5)$$

If we suppose that character *s* introduces a negentropy, we have to compare *W(v)* (*Equation 1*) the number of possible configurations of the system without *s* and $W_s(v)$ the number of configurations with *s*. The latter is given by:

$$W_s(v) = \prod_{i=1}^{n} \left(\frac{f_i.!}{\prod_{j=1}^{m} f_{ij}!} \right) \qquad (6)$$

The relationship between, *W(v)* and $W_s(v)$ can be expressed by the equation:

$$W(v) = W_s(v) * r \qquad \text{(where } r >= 1\text{)} \qquad (7)$$

The factor *r* represents the factor by which the number of possible configurations has been reduced by knowing character *s*. From equation (7), we have:

$$\log_e r = \log_e W(v) - \log_e W_s(v) \qquad (8)$$

From equation (1) and (6), we have:

$$\log_e r = \log_e \left(\frac{N!}{\prod_{j=1}^{m} f_{\cdot j}!}\right) - \log_e \left(\prod_{i=1}^{n} \left(\frac{f_{i\cdot}!}{\prod_{j=1}^{m} f_{ij}!}\right)\right) \qquad (9)$$

If we simplify equation 9, we get:

$$\log_e r = \log_e (N!) - \sum_{j=1}^{m} \log_e (f_{\cdot j}!) - \sum_{i=1}^{n} \log_e (f_{i\cdot}!) + \sum_{j=1}^{m} \sum_{i=1}^{n} \log_e (f_{ij}!) \qquad (10)$$

The calculation of the negentropy provided by an environmental factor (e.g. soil) will be given below.

If number *N* and *f* are large enough, equation (10) can be estimated by using Stirling's approximation:

$$\log_e r = \frac{1}{N} \left(N \log_e N + \sum_{j=1}^{m} \sum_{i=1}^{n} f_{ij} \log_e f_{ij} - \sum_{j=1}^{m} f_{\cdot j} \log_e f_{\cdot j} - \sum_{i=1}^{n} f_{i\cdot} \log_e f_{i\cdot}\right) \qquad (11)$$

This quantity measures the amount of uncertainty removed (or the negentropy introduced) by knowing the character *s*. It is a measure of the ecological constraints exerted by the feature *s* on the spatial location of the types of vegetation. It is also known as the mutual information between both variables *v* and *s*.

3.2.1.2 The PEGASE Procedure

In the previous section, we presented a series of mathematic equations used to measure the relationship between one environmental variable and the dependent variable. Now suppose we have *k* environmental variables and each may introduce a certain amount of negentropy into the system. Therefore, we want to study the relationship between one dependent variable and several environmental variables. *PEGASE* is a procedure that allows us first to calculate the entropies removed by all environmental variables separately, then to choose the variable which reduces the most entropy of the community types. This environmental variable is used to divide the set of cells into a number of subsets according to the value of this variable in each cell. The subsets resulting from this division are themselves further re-divided or re-classified into sub-subsets according to the same rule. Therefore, we have a stepwise divisive process and the entropy of the system is reduced

step by step. This divisive process will continue until one of the following conditions is satisfied:

(1) The number of cells in a subset is less than the minimum number needed to implement the information analysis by the equation 11 given above. The default number is taken as 30 in *PEGASE* and it may be reset by the user. If we use the default number, it means that the divisive process will stop if the number of cells of a subset is less than 30.

(2) The subset's entropy is smaller than a predefined minimum value. Usually we use the default value which is 0.30. In this case, it is considered that the subset is homogenous enough and that the outcome (type of community) is determined for all practical purpose. Even if we keep dividing this subset, we will not reduce the entropy of the system very much.

(3) The subset's size is larger than 30 and its entropy is larger than the minimum value (0.30) but none of the remaining variables provide significant negentropy. It means that "*a certain amount of uncertainty which cannot be removed by the known information remains*" (*Phipps 1981 a and b*).

This analysis provides us with two important results which can be

used to predict forest response to global warming. One is a divisive hierarchical tree-like diagram (see *Appendix A*), which shows a progressive reduction of the initial entropy introduced by environmental variables. This diagram is called a *tree*. The root of this tree is the initial whole set of cells. The leaves of this tree are the terminal subsets (*TSS*) resulting from the division of the initial set by a group of environmental variables. Between the root and the leaf, there are intermediate subsets. It is interesting that each terminal subset contains information or "knowledge", which tells what a combination of environmental variables interactively predicts a given community type when a set of specific values are associated. The total number of "knowledge" obtained equals the number of terminal subsets.

This "knowledge" can be used to predict community type cell by cell if we know the values of the environmental variables in each cell. Moreover, the same kind of prediction can be made even if the specific values of a cell are modified subject to climatic change. This predictive capability is inherent in *PEGASE;* however, the function allowing the prediction of vegetation change has not been developed so far. This predictive ability makes *PEGASE* a possible method to study forest response to global warming.

The other result is that *PEGASE* finds the sequence of the

environmental variables according to their contribution to the entropy reduction. Since we may interpret a variable's contribution to the entropy reduction as a potential prediction of community type by knowing the variable, the variable which reduces the most entropy (or uncertainty) means that we learn most about the dependent variable (community type) by knowing this variable. In other words, *PEGASE* can help us to identify which environmental variables are critical to the dependent variable (community type). This function is extremely valuable for constructing a forest model. When we construct this model, there are many environmental variables which can potentially influence forest distribution, and the *PEGASE* algorithm can select the most critical environmental variables from a set of candidates. We call those environmental variables which are good predictors of forest community type the **critical factors** or the **critical environmental variables**. A good predictor of forest community type is an independent variable which reduces a significant amount of entropy of forest community in the information analysis process.

Given its structure, the *PEGASE* procedure has some similarities with multiple regression. However, it differs from multiple regression model in some important aspects. For the multiple regression model, the relationships between one variable and other variables are deterministic relations and the variables usually

should be quantitative. In *PEGASE*, the relationship is set up by reducing the entropy of spatial distribution. These relationship is of a stochastic and non-linear nature. The variables can either be qualitative or quantitative in *PEGASE*, which makes it convenient to handle environmental variables such as soil types, soil texture, or surficial geological deposits. Moreover, the hierarchical divisive diagram is composed of a limited number of relevant combinations of categories of independent variables. These combinations predict the categories of the dependent variable. This means that this method does not consider all the possible combinations of the categories of independent variables. The prediction is based on relevant relationships between dependent and independent variables.

One more important aspect is that the prediction in *PEGASE* leaves room for uncertainty. This means that we usually cannot predict a specific community type with 100% probability when given a group of environmental variables' values. The prediction is always associated with a probability. For example, if a cell has a specific environmental conditions (e.g. the soil texture is sandy, the annual average temperature is 5.0°C and the annual precipitation is 520.0 mm), we have 85% probability to predict the community in this cell will be type *A* and 15% probability to be type *B*. It is rare that we have 100% probability to predict a cell's community when given a specific environmental conditions.

This is because of following reasons:

(1) The set of independent variables might not contain all the relevant information with regard to the spatial distribution of the dependent variable. The dependent variable may be influenced by many variables in the real world, and usually it is hard to include all of them in the analysis. In a study of the relationship between forest distribution and climatic and edaphic variables, for example, historical forest fires, clear-cut and other human disturbances exert an influence on forest pattern and weaken the association between forest and environmental variables;

(2) The nature of the relationship may well be seen as a non-deterministic one.

Because of these analytical and predictive abilities, *PEGASE* is an ideal method to analyze regional scale forest distributions.

3.2.2 A Knowledge-Based Forest Model for Predicting Forest Response to Global Warming at Regional Scales

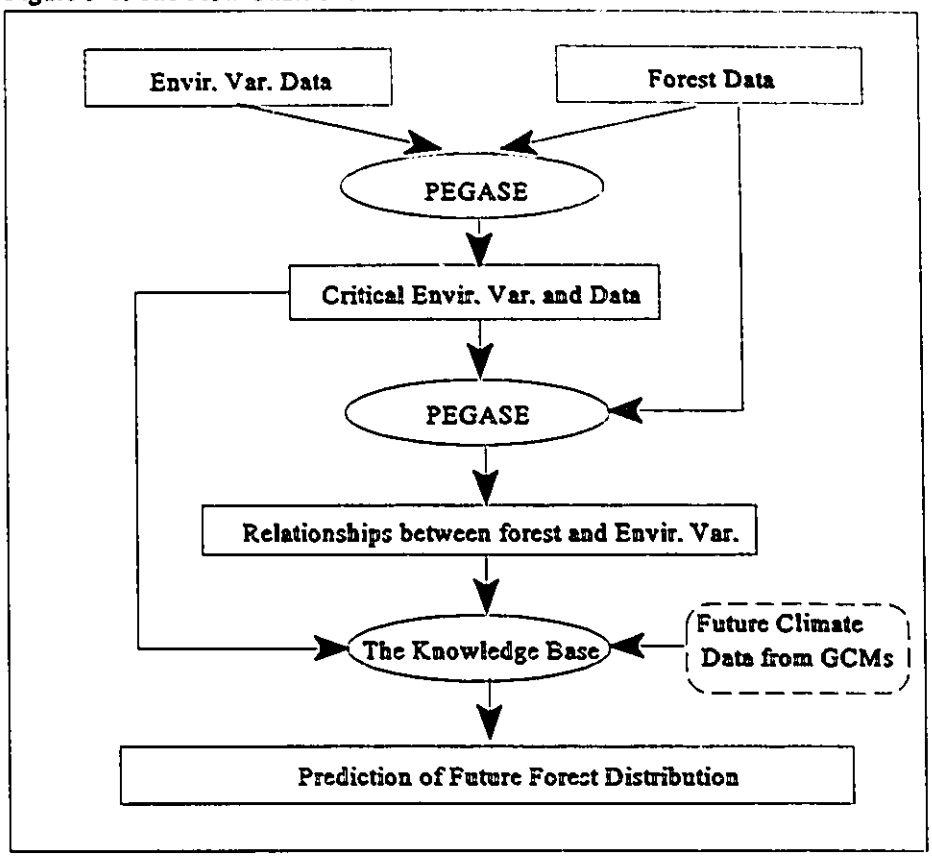
PEGASE, as mentioned above, can identify the complex relationship between environmental variables and the dependent variable

(community type). This relationship constitutes a set of "knowledge" which can be used to predict the dependent variable. After we obtain the "knowledge" between forest distribution and the environmental variables, we develop a knowledge-based forest model. For simplicity, we will refer this knowledge-based forest model as *KBFM*.

The *KBFM* was developed on a platform of IBM compatible personal computer (*80386, 80486, or higher*) under the DOS environment. The programming language used to construct this predictive model is *C/C++* and the compiler is the *BORLAND C++* version 3.1. The model is described in *Figure 3-1*.

The data base necessary to build this knowledge base will be described in detail in the next section. This data base includes data of present forest distribution and data of environmental variables related to the forest type distribution. This data base is used for running *PEGASE* in order to get the "knowledge" about the relationship between forest distribution and the environmental variables as well as predicting the future forest distribution.

Figure 3-1. The Flow Chart of the KBFM



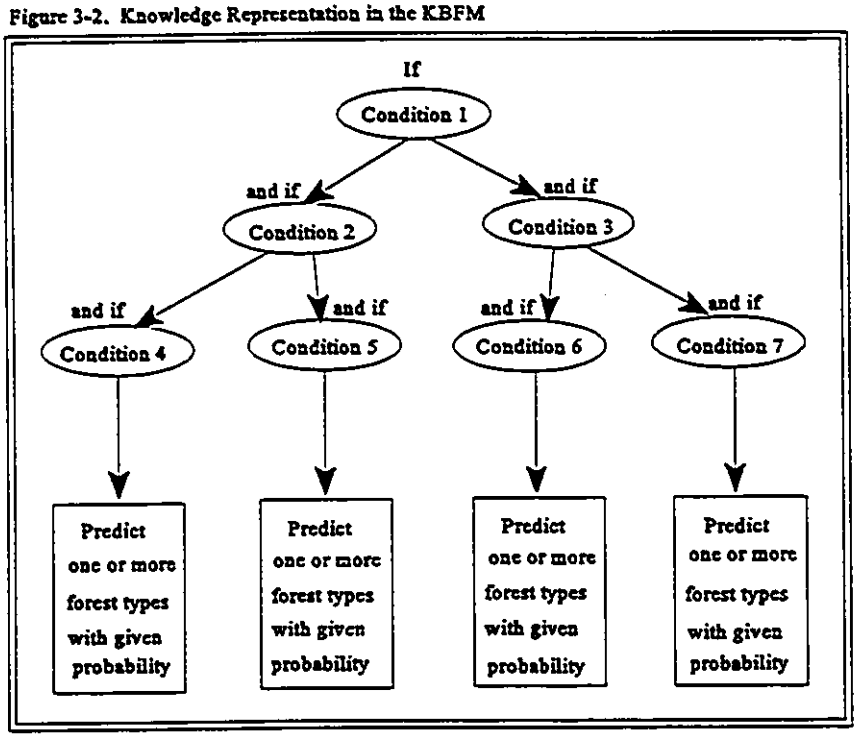
The knowledge base is the most critical part of *KBFM* since the prediction accuracy depends on the quality of this knowledge base. The concept of knowledge base comes from the expert system in computer science. A knowledge base is a subsystem of an expert system. It is "*a knowledge set (mainly expert knowledge) formally represented and stored in computer usable form*" (*Savory 1988*). One

of the basic forms of knowledge representation is the *condition-action* statement which has the form (*Kelly 1993*):

IF condition THEN actions

The *KBFM's* knowledge base has the same structure and function as expert system's knowledge base. A theoretical example of knowledge representation is indicated in *Figure 3-2*.

Figure 3-2. Knowledge Representation in the KBFM



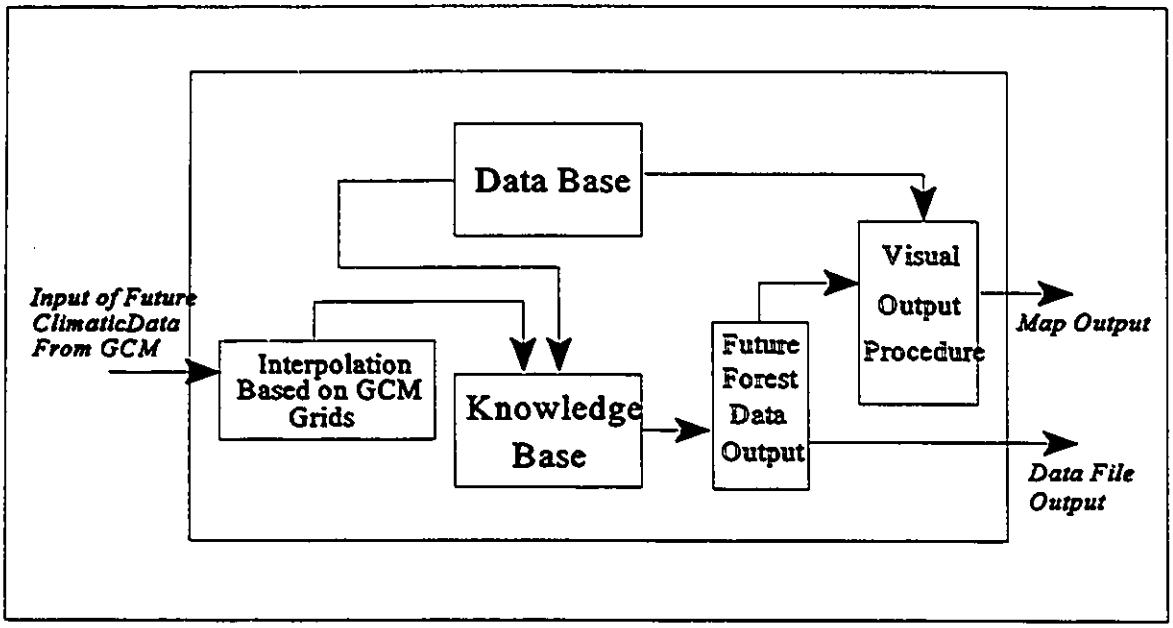
The knowledge base in *KBFM* has the same functions as the expert system's knowledge base; however, one important difference between them is that the knowledge comes from different sources. Knowledge in *KBFM*'s knowledge base come from an empirical source - the analysis of the present forest distribution and their associated environmental variables using *PEGASE*. On the contrary, knowledge in an expert system's knowledge base comes from human experts and is more subjective. This is not to say that the knowledge in an expert system is not good, but this just emphasizes the objective character of the knowledge in the *KBFM's* knowledge base.

The *GCM* prediction of future climate is an input to the *KBFM*. At this time, GCMs can only provide a global scale future climate, which makes the regional-scale studies difficult. There has been some research exploring ways to obtain regional-scale climate from *GCM* predictions (*Robock 1993*); however, there is no simple and effective solution. In this study, the global scale *GCM* predictions is transferred to the regional-scale (the research area) using interpolation. We will talk about this in chapter 5 in more detail.

The display procedures are a group of program functions which display the predictive forest distribution map and all the database information on a computer screen. These are the visual output of the *KBFM*. The visual outputs include the maps of present and the

predictive forest distribution. The overall structure of the *KBFM* is shown as *Figure 3-3*.

Figure 3-3. The structure of the KBFM

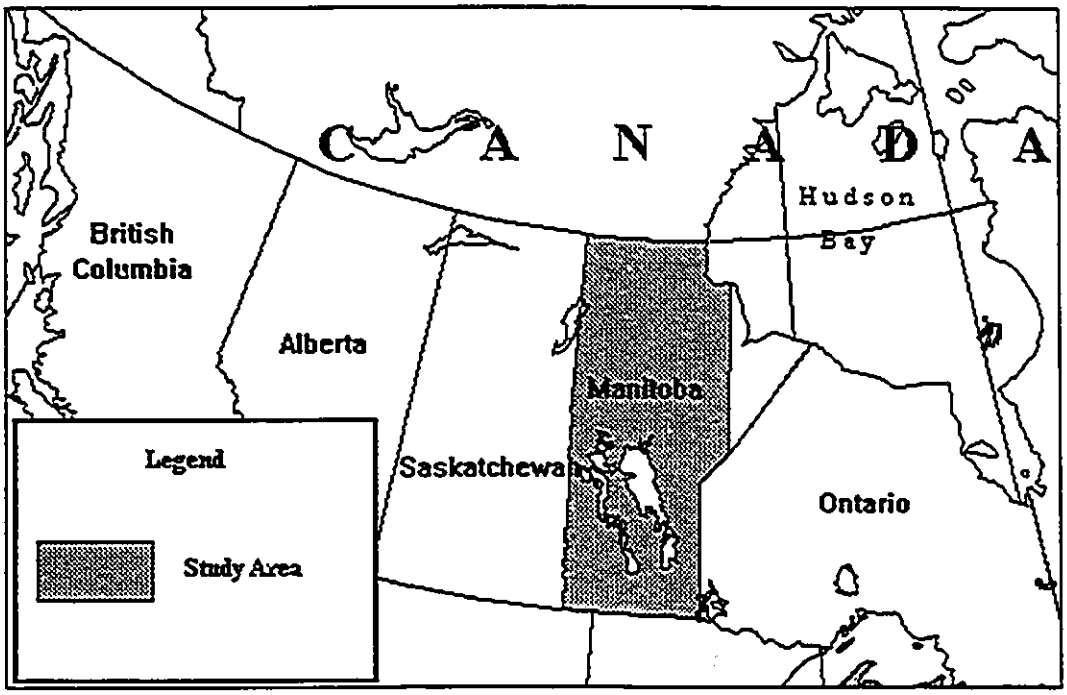


Since the construction of the *KBFM* is closely related to real data, we have only given the outline of this model here. The detailed construction of the model will be explained in Chapter 4.

3.3 Study Area

The province of Manitoba has been selected as the research area of this study (*Map 3-1*). A small part of the north-east of Manitoba was not included for two reasons. First the climate data in this area are insufficiently accurate since there are few weather stations in this area. Secondly, there are enough cells to perform adequately the analysis in terms of the methodology.

Map 3-1 The study area



The province of Manitoba was selected as the research area for two reasons: the diversity of the vegetation and the availability of a consistent data set.

Manitoba is one of the Canadian provinces where the whole range of major bioclimatic zones can be found in a relatively small area. From the south to the north, there are prairie, broad-leaf forest, boreal forest, transition zone between boreal forest and tundra, and tundra. Moreover, two forest boundaries, the southern border of the broad-leaf forest and the northern border of boreal forest, are included in this study area. Such a diversity of vegetation types is favorable to the development of the research project and the implementation of *PEGASE*.

Even though *Canadian Forest Inventory* (*Grey et al. 1989*) contains a forest data base at a national level, there are inconsistencies among the provinces. For example, the different provinces use different grid systems, the forest inventories were done in different years, and different classification systems were used. Moreover, there are inconsistencies in forest data within provinces. For example, Ontario, Quebec, and British Columbia use different grid systems within their own territories. In contrast, Manitoba is a province where a single forest data system exists over the entire territory. Other relevant data, like climatic, soils and geological data, are also available to support this study.

3.4 The Database

The data base for this research is based on the cell system of the Manitoba forest inventory. Each cell is a *Township* which has a unique spatial location code including meridian, township and range (*Table 3-1*). A normal township has an area of 92.16 km^2 and a few exceptional ones have a smaller area. All the data are stored based on cell. Each cell has a record which contains the information of forest type, climatic data, and other data. In addition, each cell also has a five byte location code which can determine the spatial location of each cell. This code is essential for mapping purpose to display the predictive forest distribution.

The data base is composed of two ASCII files. These two data files contain the same information in two formats for two purposes. The first one is a data file in which the climatic data are stored in *degrees centigrade* for *temperature* and in *millimetres* for *precipitation* (*Appendix B*). This data file records the present-day climatic data so that it can be used for calculating future climatic data. The second data file is organized in a specific format such that *PEGASE* can read it directly (*Appendix C*). In this data file, the climatic data are classified into groups. For example, the data for October temperature for a cell is TO2 (the temperature is higher than or equal to 0.00°C and lower than 0.99°C.

This data file is transformed from the first data file using a simple computer program.

3.4.1 Forest Data

Forest data of Manitoba come from two different sources. The first source is the *Canadian Forest Inventory Project* undertaken by the *Petawawa National Forest Institute* (*Gray et al. 1989*). Manitoba forest data are part of the Canadian Forest Inventory. These data had been collected by the Manitoba provincial government from 1961 to 1985 and all records were taken from a detailed, intensive forest inventory for management purposes. The corresponding information of Manitoba forest data are summarized in *Table 3-1*.

The Canadian Forest Inventory data base only covers the area from the south of the province to the northern boundary of boreal forest. The vegetation in the north of the Province is not included in the inventory. It was therefore necessary to use another source providing the vegetation of the northernmost part of the province. This source is the **"Natural Vegetation"** map in the Atlas of Manitoba (*Weir 1983*). We obtained the vegetation data of tundra and the transition zone between boreal forest and tundra from this map.

Since the forest data were collected for the management purpose,

the forest types are classified on the basis of the predominant genus. A predominant genus is the most abundant genus in a forest according to the forest coverage description. "*This is usually the first species in the cover type description of the stand in the source inventory*"(*Gray et al. 1989*). A township is not covered by one specific forest type; however, it is covered by several forest types. For each township, the forest inventory gives all the areas of different forest types classified based on predominant genus.

Table 3-1 Information relative to the Manitoba Forest Data

Province	Manitoba
Source agency	Province
Inventory	Township
Area inventoried	40565 (*10^3 ha)
Percentage of total inventoried area	6.7% of Manitoba area
Average year of inventory	1975
Inventoried between 1981-1985	30%
Inventoried between 1976-1980	19%
Inventoried between 1971-1975	20%
Inventoried between 1966-1970	25%
Inventoried between 1961-1965	7%
Cell reference system	Township(e.g. 1E02205, which means east of the 1st meridian, township 022 and range 05)

Source: "Canadian Forest Inventory 1989" (Gray et al. 1989).

There are nine types of predominant genus in Manitoba:

(1) *Spruce* (including black spruce (*Picea mariana (Mill.)BSP.*) and white spruce (*P.glauca (Moench) Voss*));
(2) *Pine* (including jack pine (*Pinus banksiana Lamb.*), red pine (*P. resinosa Ait.*), scots pine (*P. sp.*), and white pine (*P. strobus L.*));
(3) *Fir* (balsam fir (*Abies balsamea (L.) Mill.*));
(4) *Larch* (tamarack (*Larix laricina (Du Roi) K.Koch*));
(5) *Cedar* and other conifers (eastern white cedar(*Thuja occidentalis L.*));
(6) *Poplar* (including balsam poplar (*Populus balsamifera L.*), eastern cottonwood (*P. deltoides Marsh.*), largetooth aspen (*P. grandidentata Michx*), and trembling aspen (*P. tremuloides Michx.*));
(7) *Birch* (white birch (*Betula papyrifera L.*));
(8) *Maple* (black maple (*Acer nigrum Michx.*));
(9) Other broad-leaved species (including basswood (*Tilia americana L.*), black ash(*Fraxinus nigra Marsh.*), bur oak (*Quercus macrocarpa Michx.*), hackberry (*Celtis occidentalis L.*), hop-hornbeam (*Ostrya virginiana (Mill.) K.Koch*), red ash (*Fraxinus pennsylvanica Marsh*), white elm (*Ulmus americana L.*), and willow (*Salix sp.*)).

A classification of forest types based on predominant genus may not

be appropriate in forest ecological research since two species of the same genus may have different ecological requirements. In Manitoba; however, such a classification is acceptable considering the fact that no better forest data available. This is because that only spruce, pine and poplar can be the dominant genera at a township scale in this study area. If trees of other genera exist in a township, they are never dominant. There are two spruce species which are black spruce and white spruce in the study area. Black spruce and white spruce have relatively similar ecological characteristics at a regional scale. They can grow in pure stands or in mixtures with each other. Therefore, a spruce dominant forest is ecologically sensible in the study area. This is also true for the poplar species in the study area. For the pine genus, there are several pine species such as Jack pine, white pine, red pine and scots pine in the study area. Jack pine has some different ecological properties from white pine, red pine and scots pine; however, only Jack pine forms large areas forest in this region. A pine dominant forest in the study area is a Jack pine dominant forest.

As mentioned above, forest inventory gives all the areas of forest types classified based on the predominant genus within a township. We cannot directly use these data in our analysis because we need to have a specific forest type for each cell in order to run *PEGASE;* therefore, we have to generalize a forest type for each

township based on the forest inventory data. The generalization of forest types for each cell is done as follows: (1) sum up the areas of each forest type classified based on the predominant genus in a cell; (2) generalize a forest type for each cell based on the top two forest types which cover the largest areas in a cell. Seven forest types in southern Manitoba have been classified. These generalized forest types are listed as below:

(1) *VE1* - Spruce and other coniferous forest. Altogether, 1222 townships are covered by this forest. This forest type includes following categories:

(i) Spruce forest is the most dominant forest type covering more than 75% of a township area. There are 586 townships which belong to this forest type.

(ii) Spruce forest is the most dominant forest type and covers 35% to 75% of a township area. The second most dominant forest type is coniferous forest, such as pine, fir or larch forests. There are 636 townships which belong to this forest type.

(iii) Fir forest or larch forest are the most dominant forest type and spruce forest is the second most dominant. Only 8 townships belong to this forest type; therefore, they have been

grouped into spruce and other coniferous forest.

(2) *VE2* - Spruce and broad-leaf forest. Spruce forest is the most dominant forest and covers 35% to 75% of a township area. The second dominant forest is broad-leaf forest including poplar, maple, birch, or other broad-leaf trees. There are 273 townships in this category.

(3) *VE3* - Pine and other coniferous forest. There are 890 townships which are covered by this forest type and it includes two categories:

(i) Pine forest, the most dominant forest type, covers more than 75% of a township area. There are 345 townships which belong to this category.

(ii) Pine forest, the most dominant forest, covers less than 75% of township area and spruce forest is the second most dominant forest type. There are 545 townships which belong to this category.

(4) *VE4* - Pine and broad-leaf forest. There are 152 townships covered by this forest type. Pine forest is the most dominant forest and poplar forest is the second most dominant forest.

(5) *VE5* - Poplar and coniferous forest. There are 278 townships covered by this forest type. This forest is composed by two categories:

(i) Poplar is the most dominant forest. Either one of spruce, pine, fir, or larch forest is the second most dominant. There are 230 townships belong to this category.

(ii) There are 48 townships in which unspecified broad-leaf forest is the most dominant forest and spruce forest or pine forest is the second dominant forest. They are grouped into "poplar and coniferous forest" because we do not want to create a new forest type for this small number of townships.

(6) *VE6* - Poplar and other broad-leaf forest. There are 618 townships which belong to this forest type. This forest type includes three categories:

(i) Poplar forest is the most dominant forest type covering more than 75% of township area. There are 550 townships which belong to this category.

(ii) Poplar forest is the most dominant forest type covering 35% to 75% of township area. Other broad-leaf forest types are the

second dominant forest. There are 38 townships covered by this forest type.

(iii) Other broad-leaf forest is the most dominant forest in a township. Only 30 township belong to this category. Since we do not want to create another forest type for such a small number of townships; therefore, we group them into "poplar and other broad-leaf forest".

(7) *VE7* - Open land. The definition of open land is as follows: Open lands are those townships in which forest covers less than 1000 ha, which is approximately 10% of a township area. Open land is thus a township in which there is little or no forest. There are 100 townships which belong to this category. It must be noted that this category does not include prairie, tundra and the transition zone between tundra and boreal forest. Open land includes following categories:

(i) Urban areas (e.g. Winnipeg);

(ii) Those townships in which water body takes up a very high percentage of the township area;

(iii) Others, such as clear cutting and forest fires cause a township to become a open land.

From "*Natural vegetation of Manitoba*" map (Weir 1983), we obtained three other types of vegetation. These vegetation types are:

(1) *V10* - Prairie. There are 467 townships which belong to this category. This vegetation type includes four categories:(i) Tall grass prairie, it is the true prairie with trees confined to water courses; (ii) Mixed grass prairie, it is continuous with the mixed prairie of Saskatchewan; (iii) Sparsely wooded grassland, which is predominantly grassland with scattered groves of willow and aspen; and (iv) Wooded grassland, which is the transition from lightly treed areas to forest.

(2) *VE8* - Transition zone between boreal forest and tundra. There are 1597 townships which belong to this category. Conifers are mainly confined to valleys and decreasing stands on uplands.

(3) *VE9* - Tundra. There are 114 townships which belong to this category. In this area, trees are absent. The most common plants are stunted willcw, lichen, mosses, grasses and sedges.

Therefore, we generalized 10 vegetation types for this study area. *Table 3-2* gives a summary of these vegetation types including the names of vegetation types, the number of townships for each vegetation type and the percentage.

Table 3-2 Forest types and the number of corresponding townships

Vegetation types	Number of Township	Percentage(%)
VE1	1222	21.4
VE2	273	4.8
VE3	890	15.6
VE4	152	2.7
VE5	278	4.9
VE6	618	10.8
VE7	100	1.8
VE8	1597	30
VE9	114	2
V10	467	8.2
Total	5711	100

For the purpose of analysis, as we will see later, we generalized these 10 vegetation types into the following six categories:

(1) *V1* - Boreal forest. This includes spruce and other coniferous forest (*VE1*), spruce and broad-leaf forest (*VE2*), pine and other coniferous forest (*VE3*), and pine and broad-leaf forest (*VE4*);

(2) *V6* - Broad-leaf forest. It includes poplar and other coniferous forest (*VE5*), and poplar and other broad-leaf forest (*VE6*);

(3) *V7* - Open land. This is the same as the above *VE7*.

(4) *V8* - Transition zone between boreal forest and tundra. This is the same as the above *VE8*.

(5) *V9* - Tundra. This is the same as the above *VE9*.

(6) *V10* - Prairie. This is the same as the above *V10*.

Table 3-3 gives the summary of the generalized vegetation types including the names of vegetation types, the number of township for each vegetation type and the percentage.

Table 3-3 Forest types and the number of corresponding townships

Vegetation Types	Number of Townships	Percentage (%)
Boreal Forest (V1)	2537	44.4
Broad-leaf Forest (V6)	896	15.7
Open Land (V7)	100	1.8
Transition Zone (V8)	1597	28.0
Tundra (V9)	114	2.0
Prairie (V10)	467	8.2
Total	5711	100

Now we have two levels of forest groupings; one is more detailed and another is more general. Both will be used in this analysis;

however, we will choose one of them in prediction of forest response to global warming and this will be explained in Chapter 4.

3.4.2 Climatic Data

All climatic data come from the publication "***Canadian Climate Normals***" published by Atmospheric Environment Service, Environment Canada(*1982*). Most of these data are calculated as the averages of thirty years from 1951 to 1980. Since the forest data were collected in the seventies, climatic data and forest data do match well. One hundred and sixty nine weather stations' climatic data for the province of Manitoba are listed in "***Canadian Climate Normals***"; however, only forty nine weather stations' information were used (*Table 3-4*). The reason is that many weather stations have less than 30 years of climatic records. Generally, it was decided to discard climatic data from a weather station if its records length is less than thirty years. However, there are some exceptions. We clearly see from *Table 3-4* that there are not many weather stations are located in the north and north-east of Manitoba , and most of them have been recording data for less than 30 years. Under these circumstances, we do use those climatic data in the north and north-east of Manitoba even though they are based on less than 30 years' average. This may possibly impact climatic data precision; however, it is better to use these data because we need climatic data in the north and north-east of Manitoba for interpolation.

The selection of climatic variables is based on following principles: (1) data of a climatic variable must be available from "***Canadian Climate Normals***"; (2) data of a climatic variable must be available from GCMs prediction because we need future climatic data to predict forest response to global warming; (3) a climatic variable must be pure, which means that this variable cannot be represented by other variables. For example, drought is not a pure climatic variable because it is related to many other variables such as precipitation, temporal distribution of precipitation, temperature, soil texture, etc. Based on these principles, nine climatic variables were selected. These variables are:

(1) Annual temperature (normal of annual mean temperature);
(2) January temperature (normal of January daily mean temperature);
(3) April temperature (normal of April daily mean temperature);
(4) July temperature (normal of July daily mean temperature);
(5) October temperature (normal of October daily mean temperature);
(6) Annual precipitation (normal of annual mean precipitation);
(7) Rainfall in summer (normal of June, July, and August total precipitation);
(8) Total snowfall (normal of annual total snowfall);
(9) Growing degree days (normal of annual mean growing degree days). When daily average temperature is higher than 5^0C, the plant is biologically active. Growing degree days represent the length of growing season above the reference temperature (5^0C).

Table 3-4. Weather stations used in analysis.

Name of the Weather Station	Geographic Location	Years of Weather Record
Emerson	49°1'N 97°12'W	36 - 60
Sprague	49°2'N 95°38'W	64 - 65
Waskada	49°2'N 100°45'W	34 - 48
Altona	49°6'N 97°33'W	21 - 32
Morden	49°11'N 98°6'W	39 - 66
Deloraine	49°11'N 100°30'W	25 - 30
Pierson	49°11'N 101°16'W	43 - 60
Boissevain	49°14'N 100°3'W	24 - 47
Portage La Prairie	49°16'N 99°50'W	24 - 56
Morris	49°21'N 97°22'W	61 - 64
Deerwood	49°24'N 98°19'W	26 - 28
Ninette	49°24'N 99°38'W	58 - 62
Roland	49°25'N 98°0'W	25 - 27
Graysville	49°30'N 98°10'W	48 - 53
Cypress River	49°33'N 99°5'W	32 - 56
Indian Bay	49°37'N 95°12'W	63 - 65
Souris	49°37'N 100°16'W	53 - 56
St Albans	49°42'N 99°33'W	74 - 76
Virden	49°51'N 100°56'W	23 - 62
Brandon CDA	49°52'N 99°58'W	90 - 91
Winnipeg Int'l A	49°54'N 97°14'W	42 - 43
Rivers A	50°1'N 100°19'W	32 - 33
Seven Sisters Falls	50°7'N 96°1'W	18- 20
Hamiota	50°11'N 100°37'W	52 - 58
Minnedosa	50°16'N 99°50'W	84 - 90

Birtle	50°25'N 100°50'W	61 - 65
Great Falls	50°28'N 96°0'W	57 - 59
Gimli	50°38'N 97°3'W	27 - 28
Russell	50°46'N 101°17'W	68 - 74
Arborg	50°59'N 97°6'W	19 - 20
Dauphin A	51°6'N 100°3'W	37 - 38
Moosehorn	51°18'N 98°37'W	31 - 42
Swan River	52°7'N 101°16'W	25 - 48
Grand Rapids	53°11'N 99°16'W	11 - 14
Pasquia Projet PFRA	53°43'N 101°35'W	22 - 25
The Pas	53°49'N 101°15'W	57 - 58
Island Lake	53°52'N 94°40'W	10 - 11
The Pas A	53°58'N 101°6'W	37 - 38
Norway House Forestry	54°0'N 97°48'W	9 - 11
Wanless	54°11'N 101°21'W	29 - 36
Flin Flon	54°46'N 101°51'W	49 - 53
Wabowden	54°55'N 98°38'W	26 - 28
Thompson A	55°48'N 97°52'W	13 - 14
Laurie River Power Site	56°14'N 100°59'W	9 - 10
Gillam	56°21'N 94°42'W	17 - 20
Lynn Lake	56°51'N 101°2'W	12 - 15
Lynn Lake A	56°52'N 101°4'W	12 - 13
Brochet A	57°54'N 101°41'W	29 - 31
Churchill A	58°45'N 94°4'W	37 - 38

These nine climatic variables give a relatively complete and detailed description of the climatic conditions. Annual temperature is the average temperature for the twelve month. Growing degree days represents the length of growing season. Both of them give a general picture of the temperature of an area. January temperature represents the temperature of the coldest month and July temperature represents the temperature of the warmest month. April and October temperatures are the intermediate months between the coldest and the warmest months. January, April, July, and October temperatures can very well describe the temperature regime of an area. Annual precipitation gives the general picture of the water availability of an area. Rain-fall in summer and annual snowfall can roughly describe precipitation regime.

As said previously, the climatic data come from 49 weather stations in Manitoba (*Table 3-4*) and most of them concentrated in the south and southwest of the study area. For running *PEGASE*, the climatic data for each cell (township) are needed; however, we do not have weather station in every cell. This problem can be solved by spatial interpolation method.

The procedure of estimating the value of properties at unsampled sites within the area covered by existing point observations is called interpolation. "*The rationale behind spatial interpolation*

is the very common observation that, on average, points that are close together in space are more likely to have similar values of a property of interest than points further apart" (Burrough 1986). It is possible to make a precise statements about the climatic data at townships in which there are not weather station based on the climatic data of those 49 weather stations.

Most GIS software provide interpolation function. Considering the fact of software availability limitation (*IDRISI* and *ARC/INFO*) and the fact that our research area is divided into a grid system, we used the interpolation function provided by *IDRISI* (a raster-based GIS software, version 3.0) to obtained the climatic data for all the townships for the above nine variables based on these 49 weather stations. Since the quality of spatial interpolation depends on the number of reference points used, the climatic data for those townships in the north and north-east are not considered to be as good as those in the south and south-west. However, there is no any other method which can give more precise climatic data.

In order to use the climatic data in *PEGASE*, the climatic data need to be classified into several classes. The following tables (*Table 3-5* to *Table 3-13*) give the range of climatic data of each class and its code for all the climatic variables.

Table 3-5 The classification of annual temperature

Code of Annual Temperature Class	Temperature Range
TA1	<-7.00°C
TA2	-7.00 to -6.01°C
TA3	-6.00 to -5.01°C
TA4	-5.00 to -4.01°C
TA5	-4.00 to -3.01°C
TA6	-3.00 to -2.01°C
TA7	-2.00 to -1.01°C
TA8	-1.00 to -0.01°C
TA9	0.00 to 0.99°C
T10	1.00 to 1.99°C
T11	2.00 to 2.99°C
T12	>=3.00°C

Table 3-6 The classification of October temperature

Code of October Temperature Class	Temperature Range
TO1	<0.00 °C
TO2	0.00 to 0.99°C
TO3	1.00 to 1.99°C
TO4	2.00 to 2.99°C
TO5	3.00 to 3.99°C
TO6	4.00 to 4.99°C
TO7	5.00 to 5.99°C
TO8	6.00 to 6.99°C
TO9	>=7.00 °C

Table 3-7 The classification of July temperature

Code of July Temperature Class	Temperature Range
UL1	<12.00°C
UL2	12.00 to 12.99°C
UL3	13.00 to 13.99°C
UL4	14.00 to 14.99°C
UL5	15.00 to 15.99°C
UL6	16.00 to 16.99°C
UL7	17.00 to 17.99°C
UL8	18.00 to 18.99°C
UL9	19.00 to 19.99°C
U10	>= 20.00°C

Table 3-8 The classification of April temperature

Code of April Temperature Class	Temperature Range
AP1	<-6.00°C
AP2	-6.00 to -5.01°C
AP3	-5.00 to -4.01°C
AP4	-4.00 to -3.01°C
AP5	-3.00 to -2.01°C
AP6	-2.00 to -1.01°C
AP7	-1.00 to -0.01°C
AP8	0.00 to 0.99°C
AP9	1.00 to 1.99°C
A10	2.00 to 2.99°C
A11	>= 3.00°C

Table 3-9 The classification of January temperature

Code of January Temperature Class	Temperature Range
JT1	<-28.00°C
JT2	-28.00 to -27.01°C
JT3	-27.00 to -26.01°C
JT4	-26.00 to -25.01°C
JT5	-25.00 to -24.01°C
JT6	-24.00 to -23.01°C
JT7	-23.00 to -22.01°C
JT8	-22.00 to -21.01°C
JT9	-21.00 to -20.01°C
J10	-20.00 to -19.01°C
J11	-19.00 to -18.01°C
J12	>= -18.00°C

Table 3-10 The classification of annual precipitation

Code of Annual Precipitation Class	Precipitation Range
RO1	<420.0mm
RO2	420.0 to 439.9mm
RO3	440.0 to 459.9mm
RO4	460.0 to 479.9mm
RO5	480.0 to 499.9mm
RO6	500.0 to 519.9mm
RO7	520.0 to 539.9mm
RO8	>=540.0mm

Table 3-11 The classification of growing degree days

Code of Growing Degree Days Class	Range of Growing Degree Days
G01	<800.0
G02	800.0 to 899.9
G03	900.0 to 999.9
G04	1000.0 to 1099.9
G05	1100.0 to 1199.9
G06	1200.0 to 1299.9
G07	1300.0 to 1399.9
G08	1400.0 to 1499.9
G09	1500.0 to 1599.9
G10	1600.0 to 1699.9
G11	1700.0 to 1799.9
G12	>=1800.0

Table 3-12 The classification of rainfall in summer

Code of Rainfall Class	Range of Rainfall
RG1	<200.0mm
RG2	200.0 to 224.9mm
RG3	225.0 to 249.9mm
RG4	250.0 to 274.9mm
RG5	275.0 to 299.9mm
RG6	>=300.0mm

Table 3-13 The classification of snowfall

Code of Snowfall Class	Range of Snowfall
SN1	<100.0mm
SN2	100.0 to 119.9mm
SN3	120.0 to 139.9mm
SN4	140.0 to 159.9mm
SN5	160.0 to 179.9mm
SN6	180.0 to 199.9mm
SN7	200.0 to 219.9mm
SN8	220.0 to 239.9mm
SN9	>=240.0mm

3.4.3 Data of Non-climatic Variables

The non-climatic variables used in this study include soil type, soil texture, surficial geological deposit, and topography.

3.4.3.1 Data of Soil Texture

Both soil type and soil texture data are available in the province of Manitoba. The use of soil type in an environmental study may raise a problem, since the formation of a soil type is a function of climatic conditions, vegetation, and parent material. A link

between soil type and vegetation would be difficult to interpret. In this case, it would be more appropriate to say that the "vegetation" determines the "type of soil" rather than the other way around. Although soil type has been proven to be a good predictor of vegetation distribution when it is included in the analysis, we discarded it since it cannot be considered as an independent environmental variable. However, soil properties are important to vegetation. Soil texture was chosen as an environmental variable for two reasons. First, it has an important influence on soil moisture. Soil capacity for moisture can be very different for various texture types, such as clay soil and sandy soil, even when both receive the same amount of precipitation. Soil moisture can influence plant growth, for example, jack pine forests are frequently found in sandy soil. Drier soils have higher fire frequency which influence the development of Jack pine forest. Secondly, soil texture, unlike soil type, is an independent environmental variable.

Data of soil texture are taken from the map of "***Manitoba Soil Landscapes***", which is part of the "***Soil Landscapes Of Canada***" project carried out by the Land Research Centre, Research Branch, Agriculture Canada (*1986*). Soil textures are grouped into 7 categories as shown in *Table 3-14*:

Table 3-14. Soil texture data

Code of Soil Texture Class	Soil Texture Type
ST1	Sand or sandy loam
ST2	Loam or clay loam
ST3	Clay
ST4	Mesic sedge material or Mesic woody forest
ST5	Fibric sphagnum material
ST6	Hardrock outcrop (granite)
ST7	Hardrock outcrop (limestone)

ST1, *ST2*, and *ST3* are classified according to the size of soil particles. *Sand or sandy loam* (*ST1*) includes sand, loamy sand, gravelly sand, fine sandy loam, sandy loam, and gravelly sandy loam. *Loam or clay loam* (*ST2*) very fine sandy loam, silt loam and loam. *Clay* (*ST3*) includes clay, silty clay and heavy clay (*Land Resource Research Centre, Research Branch, Agriculture Canada, 1986*). Clay has a higher percentage of fine soil particles which results in a higher soil water holding capacity than sand or sandy soil (*ST1*). *ST4* and *ST5* are not classified by the size of soil particles; however, they are classified by their specific environments. *Mesic sedge material and mesic woody forest* (*ST4*) represents mesic to humid soils associated with forest. *Fibric sphagnum material* (*ST5*) represents an organic material in a

incomplete degree of decomposition in which the fibric materials are readily identifiable as to botanical origin. *ST6* and *ST7* are rock outcrop and are non-soil.

3.4.3.2 Data of Surficial Geological Deposit

The surficial geological deposit is included as a potential environmental variable. In the study of forest response to global warming as it may be considered as an indicator of soil parent material. Parent material is the mineral body which existed before pedogenetic processes were initiated and upon which these processes acted to produce a soil under the influence of climate and vegetation. Parent material, as we know, is one of the major factors which influences soil formation and properties. Since we cannot get the parent material itself, we use surficial geological deposit as a substitute for parent material.

Data of surficial geological deposit were taken from the ***"Surficial Geological Map of Manitoba"*** compiled by *Manitoba Mineral Resources Division (1981)*. There are 12 categories of surficial geological deposits in Manitoba which are listed in *Table 3-15*.

Table 3-15. Surficial geological deposit data

Code of Surficial Geological Deposit Class	Surficial Geological Deposit Types
GE1	Organic deposits
GE2	Alluvial deposits
GE3	Beach deposits
GE4	Nearshore, intertidal deposits
GE5	Beach and nearshore deposits
GE6	Deep basin deposits
GE7	Deltaic deposits
GE8	Glaciofluvial deposits
GE9	Glacial deposits, mainly derived from Palaeozoic carbonate rock
G10	Glacial deposits, mainly derived from Mesozoic shales
G11	Glacial deposits, mainly derived from Precambrian bedrock
G12	Bedrock

Surficial geological deposits are classified based on the deposit environment and the geological time. Four types of environments are classified:

(1) Nonglacial environment, which includes organic deposits (*GE1*), alluvial deposits (*GE2*), beach deposits (*GE3*) and nearshore and intertidal deposits (*GE4*). The corresponding geological time is

Holocene.

(2) Preglacial and glacial environment, which includes beach and nearshore deposits (*GE5*), deep basin deposits (*GE6*), deltaic deposits (*GE7*) and glaciofluvial deposits (*GE8*). The geological time of these deposits were before the Holocene and after Late Wisconsinan.

(3) Glacial environment, which includes all the glacial deposits (*GE9, G10 and G11*). The corresponding geological time is Late Wisconsinan.

(4) Nonglacial environment, which includes the bedrock (*G12*). The geological time is pre-Quaternary.

The geological deposits in different deposit environments and at different geological times usually have: (1) different composition of clay, silt, sand, gravel, and till; (2) different thicknesses. *Organic deposits (GE1)* include marsh, fen, swamp and bog deposits up to 6 metre thick. It is characterized by seasonal flooding and it is generally found in Precambrian terrain and overlying fine-textured preglacial lake and glacial till deposits. *Alluvial deposits (GE2)* include sand, silt gravel and clay, which can be 1-30 metre thick. It deposited on flood plains along river valleys,

on floors of glacial spillways, and as alluvial fans along the base of escarpments. *Beach deposits (GE3)* are composed of sand and gravel, 1-3 metre thick, forming distinct ridges and including areas of littoral sand and offshore bars. *Nearshore, intertidal deposits (GE4)* include coarsening upward sequence of clay, silt and sand. It is 1-14 metre thick and deposited as a flat plain. *Beach and nearshore deposits (GE5)* are composed of sand and gravel, 1-4 metre thick, forming distinct ridges, sand blankets (0-1 metre thick) reflecting form of underlying topography and areas of wave-washed till and exposed bedrock. *Deep basin deposits (GE6)* include silt, sand and clay deposits, 1-30 metre thick, forming extensive lake plains and a discontinuous veneer reflecting the form of the underlying topography. There are areas with evidence of extensive iceberg scouring, areas masked by thin alluvium, and areas of wave-washed till and exposed bedrock. *Deltaic deposits (GE7)* include sand, silt and gravel and it is 1-30 metre thick, coarsening towards the apex. The surface is flat and gently sloping and there is extensive dune development in places. *Glaciofluvial deposits (GE8)* include gravel, sand and silt, 1-100 metre thick, deposited in an ice marginal environment. It includes subaqueous outwash deposited in glacial lakes, and some late glacial valley fill. *Glacial deposits, mainly derived Palaeozoic carbonate rock (GE9)* are 1-10 metre thick, highly calcareous and the typical composition is 16 percent gravel, 21 percent sand, 44 percent silt and 19

percent clay. *Glacial deposits, mainly derived from Mesozoic shales (G10)* represent a glacial environment. They are 1-10 metre thick, moderately calcareous and the typical composition is 13 percent gravel, 33 percent sand, 31 percent silt and 23 percent clay. *Glacial deposits, mainly derived from Precambrian bedrock (G11)* are 1-5 metre thick and the typical composition is 28 percent gravel, 46 percent sand, 22 percent silt and 4 percent clay. *Bedrock (G12)* are Precambrian, Palaeozoic and Mesozoic bedrock of various lithologies.

The detailed information of the surficial geological deposits is available from the "***Surficial Geological Map of Manitoba***" (*Manitoba Mineral Resources Division 1981*).

3.4.3.3 Topography Data

We include topography as a potential environmental variable in the study of forest response to global warming because topography may possibly impact the forest distribution at a regional scale. We do not know to what extent topography can contribute to the forest distribution since the topography does not greatly vary in the study area. It is included as a candidate and will be tested using the *PEGASE* to see if it is a critical environmental variable.

Topography data were taken from the map of "***Manitoba Relief***" published by Surveys and Mapping Branch, Province of Manitoba in 1979. The relief is classified into five groups according to the elevation of the township and the classification information is listed in *Table 3-16*:

Table 3-16. Topography data

Code of Topography Class	Topography Classification
LF1	the elevation is higher than 1600 ft
LF2	the elevation is between 1300 to 1600 ft
LF3	the elevation is between 1000 to 1300 ft
LF4	the elevation is between 500 to 1000 ft
LF5	the elevation is between 0 to 500 ft

In total, twelve environmental variables are included in the data base. Some other environmental variables could be very interesting to be included in this study. For example, forest fire may be a major ecological factor with regard to forest development at a regional scale in boreal forest. In terms of the methodology, it is possible to include this variable in our model; however, it is not

included because data with the necessary level of detail were not available. Similarly, human impact on the forest distribution pattern, such as clear cut and urbanization, is not taken into account in this study. The exclusion of these variables, as we mentioned before, will weaken the association between forest distribution and environmental variables.

Chapter 4. The Construction of the Knowledge Based Forest Model

In the previous chapter, we have introduced the methodology and the procedure of *KBFM*. Since this model is closely related to the particular forest and environmental variable data we could not explain it in detail. In this chapter, we will demonstrate the processes of *KBFM* construction at a regional scale so that the reader can have a better understanding of this model. Its construction includes following three stages:

(1) Selection of the critical environmental variables;

(2) Construction of the *KBFM's* knowledge base;

(3) Test of the validity of the *KBFM*.

4.1 Selection of the Critical Environmental Variables.

As we mentioned in chapter 3, a critical environmental variable is the one which is a good predictor of vegetation distribution. A non-critical environmental variable, on the contrary, is the one

which is not a good predictor of vegetation distribution. We want to know if those 12 environmental variables we selected in chapter 3 are the critical environmental variables.

The purpose of the selection of the critical environmental variables is to discard those environmental variables which do not play a significant role as predictors of the forest types and keep only variables playing a key role in the *KBFM*. This deletion will make the *KBFM's* knowledge base more concise, less redundant, and easier to computerize. Of course, the selection process should guarantee no significant loss in the precision of the prediction. In this selection process, we have to answer the following question: are there some variables which do not contribute much to the reduction of the entropy of forest distribution so that we can eliminate them without losing much useful information?

The procedure of critical variable selection includes following steps:

(1) For all environmental variables, find their contribution to the reduction of the forest distribution entropy using *PEGASE;* and rank them in an order of negentropy importance;

(2) Delete the environmental variable which reduces most forest

distribution entropy in step 1, then re-run *PEGASE* excluding this variable. Then repeat step 1 using the rest of the environmental variables, until the condition in (3) occurs;

(3) Stop this procedure when the remaining environmental variables cannot significantly reduce forest distribution entropy. This means that the remaining environmental variables are not critical environmental variables.

We will see that this analysis process can be very helpful for determining the pattern of environmental variables' contribution to the reduction of forest distribution entropy.

4.1.1 Analysis Processes to determine the Critical Environmental Variables

We run *PEGASE* to select the critical environmental variables following the procedure of critical variable selection described above. In total, eight steps were successively analyzed. *Table 4-1* gives detailed information of the variables used in these eight steps.

Table 4-1. The Explanation of the critical environmental variable selection

Steps	Variables excluded from the analysis
Step 1	None(All the 12 environmental variables are included)
Step 2	OctTem
Step 3	OctTem, JulTem
Step 4	OctTem, JulTem, AnnTem
Step 5	OctTem, JulTem, AnnTem, AprTem
Step 6	OctTem, JulTem, AnnTem, AprTem, JanTem
Step 7	OctTem, JulTem, AnnTem, AprTem, JanTem, GroDeg
Step 8	OctTem, JulTem, AnnTem, AprTem, JanTem, GroDeg, SoiTex

Note: OctTem stands for October temperature. JulTem stands for July temperature. AnnTem stands for Annual average temperature. AprTem stands for April temperature. JanTem stands for January temperature. GroDeg stands for growing degree days. SoiTex stands for soil texture.

Running the *PEGASE* analysis with the 12 environmental variables (*Step 1*), the initial entropy is 1.982 with a final entropy 0.784, representing 60.5% (1.199) of negentropy. The final entropy (0.784) means that these 12 environmental variables do not provide all the information of forest distribution. There are some other environmental variables or factors (such as human disturbance, forest fire, or random factor, etc.) influencing forest distribution.

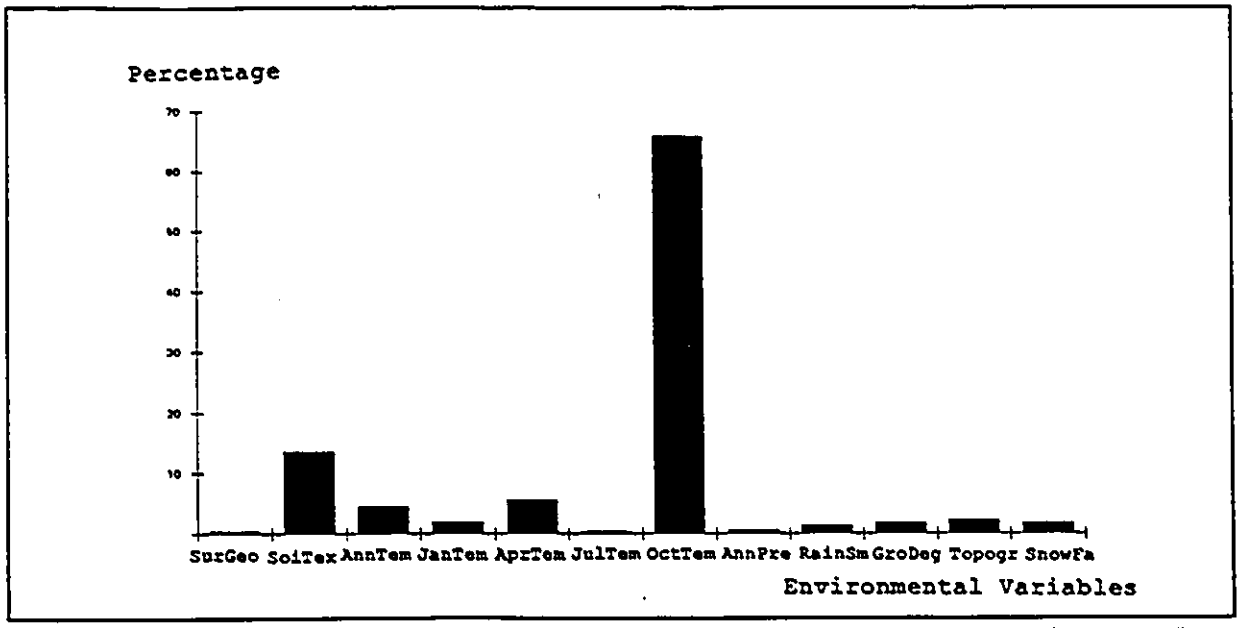
Among the 12 variables, October temperature is the most critical variable and contributes 65.9% of the total negentropy. The second

critical variable is soil texture which contributes 13.6% of the negentropy. *Figure 4-1* gives all variables' contribution to the entropy reduction. At this point we know October temperature is the most critical environmental variable. We re-run the analysis by excluding October temperature (*Step 2*) and we find that July temperature becomes the most critical variables among the remaining 11 variables. Soil texture still remains the second critical variable (*Figure 4-2*). We keep doing this analysis step by step and excluding the most critical variable until only 4 environmental variables are left. These 4 variables are annual precipitation, rain in summer, topography, and surficial geological deposits. We stop the analysis because these 4 variables never contribute significantly to the reduction of entropy. *Figure 4-3 to figure 4-8* give environmental variables' contribution the reduction of entropy for step 3 to step 8 respectively.

From the above eight analytical steps, we have following conclusions:

(1) In this study area, temperature (seasonal temperatures or annual temperature) or temperature-related variables (growing degree days) are the critical variables controlling the spatial distribution of forest. In all the steps(*step 1 to step 6*) shown in Table 4-3, this group of variables display the highest negentropy

Figure 4-1. Environmental variables' contributions to the reduction of forest distribution entropy (Step 1)



Note: SurGeo, SoiTex, AnnTem, JanTem, AprTem, JulTem, OctTem, AnnPre, RainSm, GroDeg, Topogr, and SnowFa stand for surficial geological deposit, soil texture, annual temperature, January temperature, April temperature, July temperature, October temperature, annual precipitation, rainfall in summer, growing degree days, topography, and snowfall respectively in Table 4-1 to Table 4-8.

Figure 4-2. Environmental variables' contributions to the reduction of forest distribution entropy (Step 2)

Figure 4-3. Environmental variables' contributions to the reduction of forest distribution entropy (Step 3)

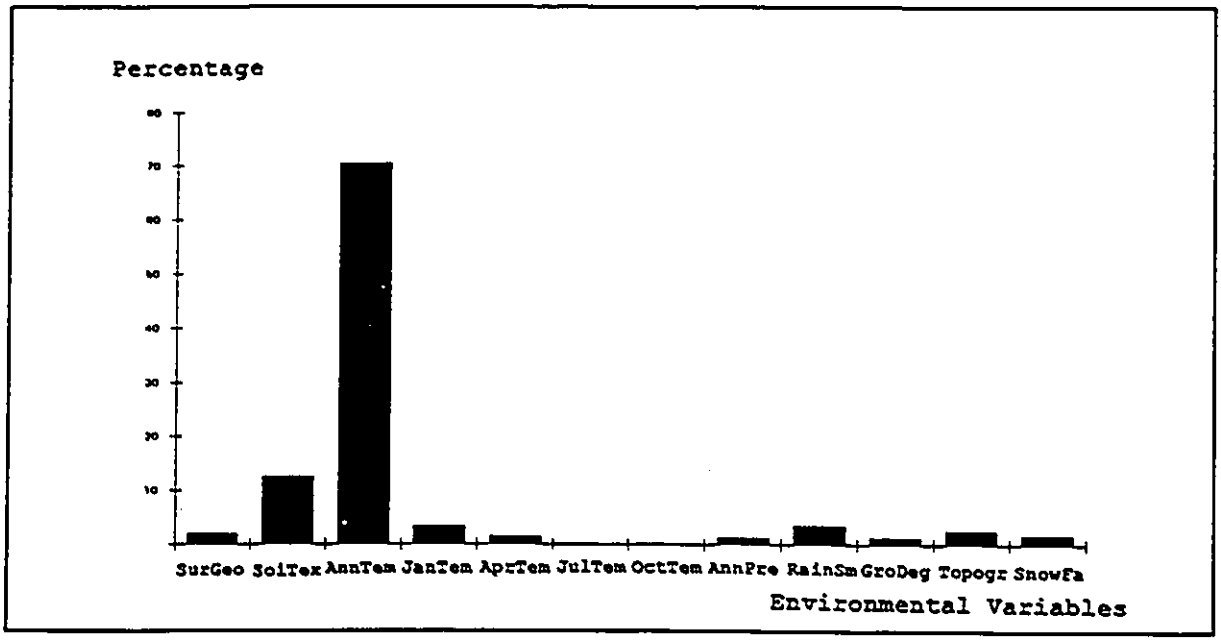


Figure 4-4. Environmental variables' contributions to the reduction of forest distribution entropy (Step 4)

Figure 4-5 Environmental variables' contribution to the reduction of forest distribution entropy (step 5)

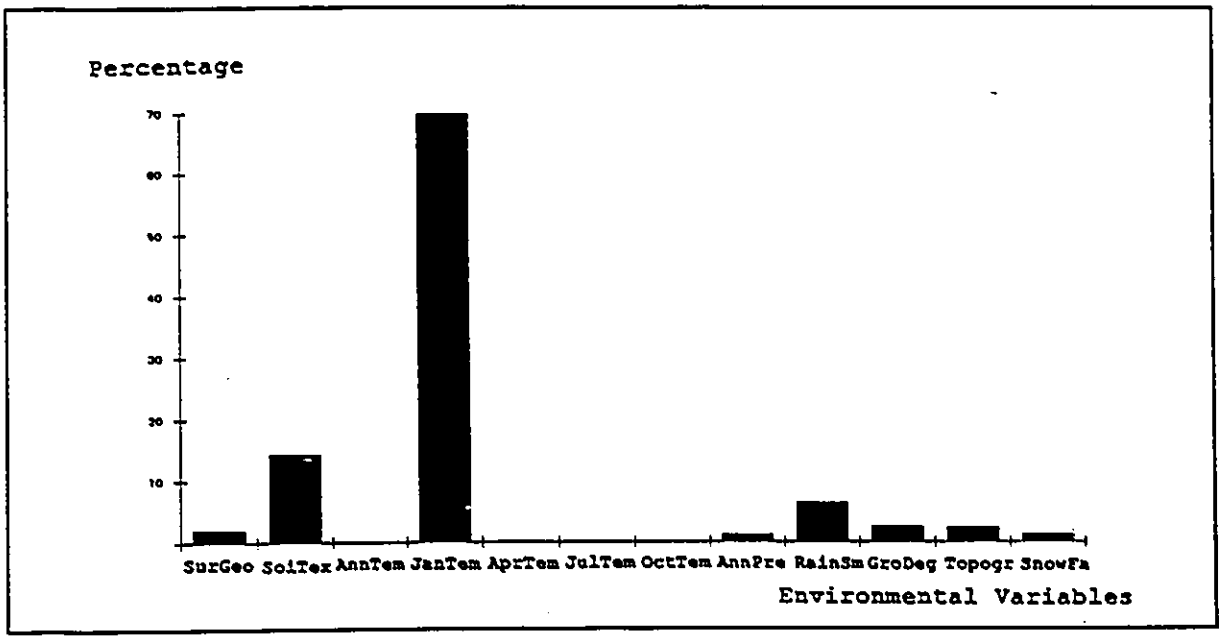


Figure 4-6. Environmental variables' contributions of the reduction of forest distribution entropy (Step 6)

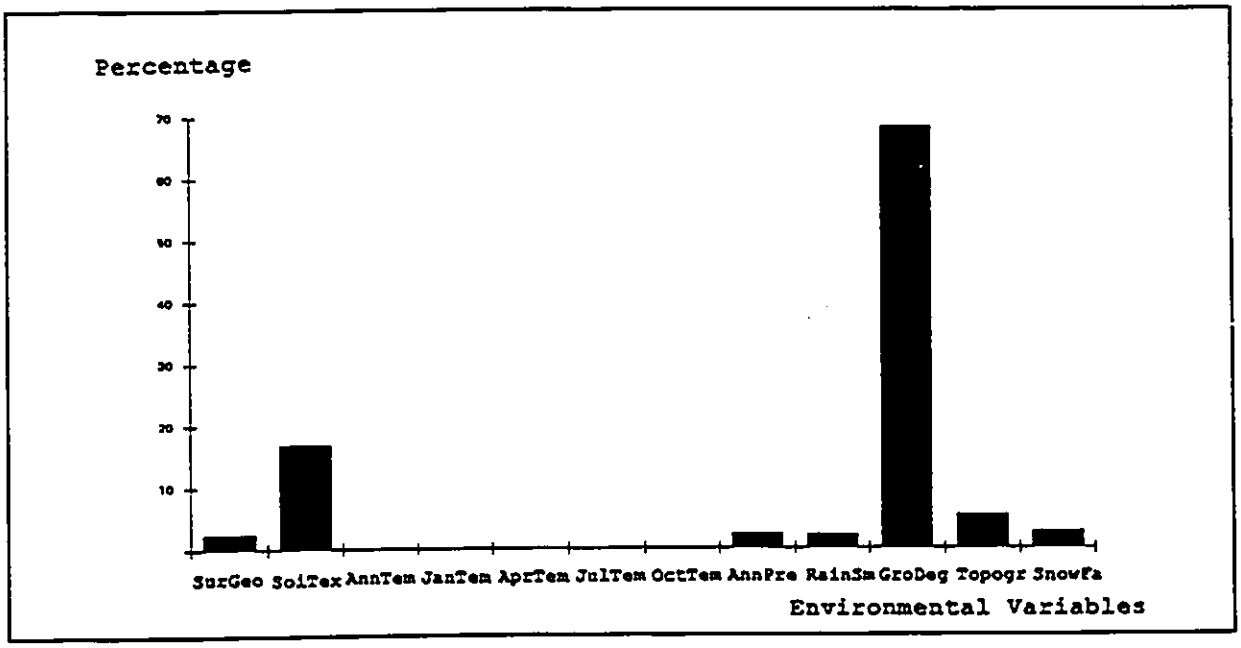


Figure 4-7. Environmental variables' contributions of the reduction of forest distribution entropy (Step 7)

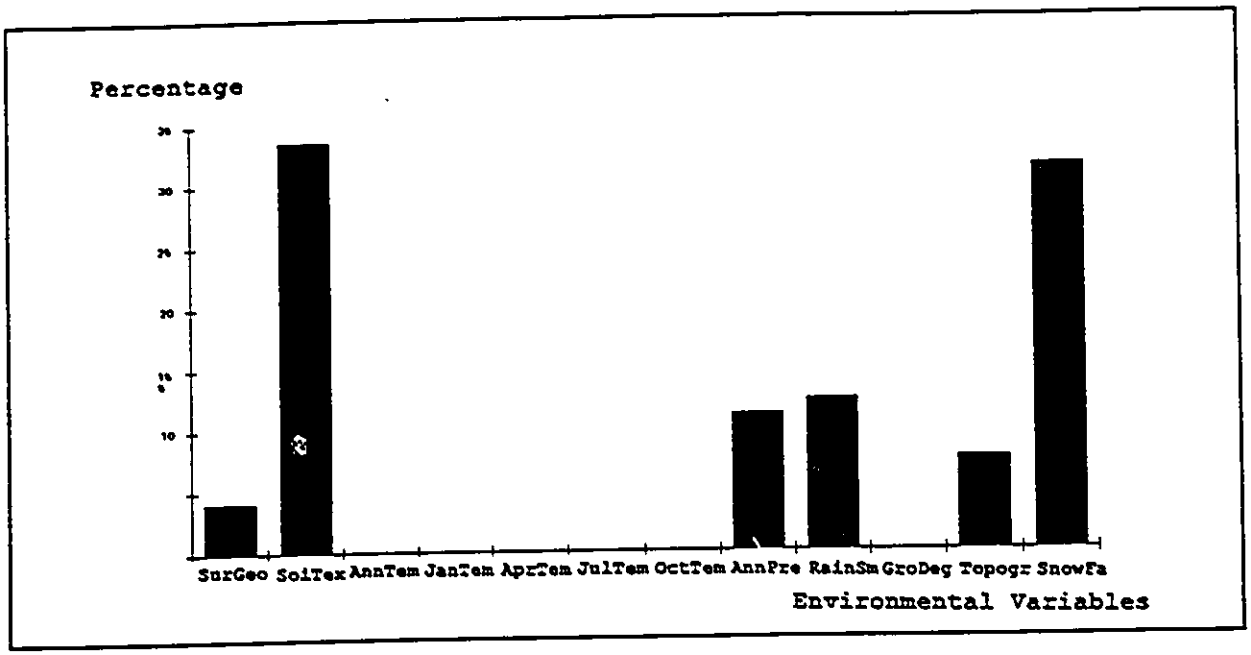
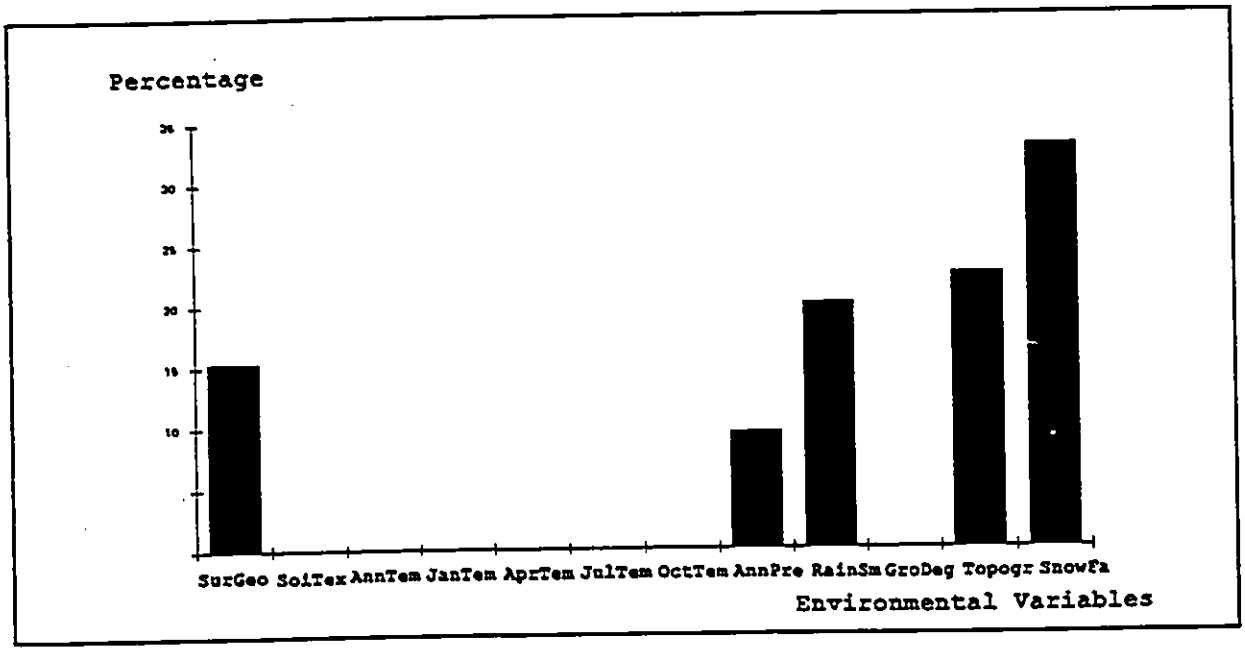


Figure 4-8. Environmental variables' contributions of the reduction of forest distribution entropy (Step 8)



values (*Table 4-2*). The pattern of the reduction of the entropy shows that one variable of this group contributes 60-70% of the negentropy while the rest variables of this group together contribute only 1-5% of the negentropy. This is due to the fact that temperature or temperature-related variables carry similar relative information. These climatic variables are strongly correlated and contain redundant information. Therefore, it is not necessary to include all the temperature and temperature-related variables in the predictive model.

Table 4-2. The most and the second most critical environmental variables

Steps	The Most Critical Variable	The Second Most Critical Variable
Step 1	October Temperature	Soil Texture
Step 2	July Temperature	Soil Texture
Step 3	Annual Temperature	Soil Texture
Step 4	April Temperature	Soil Texture
Step 5	January Temperature	Soil Texture
Step 6	Growing Degree Days	Soil Texture
Step 7	Soil Texture	Snow Fall
Step 8	Snow Fall	Topography

Among the all temperature and temperature-related variables, October temperature is found to be the most critical variable in

this study. We will discuss this late on.

(2) From step 1 to step 6, soil texture consistently appears as the second most critical variable. When a critical temperature or temperature-related variable is deleted, soil texture does not replace the first predictor but remains the second most critical variable until all the temperature-related variables have been deleted. As shown in *Table 4-2*, soil texture becomes the most critical variable in step 7 when all the temperature and temperature-related variables are deleted.

From this analysis, we may draw the following conclusions:

(1) Temperature and temperature-related variables are the most critical environmental variables. They contain redundant information; therefore, it will not be necessary to keep all of them in the KBFM.

(2) Soil texture is also a critical environmental variable. It always ranks the second most critical variable. Information about forest distribution provided by soil texture is independent of temperature and temperature-related variables.

(3) Three precipitation-related variables - snowfall, rainfall in

summer and annual precipitation - contribute about 5-7% of the total negentropy. Snowfall contributes more to the negentropy than rain in summer or annual precipitation (*Step 8*).

(4) Surficial geological deposit and topography contribute about 1-2% to the negentropy individually. They never appear as critical environmental variables. It does not make difference whether we include them in the *KBFM* or not since their contributions are negligible. Therefore, we can discard these two variables.

4.1.2 The Final Selection of the Critical Variables

Based on the previous analysis of the environmental variables' contribution to the reduction of forest distribution entropy, we include following variables in the final predictive model:

(1) January temperature;
(2) April temperature;
(3) July temperature;
(4) October temperature;
(5) Soil texture;
(6) Annual precipitation.

Annual temperature and growing degree days were not included since information carried by these two variables is also expressed by January, April, July and October temperatures. In other words, annual temperature and growing degree days are more or less redundant with January, April, July and October temperatures. We will lose very little information when these two variables are discarded. January, April, July and October temperatures are selected because they can very well represent the annual temperature pattern. Moreover, the data of these variables are available from the prediction of GCMs.

Among the three precipitation-related variables, annual precipitation is chosen although it provides less negentropy than snowfall and summer rainfall. However, we believe that snowfall and summer rainfall represent, to some extent, temperature-related factors. Annual precipitation is obviously the most independent from the temperature-related factor and was chosen.

Obviously, soil texture must be included because it always ranks as the second most critical variable. Other non-climatical variables will not be included since they are not critical environmental variables.

Among the above six variables, the combination of July temperature, annual precipitation and soil texture can very well describe the

probability of drought. Similarly, the combination of April and October temperature can very describe the frost-free days. Therefore, we can treat July temperature, annual precipitation and soil texture as the proxies for probability of drought and April temperature and October temperature as the proxies for frost-free days.

4.1.3 Suitability of the Selected Environmental Variables

Now we can demonstrate that the selection of the above six environmental variables and the deletion of the others will not reduce the quality of the prediction of the forest distribution. We simply run *PEGASE* with the six selected environmental variables and compare the result with the result when we include all the twelve environmental variables. We find that we get almost the same amount of information with the six selected environmental variables as we did with those twelve variables (*Table 4-3*). When 10 forest types are used, the entropy removed drops only 1.1% after we discarded the non-critical environmental variables. It decreases 2.3% when 6 forest types are used.

Table 4-3 Comparison of the PEGASE outcomes before and after selection of critical environmental variables.

Forest Classification Groups	Before or After Var. Selection	Number of *TSS*	Initial Entropy	Final Entropy	Entropy Removed (%)
The case when 10 forest types are used	Before (12 var.)	108	1.984	0.784	60.5
	After (6 var.)	107	1.985	0.807	59.4
The case when 6 forest types are used	Before (12 Var.)	104	1.414	0.404	71.4
	After (6 Var.)	76	1.415	0.483	69.1

Note: The case when 10 forest types are used includes VE1, VE2, VE3, VE4, VE5, VE6, VE7, VE8, VE9, V10 and the case when 6 forest types are used includes V1, V6, V7, V8, V9, V10. A TSS is a terminal subset coming from PEGASE analysis.

4.1.4 Suitability of the Selected Environmental Variables at a Scale of Study Area

In the previous chapter, we have mentioned that the selection of environmental variables in the study of forest response to climatic change depends on the scale of the study area. In this section, we will demonstrate this assumption and show that the selection of the

6 environmental variables are suitable to the current study scale by the following analysis. This analysis help us to understand the impact of the scale of study area on the selection of environmental variables.

We chose 9 sampling areas which correspond to a set of decreasing size, as described in *Table 4-4*. The first sampling area corresponds to the entire research area which is approximately 126 by 51 townships (we have 126 rows and each row has not a fixed number of townships because some of them are water body). The 8 other sampling areas exclude a part of the northernmost research area. For example, sample area 2 is an area excluding the northmost 26 rows of townships (approximately 100 by 51 townships remain) and sample area 9 excludes the northmost 100 rows of townships (approximately 26 by 51 townships remain). Table 4-4 gives the information of these 9 sampling areas.

We successively investigate the relationship between forest and environmental variables in each of these sampling areas. The purpose is to explore how the relationship between forest and environmental variables will be affected by the change in the sampling area.

Table 4-4 The explanation of the sampling area

Sampling Areas	Size (row by column)	Area (1000 Square KM)	Description
Sample Area 1	126 by 51	590	The entire research area
Sample Area 2	100 by 51	470	Excludes the northmost 26 rows.
Sample Area 3	88 by 51	410	Excludes the northmost 40 rows
Sample Area 4	76 by 51	357	Excludes the northmost 50 rows
Sample Area 5	66 by 51	310	Excludes the northmost 60 rows
Sample Area 6	56 by 51	260	Excludes the northmost 70 rows
Sample Area 7	46 by 51	216	Excludes the northmost 80 rows
Sample Area 8	36 by 51	169	Excludes the northmost 90 rows
Sample Area 9	26 by 51	122	Excludes the northmost 100 rows

Note: The information of the size and area in this table are approximate values.

By doing the *PEGASE* analysis for all the nine sampling areas, we find that the change of the study area has a great impact on environmental variables' contribution to the reduction of the entropy of forest distribution (*Figure 4-9 to Figure 4-11*). *Figure 4-9* shows the changes of initial entropy and final entropy as the sampling area changes. The ratio of the removed entropy, which equals to the initial entropy minus the final entropy divided by the initial entropy increase as the study area increase. This means

that the selection of the environmental variables and the forest types under the study area (sample area 1 in *Table 4-4*) reach a point where we get the most knowledge about the forest types. This figure also suggests that the environmental variables should be different when dealing with a smaller area (e.g. sample area 9 in *Table 4-4*) in order to reduce the entropy significantly. *Figure 4-10* shows the contribution of environmental variables to the reduction of forest distribution entropy (Six other variables are not shown in this figure because their contribution is small). We see that soil texture is the most critical environmental variable when the sampling area is equal to or smaller than 260,000 square KM (sample area 6 in *Table 4-4*) dominantly in prairie or transitional prairie and southern boreal forest. About 33 - 65% of the entropy was removed by soil texture. However, as the area increases, soil texture becomes the second most critical variable when October temperature becomes the most critical variable. When we classify the environmental variables into climatic variables and non-climatic variables, we obtain *Figure 4-11* which shows the contributions of these two group variables. From this figure, we find a threshold area in this study, about 235,000 square KM, which differentiates the contributions of climatic and non-climatic variables to the reduction of forest distribution entropy. When the sampling area is smaller than this threshold (sample areas 7-9 in *Table 4-4*), non-climatic variables are predominant in the entropy

Figure 4-9. The change of initial and final entropies as sampling area changes.

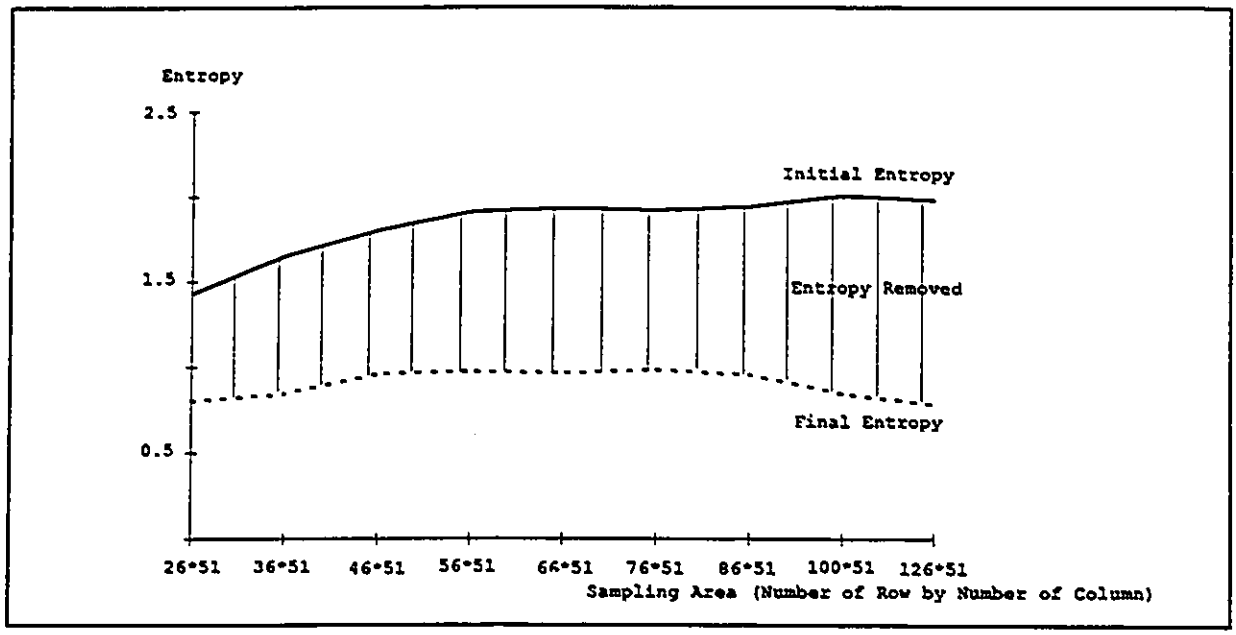


Figure 4-10. Variables' contributions to the reduction of forest distribution entropy as sampling area changes

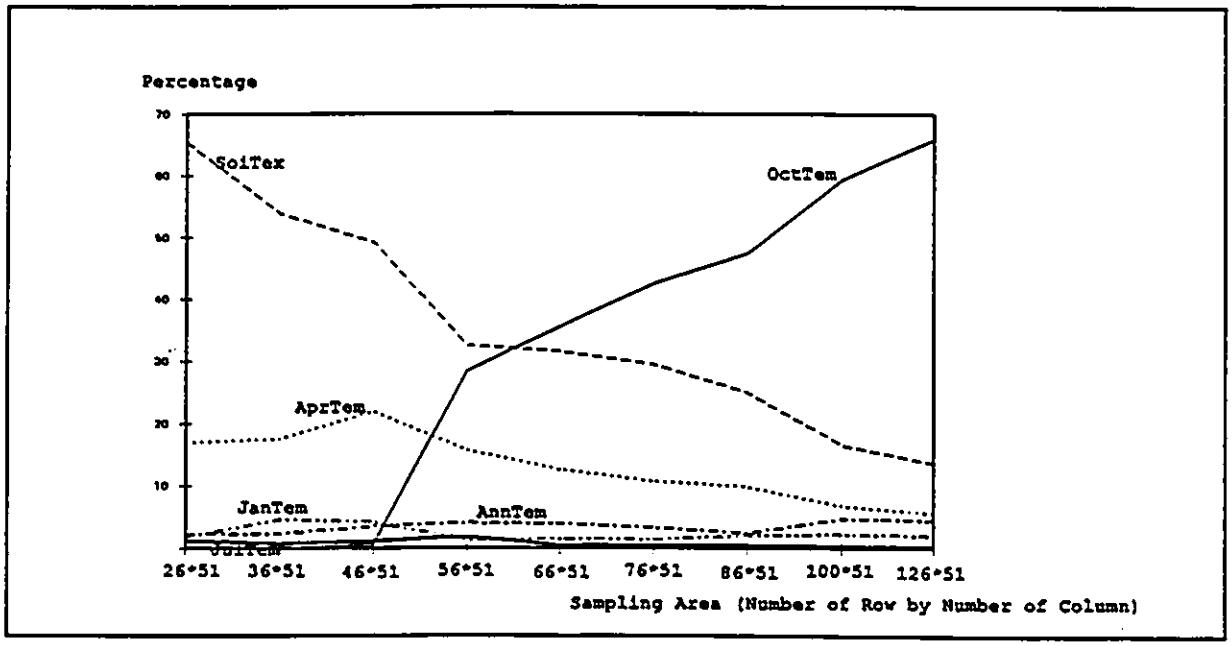
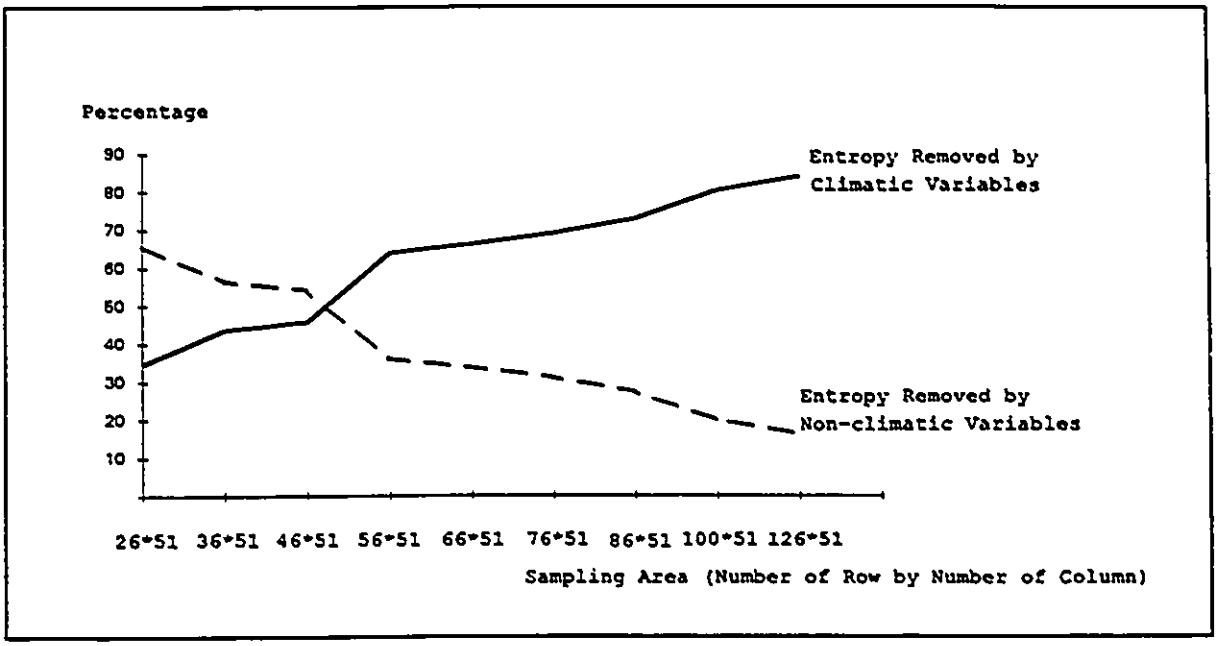


Figure 4-11. Climatic and non-climatic variables' contributions to the reduction of forest distribution entropy



reduction as compared to climatic variables, and when the sampling area is greater than this threshold (sample areas 1-6 in *Table 4-4*), the situation is reversed: climatic variables become predominant in entropy reduction whereas non-climatic variables become secondary.

This pattern can be easily understood. Within a relatively small research area, climate does not change greatly. In other words, all the cells within this area have relatively similar climatic conditions. Consequently, climatic variables cannot be good predictors of forest distribution. Conversely, non-climatic

variables, such as soil texture and topography, can display a spatial variation in such a small area and, in addition, they can be strongly related to the variation of forest types. Therefore, non-climatic variables can potentially be good predictors of forest distribution in small sampling area. As the size of sampling area increases, climate becomes more and more spatially variable and therefore the contribution of climatic variables to the explanation of forest distribution increases. When the sampling area reaches a threshold, climatic and non-climatic variables contribute approximately the same amount of negentropy to the forest distribution. If we keep increasing the size of the sampling area, climatic variables will contribute more than non-climatic variables to the forest distribution.

Since the research area of this thesis is larger than the threshold area, climatic variables are more important. Our final selection of the environmental variables follows these findings.

This analysis demonstrates a point which forms the basis of this thesis, that is the selection of the environmental variables for predicting forest response to global warming must be related to the study area. In other words, the selection of the suitable environmental variables corresponding to the scale is important in the construction of the predictive model.

These findings challenge the application of local-scale forest models (Gap models) in a regional scale study. These findings also justify the importance of developing a regional-scale model in a regional-scale study.

As we found that October temperature is the most critical environmental variable. We can understand this better after the analysis of variables' contribution to the reduction of forest distribution entropy when changing the sampling area. October temperature is the most critical environmental variable since it removed the most entropy among all the variables. It is the best predictor of the distribution of all the vegetation types in the study area in general. From *Figure 4-10*, we can see that October temperature is not the best predictor of the vegetation types in the southern study area, where the vegetation types are prairie and broad-leaf forest. However, October is the best predictor of the vegetation types in the northern study area, where the vegetation types are tundra, the transition zone between tundra and boreal forest, and boreal forest. In average, October temperature is the best predictor of the vegetation types in the study area. The fact that October temperature is the most critical environmental variable suggests that the length of growing season is more important than any of the variables actually used.

4.2 Construction of KBFM's Knowledge Base.

The construction of the knowledge base is a process of finding the relationships between forest distribution and environmental variables, and computerizing these relationships. First of all, we have to decide which forest groups we are going to use.

4.2.1 Selection of the Group of Forest Types.

As mentioned in chapter 3, we have two ways (or levels) to group the forest types. One is more detailed forest types and the other is less detailed forest types. The reasons for using two groups of forest types are:

(1) We have an option to choose one of them such that we can construct a *KBFM* with higher prediction quality.

(2) We will explore if it is possible to have a *KBFM* which predicts the detailed forest distribution, and at the same time with higher prediction quality.

We now want to decide which level of forest group we will finally use to construct the knowledge base and the predictive model.

Again, *PEGASE* is used to make the decision. We ran *PEGASE* for these two group of forest types individually using the selected six environmental variables. The outcomes are shown as *Table 4-5*.

Table 4-5. The outcomes of running *PEGASE* for two groups of forest types.

Fores Type Grouping	Forest Types	Number of TSS	Initial Entropy	Final Entropy	Entropy Removed (%)
Group A	VE1, VE2, VE3, VE4, VE5, VE6, VE7, VE8, VE9, V10	107	1.985	0.807	59.4
Group B	V1, V6, V7, V8, V9, V10	76	1.415	0.483	69.1

Note: TSS stands for the terminal subset which has been explained in Chapter 3.

From *Table 4-5*, we find that about 10 percent more of entropy is removed when we use 6 classes of forest types (*Group B*) rather than 10 classes of forest types (*Group A*). The higher the amount of entropy removed, the more knowledge we get from the environmental variables about the forest distribution, and the higher the quality of the prediction. This means we have higher probability to do better prediction using the six selected environmental variables to predict the forest types in *Group B* than using the same variables to predict forest types in *Group A*. This conclusion is logical because we have a higher reliability to predict more general forest

types than to predict more detailed forest types using the same environmental variables. The prediction model will be constructed using according to group B for the following two reasons:

(1) We have higher reliability to predict forest types of *Group B* than forest types in *Group A* since 10% more of entropy is removed when we use the forest types in *Group B* than in *Group A;*

(2) The knowledge base will be much simpler when we use forest types in *Group B* than *Group A* since the former has only 76 terminal subsets (*TSS*), and the letter has 107 *TSS*. As mentioned in chapter 3, *TSS* are characterized by: (a) a specific combination of classes of environmental variables, and (b) a tendency to predict a particular vegetation type.

The disadvantage for choosing Group B is that our prediction is less detailed in terms of vegetation.

The detailed information of the outcomes of running *PEGASE* for the forest types in Group B are shown in *Table 1-6*.

Table 4-6. The outcomes of *PEGASE* analysis for Group B forest types

Information Items	Values
Initial Entropy	1.415
Final Entropy	0.438
Negentropy	0.977
Adjusted Negentropy	0.987
Entropy Removed (%)	69.1
Adjusted Entropy Removed (%)	69.8
Number of TSS	76

4.2.2. The Knowledge Base of the Relationship between Forest Distribution and Environmental Variables

Having decided which environmental variables and which forest type class we will use, we run *PEGASE* again with the selected environmental variables and the selected forest types. The final outcome is a hierarchical and divisive tree with a number of terminal subsets (*TSS*). Each *TSS* is differentiated from others by a specific combination of the classes of environmental variables. For example, *TSS30* in *Table 4-7* is composed of 70 cells. Among these 70 cells, 59 cells are currently occupied by boreal forest, 10 cells are currently taken by broad-leaf dominated forest (*V6*) and 1 cell is openland (*V7*). *TSS30* was differentiated by *TO6* (October temperature 4.0-4.99 ^{0}C) and *AP8* (April temperature 0.0-

0.99°C) and *UL8*(July temperature 18.0-18.99°C). *TSS30* can be interpreted as this: if a cell has a combination of the classes of environmental variables such as *TO6+AP8+UL8*, this cell has 84.3% probability to be boreal forest (*V1*), 14.3% broad-leaf forest (*V6*), and 1.4% open land(*V7*). From this example, we can see that each *TSS*, in fact, provides a piece of knowledge which tells the relationship between forest and environmental variables. All the *TSS* obtained from *PEGASE* constitute *KBFM*'s knowledge base. We obtained a total of 76 *TSS* from the *PEGASE* analysis and all these *TSS* are listed in *Table 4-7*. Therefore, we obtained 76 pieces of knowledge which together defined the relationship between forest distribution and the critical environmental variables in study area.

From *Table 4-7*, we see that some *TSS* can be specified only by one environmental variable. For example, *TSS1* (V8 91%, V9 9%)is differentiated by *TO1* (October temperature < 0.0°C). In this case, the relationship between forest and environmental variables is very simple and it means we can predict forest (V8 and V9) with one environmental variable. However, most of the other *TSS* are specified by a combination of three to five environmental variables; this means that the relationships between forest and environmental variables are far more complicated.

Table 4-7. The knowledge base of the KBFM

TSS	Combinations of Classes of Environmental Variables	Number of Cells	Predictive Probability of Forest Types
TSS1	TO1	1317	V8(91%)+V9(9%)
TSS4	TO2+JT2	249	V8(96%)+V1(4%)
TSS5	TO2+JT4	40	V1(88%)+V6(12%)
TSS15	TO2+JT3+RO2	50	V8(52%)+V1(42%)+V7(6%)
TSS16	TO2+JT3+RO4	101	V1(58%)+V8(42%)
TSS17	TO2+JT3+RO5	243	V1(72%)+V8(27%)+V7(1%)
TSS18	TO2+JT3+RO6	121	V1(98%)+V8(2%)
TSS19	TO2+JT3+RO7	60	V1(97%)+V8(2%)+V7(1%)
TSS20	TO2+JT3+RO8	6	V1(100%)
TSS48	TO2+JT3+RO3+AP2	24	V8(67%)+V1(21%)+V7(12%)
TSS49	TO2+JT3+RO3+AP3	61	V1(82%)+V7(12%)+V7(6%)
TSS2	TO3	350	V1(95%)+V7(5%)
TSS3	TO4	443	V1(95%)+V7(5%)
TSS6	TO5+ST1	86	V1(95%)+V6(4%)+V10(1%)
TSS7	TO5+ST2	117	V1(81%)+V7(14%)+V6(5%)
TSS8	TO5+ST6	153	V1(95%)+V6(3%)+V7(2%)
TSS9	TO5+ST7	28	V1(100%)
TSS21	TO5+ST3+AP4	18	V1(78%)+V6(22%)
TSS22	TO5+ST3+AP5	52	V1(96%)+V7(4%)
TSS23	TO5+ST3+AP6	8	V1(100%)
TSS24	TO5+ST3+AP8	19	V1(58%)+V6(37%)+V7(5%)
TSS25	TO5+ST4+JT5	66	V1(76%)+V7(24%)
TSS26	TO5+ST4+JT6	58	V1(66%)+V7(44%)
TSS27	TO5+ST4+JT7	87	V7(46%)+V1(40%)+V6(24%)

TSS10	TO6+AP7	71	V1(82%)+V7(13%)+V6(5%)
TSS11	TO6+A10	104	V6(65%)+V10(35%)
TSS28	TO6+AP8+UL6	2	V6(100%)
TSS29	TO6+AP8+UL7	28	V1(50%)+V6(25%)+V7(25%)
TSS30	TO6+AP8+UL8	70	V1(84%)+V6(26%)
TSS31	TO6+AP9+ST1	16	V6(44%)+V7(25%)+V10(19%)
TSS32	TO6+AP9+ST3	26	V6(62%)+V1(27%)+V7(8%)
TSS33	TO6+AP9+ST4	81	V1(73%)+V6(20%)+V7(7%)
TSS34	TO6+AP9+ST5	25	V1(68%)+V6(28%)+V7(4%)
TSS35	TO6+AP9+ST6	68	V1(97%)+V7(3%)
TSS50	TO6+AP9+ST2+RO1	1	V6(100%)
TSS51	TO6+AP9+ST2+RO2	16	V6(82%)+V118%)
TSS52	TO6+AP9+ST2+RO3	41	V1(54%)+V6(46%)
TSS53	TO6+AP9+ST2+RO4	71	V6(52%)+V1(38%)+V7(7%)
TSS54	TO6+AP9+ST2+RO6	31	V6(84%)+V1(10%)+V10(6%)
TSS55	TO6+AP9+ST2+RO7	19	V6(95%)+V1(5%)
TSS72	TO6+AP9+ST2+RO5+JT8	11	V6(82%)+V1(9%)+V10(9%)
TSS73	TO6+AP9+ST2+RO5+JT9	51	V6(75%)+V7(18%)+V10(6%)
TSS74	TO6+AP9+ST2+RO5+J10	10	V6(50%)+V10(50%)

TSS12	TO7+ST4	76	V1(57%)+V6(42%)+V10(1%)
TSS13	TO7+ST5	7	V1(57%)+V6(29%)+V10(14%)
TSS36	TO7+ST1+UL8	60	V6(73%)+V1(20%)+V10(7%)
TSS37	TO7+ST1+JT8	5	V6(100%)
TSS38	TO7+ST2+J11	67	V10(64%)+V6(38%)
TSS39	TO7+ST2+J12	15	V10(80%)+V6(20%)
TSS40	TO7+ST3+AP9	13	V10(54%)+V6(31%)+V1(15%)
TSS41	TO7+ST3+A10	58	V6(59%)+V10(17%)+V1(17%)
TSS42	TO7+ST3+A11	19	V10(90%)+V6(5%)+V7(5%)
TSS43	TO7+ST6+AP9	82	V1(99%)+V7(1%)
TSS56	TO7+ST1+UL9+RO3	4	V10(100%)
TSS57	TO7+ST1+UL9+RO4	24	V10(71%)+V6(29%)
TSS58	TO7+ST1+UL9+RO5	16	V6(56%)+V10(44%)
TSS59	TO7+ST1+UL9+RO6	6	V6(67%)+V10(33%)
TSS60	TO7+ST1+UL9+RO7	18	V6(83%)+V1(11%)+V10(6%)
TSS61	TO7+ST1+UL9+RO8	3	V6(67%)+V1(33%)
TSS62	TO7+ST2+JT9+UL8	112	V6(83%)+V1(12%)+V10(3%)
TSS63	TO7+ST2+JT9+UL9	31	V6(65%)+V10(32%)+V7(3%)
TSS64	TO7+ST2+J10+RO2	1	V10(100%)
TSS65	TO7+ST2+J10+RO3	6	V10(100%)
TSS66	TO7+ST2+J10+RO4	79	V10(79%)+V6(21%)
TSS67	TO7+ST2+J10+RO6	41	V6(63%)+V10(37%)
TSS68	TO7+ST2+J10+RO7	22	V6(96%)+V10(4%)
TSS69	TO7+ST2+J10+RO8	3	V6(100%)
TSS70	TO7+ST6+A10+JT9	21	V1(100%)
TSS71	TO7+ST6+A10+J10	61	V1(72%)+V6(28%)
TSS75	TO7+ST2+J10+RO7+UL8	33	V6(82%)+V10(9%)+V7(9%)
TSS76	TO7+ST2+J10+RO5+UL9	25	V6(56%)+V10(44%)
TSS14	TO8+A10	28	V6(50%)+V10(25%)+V1(25%)
TSS44	TO8+A11+ST1	44	V6(82%)+V10(18%)
TSS45	TO8+A11+ST2	132	V10(67%)+V6(33%)
TSS46	TO8+A11+ST3	77	V10(94%)+V6(6%)
TSS47	TO8+A11+ST4	4	V6(100%)
*	TO9+(<= RO6)	None	V10(100%)
*	TO9+(>= RO7)	None	V6 (100%)

* *Knowledge obtained from the south of the research area*

All combinations of environmental classes come from the analysis of the present forest distribution under current climate conditions. When this knowledge base is used to predict future forest

distribution, a problem may arise if the combination of future climatic variables cannot be found under the current climate. This is due to the fact that global warming causes seasonal temperature pattern change (e.g. winter temperature usually increases higher than summer temperature in high latitude area). The solution consists in choosing the combination of current climatic variables which is the closest to the future combination. This will, to some extent, lower the quality of prediction. However, this problem is unavoidable in all empirical models. This is also the difficulty in predicting forest response to global warming.

Considering the fact that this knowledge base is used for predicting the forest response to global warming, we will have a problem if we only use these 76 pieces of knowledge. We are unable to predict the future forest in the southern research area because the temperature in this area is predicted to increase to a level which does not exist so far in the research area. Therefore, none of the 76 pieces of knowledge will be applicable. The solution to this problem consists in finding more knowledge from the warmer area of North Dakota and Minnesota in the United State (National Atlas - Forest Types 1967, Climate Atlas of North and Central America 1979) and add them to the knowledge base. We would then have to use a *PEGASE* analysis in order to obtain the good quality "knowledge" about the forest distribution in the areas of North

Dakota and Minnesota. To obtain these data and redo this analysis is beyond the scope of this study. Because of the simplicity of the relationship between forest and environmental variables in the North Dakota and Minnesota areas, we will instead use a simple method of map overlay to specify the relationship between forest and climatic variable. We did not include the soil texture in map overlap analysis because of two reasons:

(1) The data are not available;
(2) More important, the boundary between the two vegetation types in that area, the prairie and the broad-leaf forest, is mainly controlled by the annual precipitation, which is very clear when we overlay the vegetation map and the annual precipitation map. This is also the case in the south of our research area.

Two pieces of knowledge were created by the map overlap method. These two pieces of "knowledge" are:

(1) If the October temperature is greater than or equal to 7.00 degree centigrade (TO9) and the annual precipitation is greater than 520.0 millimetre, the forest type will be broad-leaf forest with 100% probability;

(2) If the October temperature is greater than or equal to 7.00 degree centigrade (TO9) and the annual precipitation is less than

520.0 millimetre, the vegetation type will be prairie with 100% probability.

In the *KBFM*, October temperature is the most important variable controlling forest distribution (*Table 4-8*). A specific category of October temperature corresponds to one or more forest types. For example, when October temperature falls into the category of *TO1* (October temperature <0.0°C), the only possible forest types are tundra and the transition zone; when October temperature falls into the categories of *TO3* (October temperature >=1.00°C and <1.99°C) and *TO4* (October temperature >=2.00°C and <2.99°C), the only possible forest type is boreal forest (*V1*) and some openland (*V7*). For the categories of *TO1*, *TO3*, and *TO4*, we can make good prediction of forest distribution just by knowing October temperature. However, a good prediction becomes difficult for the categories of *TO2* (October temperature >=0.00°C and <0.99°C), *TO5* (October temperature >=3.00°C and <3.99°C), *TO6* (October temperature >=4.00°C and <4.99°C), *TO7* (October temperature >=5.00°C and <5.99°C), and *TO8* (October temperature >=6.00°C and <6.99°C) (*Table 4-8*). In these cases, we need additional information. The uncertainty of the categories of *TO2*, *TO5*, *TO6*, *TO7*, and *TO8* can be greatly reduced by adding other environmental variables. As *Table 4-8* shows the category *TO2*, for example, *V1* and *V8* can be well differentiated by adding the variable January temperature. Under the condition of *TO2*

+ *JT2 (TSS4)*, we have a 96% probability of predicting the transition zone between boreal forest and tundra (*V8*). We have a 88% probability of predicting V1 under the condition of *TO2 + JT4 (TSS5)*. For the category of *TO2 + JT3*, there is still some uncertainty, which is reduced by adding annual precipitation. When the annual precipitation is higher (*TSS16 - TSS20*), it favors boreal forest (*V1*); when the annual precipitation is lower (*TSS15, TSS48*), it favors the transition zone between boreal forest and tundra (*V8*).

Table 4-8 October temperature and its prediction of forest types

October temperature	Number of cells	Forest types and probability
TO1	1317	V8(91%),V9(9%)
TO2	955	V1(56%),V8(41%),V7(2%),V6(1%)
TO3	350	V1(95%),V7(5%)
TO4	443	V1(95%),V7(5%)
TO5	692	V1(80%),V7(14%),V6(6%)
TO6	743	V1(46%),V6(41%),V6(7%),V7(6%)
TO7	908	V6(47%),V10(26%),V1(26%),V7(1%)
TO8	285	V10(61%),V6(36%),V1(3%)

In the north of the research area, soil texture did not contribute much to the reduction of the entropy and this fact can be found in the categories of *TO1, TO2, TO3 and TO4*. However; soil texture becomes very important in the south and this fact can be found in the categories of *TO5, TO6, TO7, and TO8*. Comparing *TSS31, TSS32, TSS33, TSS34,* and *TSS35* in *Table 4-7*, for example, a combination of *TO6 + AP9* plus different soil texture will predict different forest. Sand or sandy loam soil texture (*TO6+AP9+ST1*) predicts broad-leaf forest (44%) and prairie (19%). Clay soil texture (*TO6+AP9+ST3*) predicts broad-leaf forest (62%) and boreal forest (27%). Loam soil texture (*TO6+AP9+ST2*) also predicts broad leaf forest (66%) and boreal forest (23%). Other soil textures (*ST4, ST5 and ST6*) mainly predict boreal forest (68-97%).

Precipitation is an important variable differentiating prairie from broad-leaf forest. Comparing *TSS56, TSS57, TSS58, TSS59, TSS60* and *TSS61* in *Table 4-7*, when the annual precipitation is less than 480mm (*RO3, RO4*), we have very high probability of predicting prairie *(TSS56 100% and TSS57 71%)*; however, when the precipitation is higher than 480 mm (*RO5, RO6, RO7, and RO8*), we have higher probability of predicting broad-leaf forest (*TSS58 56%, TSS59 67%, TSS60 83%, TSS 61%*). *TSS64, TSS65, TSS66, TSS67* and *TSS68* have exactly the same functions. Precipitation is also important in the two pieces of knowledge from the area of North Dakota and

Minnesota. The slight difference is that 520mm annual precipitation is associated with the boundary of broad-leaf forest and prairie because higher temperature favors higher evaporation.

4.2.3 Construction of the Computerized Knowledge Base of KBFM.

Computerization of the knowledge base is a process of translating each *TSS* in *Table* 4-7 into a knowledge in a condition-action statement. A condition-action statement contains a logical condition which controls the program to go to one specific branch according to the the boolean value of the logical condition. Typically, it is like:

```
if ( logical condition 1 = TRUE)
     forest type = a;
else if (logical condition 2 = TRUE)
     forest type = b;
.......
.......
else
     forest type = n;
```

The logical condition here is the combination of boolean equations of the critical environmental variables. For example, the knowledge provided by TSS12 (TO7+ST4 predicts V1(57%)+V6(42%)+V10(1%)) in *Table 4-7* can be computerized using programming language *C* as:

```
if((OctTem>=5.0)&&(OctTem<6.0)&&(SoilTex==ST4))
{
     if((r=rand()%10)<6)
          return 1;
     else
          return 6;
}
```

This section of a program means that if a cell satisfies the following conditions: *October Temperature >= 5.0 and October Temperature < 6.0* (October temperature is greater than or equal to 5.0°C and less than 6.0°C) and *Soil Texture ==ST4* (soil texture is *Mesic Woody Forest*), this cell has a 60% probability to be boreal forest (*V1*) and a 40% probability to be broad-leaf forest (*V6*). The prairie is ignored since the actual probability of being a prairie is only 1%. Here, a random number generator function in programming language *C*, ***rand()***, is introduced to determine the forest type of this cell in such a way: the value of r (*r = rand() mod 10*) is obtained by taking the module of a random integer number by 10, which always yields an integer value with a range of 0 to 9. If it is a value less than 6, then it returns forest type V1 represented

by integer 1, otherwise it returns forest type V6 represented by integer 6.

TSS1 is an exceptional case when it was computerized. *TSS1* says that we have approximately a 90% probability to predict the transition zone between tundra and boreal forest (V8) and a 10% probability to predict the tundra (V9) when the October temperature of a given cell is lower than 0.0^{0}C. If we use the method described above to computerize it, we will not be able to predict a spatially continuous tundra but the predicted tundra will evenly scatter inside the transition zone between tundra and boreal forest (V8). The solution is that we do not use a random number generator function to decide a cell to be tundra (V9) or the transition zone between tundra and boreal forest (V8) instead we use a cell's October temperature to predict. If the October temperature of a cell is lower than or equal to -0.9^{0}C, this cell will be tundra (V9) and if the October temperature of a cell is lower than 0.0^{0}C and higher than -0.9^{0}C, this cell will be the transition zone between tundra and boreal forest (V8). The value of -0.9^{0}C was obtained by spatial boundary matching.

This exceptional computerization of *TSS1* makes the prediction of present tundra distribution reasonable. However, it will not be applied to the prediction of forest response to global warming

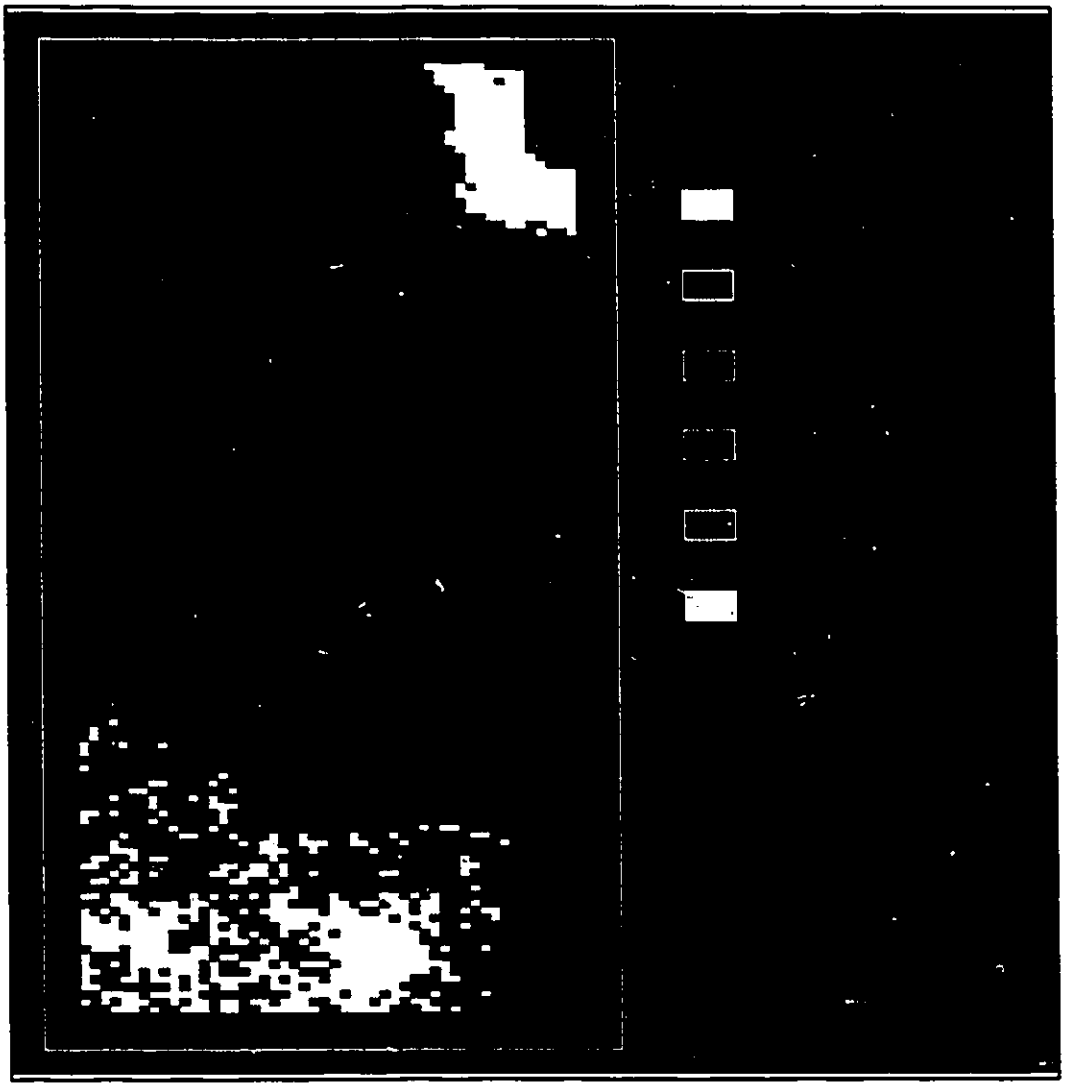
because there will be no any cell in which the October temperature is lower than 0.0°C (TO1) after global warming.

4.3 Validity Test of the KBFM

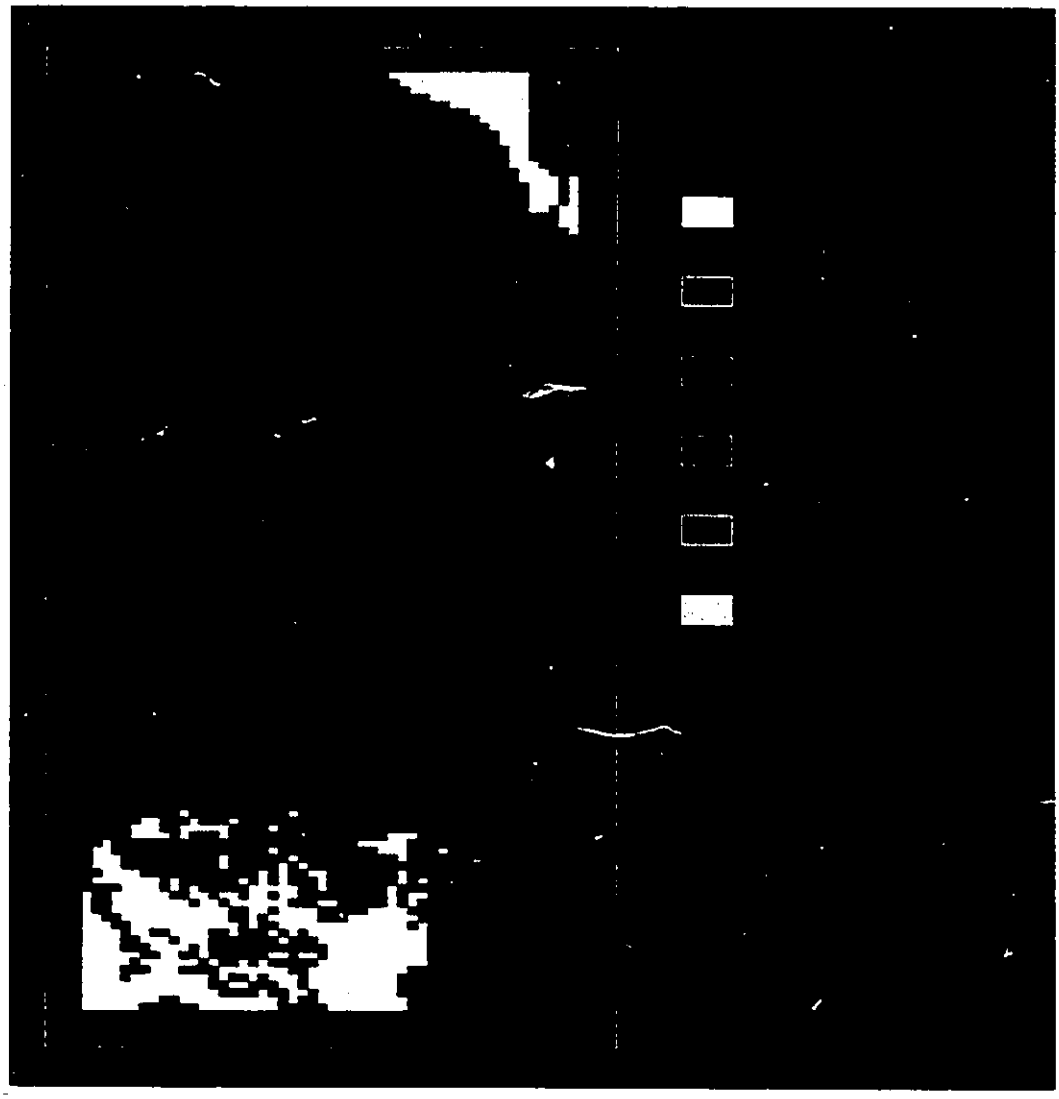
It is logical to assume that a model would not be reliable if it cannot predict the present forest distribution well. Under this assumption, we can test the validity of the *KBFM* by comparing present forest distribution with the predicted forest distribution under the present climatic condition (a 0.0 degree centigrade temperature change and 0.0 mm precipitation change).

Comparing the predictive forest distribution (*Map 4-1*) with the current forest distribution (*Map 4-2*), we find the boundaries between forest types match very well. We also find that the total surface area of each forest types agree very well(*Table 4-9*). The areas of the two forest distributions can be compared using *cell difference ratio*, which is the difference in the number of cells between predictive forest and current forest divided by the number of cell of specific forest type of present forest distribution. The predicted broad-leaf forest, prairie, and transition zone are very close to the current forest. Their cell difference ratios are as low as 1.9%, 3.0%, and 3.8% respectively. Boreal forest has a relatively higher cell difference rate -6.8%, but it is still

Map 4-1 The predictive forest distribution under the condition of zero climate change.



Map 4-2. The current forest distribution



acceptable. Tundra has the highest difference rate 64.0%. This is due to the fact we only use one environmental variable, October temperature, to predict tundra in our model and to the fact that the total number of tundra cells is relatively small.

All these figures show that the *KBFM* can predict the area of the prairie, broad-leaf forest, boreal forest, and the transition zone between boreal forest and tundra quite well. This predictive outcomes demonstrates the validity of this *KBFM*.

Table 4-9. Comparison of the areas of the present and the predicted forest

Forest types	Present Forest (Cell Number)	Predicted Forest (Cell Number)	Difference (Cell Number)	Cell Difference Ratio
Boreal forest	2425	2261	-164	-6.8%
Broad-leaf forest	893	910	17	1.9%
Tundra	114	187	73	64.0%
Transition zone	1597	1657	60	3.8%
Prairie	467	481	14	3.0%

Note: The positive signs means the predictive forest area is larger than the current forest area and negative sign means reverse.

Chapter 5. The Prediction of Forest Responses to Global Warming

5.1 GCMs and Future Climatic Data.

In order to predict future forest distribution as a response to global warming, we need to develop scenarios of future climatic change. Future climatic data can be drawn from the predictions of GCMs. As mentioned in Chapter 1, there are several versions of GCMs.

In this research, the future climatic data come from the second generation of the Canadian Climatic Centre GCM (*CCC* GCM) provided by the Canadian Climatic Centre. We use the *CCC* GCM's prediction for several specific reasons:

(1) It has a higher grid point resolution than many other GCMs with a transform grid of 3.75° longitude by 3.75° latitude. Since this research is carried out at regional scale, higher resolution is highly desirable and will provide better estimates of the predicted vegetation. However, even this model has a quite coarse resolution for this study.

(2) The model accounts for full diurnal and annual cycles.

(3) Ocean and sea-ice treatments involving the specification of ocean transports permits a good simulation of $1*CO_2$ (present-day) ocean temperature distribution and ice boundary.

(4) It has a modified treatment of land surface processes and hydrology which is somewhat more sophisticated than the usual "bucket" treatment.

(5) There is a parameterization of cloud optical properties feedback, which is important on cloud properties. This is a persistent problem in GCMs (*Boer et al. 1991*).

The basic average features of the simulated climate change due to green-house gases by the CCC GCM are:

(1) A globally average surface temperature increase of 3.5 degrees with the increase over the land larger than that over ice-free ocean;

(2) A precipitation and evaporation increase of 4% over the globe with the increase in precipitation larger over the ocean and the increase in evaporation larger over the land;

(3) An average decrease in soil moisture of the order of 6.6% with

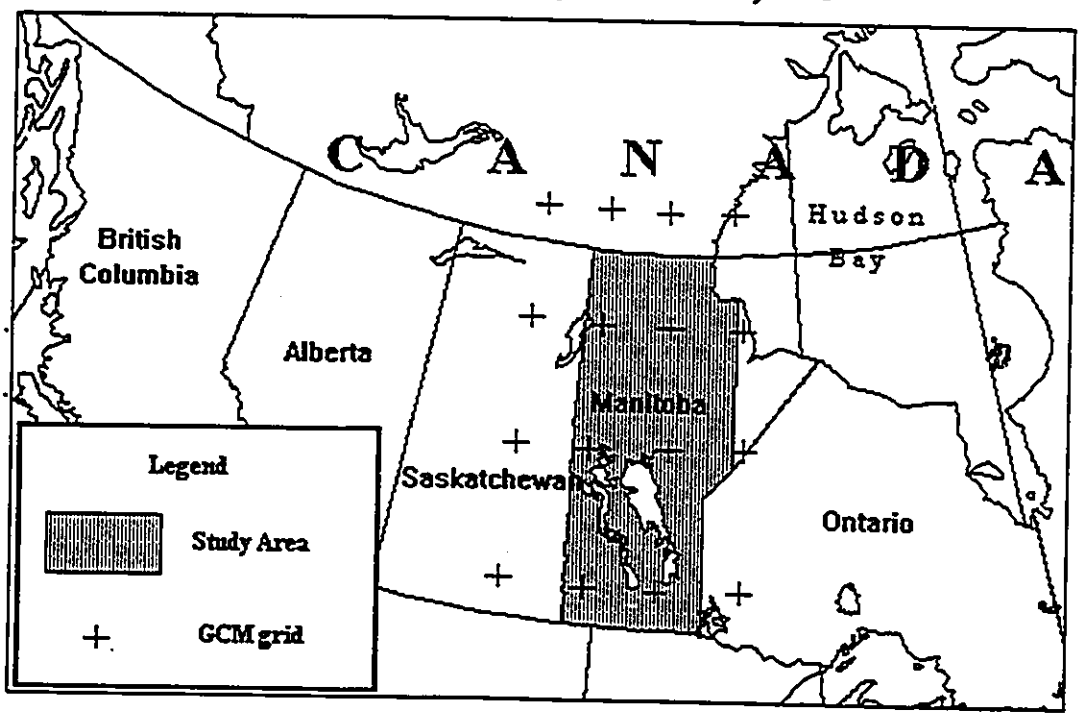
larger values occurring in northern hemisphere during summer;

(4) A decrease in cloud cover of about 2.2%;

(5) A major decrease of the mass of sea-ice amounting to 66%, and

(6) Marked changes in other quantities in the polar regions (*Boer et al. 1991*).

Map 5-1 The location of the CCC GCM grid points and study area



Source: Application of the Canadian Climate Centre General Circulation Model output for rigional impact studies, Guidelines for Users (Cohen 1991)

Six *CCC* GCM grid points fall within the research area and 14 others surround this study area. *Map 5-1* shows the location of all the grid points within the research area and 10 surrounding grid points. Four surrounding grid points in the south of the study area cannot be shown in *Map 5-1* because they are out of the map. For simplicity, we give each grid point within the study area a short name according to its spatial location in the study area. The two grid points in the south are called the southwest (SW) and southeast (SE). The two grid points in the north are called the northwest (NW) and northeast (NE). The rest two grid points in the middle are called the central-west (CW) and central-east (CE). The magnitudes of temperature increase is obtained by subtracting the predicted climate of $2*CO^2$ with the climate of $1*CO^2$. The climate change for the 6 grid points are shown in *Table 5-1*. Adding these changes to the current climate data, we get the future climate data for these 6 grids. We can see a pattern of the climate change in our research area from *Table 5-1*. The basic changes predicted for these 6 grids are:

(1) Precipitation increases for all these 6 grids. In the north, the precipitation increases about 100mm, which is higher than that in the south with an increase about 30mm;

Table 5-1. Climate change for the six grids in the study area

Grid Geographic Location	January Tem. Increase	April Tem. Increase	July Tem. Increase	October Tem. Increase	Annual Precip. Increase
50.1°N, 101.3°W (Southwest)	7.2°C	8.5°C	4.8°C	3.1°C	6.8
50.1°N, 97.5°W (Southeast)	7.1°C	6.4°C	4.3°C	2.9°C	33.9
53.8°N, 101.3°W (Central-west)	4.3°C	2.6°C	4.2°C	2.2°C	29.5
53.8°N, 97.5°W (Central-east)	4.9°C	2.1°C	4.5°C	1.9°C	16
57.5°N, 101.3°W (Northwest)	3.1°C	0.9°C	2.9°C	1.6°C	111.6
57.5°N, 97.5°W (Northeast)	3.5°C	1.4°C	3.3°C	1.5°C	118.7
Average	5.0°C	3.7°C	4.0°C	2.2°C	52.8

Note: Tem. stands for temperature. Precip. stands for precipitation. The unit for precipitation change is millimetre.

(2) Temperature increases for all these 6 grids. It is clear that temperature increases more to the south than in the north. For example, January temperature increases up to 7 degree centigrade in the south; however, it increases only 3 to 3.5 degree centigrade in the north;

(3) January and July temperature increases are higher than those of April and October temperature. For example, the average temperature increase for January is 5 degree centigrade; however, the average temperature increase for October is 2.2 degree centigrade.

The climatic data predicted by *CCC* GCM are questionable if we just consider those data in our research area in isolation. For those 6 grid points in the study area (*Map 5-1*), April temperature, for example, increases by 8.5°C in the Southwest grid point (50.1N, 101.3W); however, it only increase 0.9°C in the Northwest grid point (57.5N, 101.3W). This is due to the fact that GCMs give a relatively satisfactory picture of future climate at global scale; however, they have much less certainty on a regional scale (*Robock et al. 1993, Hansen et al. 1991*). Because of the limited resolution of GCMs, we cannot get good quality future climate data at regional scale. In this research, we deal with this by choosing three climate scenarios about the way in which global change in regionally distributed:

(1) Scenario 1. We assume that climate will change uniformly in our research area. We use the average climate change data of the six grid points (*Table 5-1*) as the climatic change data for the entire research area, and we get future climatic data by adding the average climatic change data (both temperature and precipitation) to the current climatic data. The advantage of this scenario is that it removes the large variation among grid points. The disadvantage is that it also removes the pattern of climatic change. Moreover, a uniform climate change over the entire study area seems unrealistic.

(2) Scenario 2. We assume climate changes homogeneously at the same latitude. This means that areas at the same latitude undergo an increase of the same magnitude for temperature and precipitation. Under this assumption, we use the average of the climate change data of the two grid points in the same latitude to represent the climate change at this latitude (*Table 5-2*). Future climatic data for all the areas can be obtained by interpolation based on the data in *Table 5-2*. The advantages of this scenario are: (a) it keeps the South-North climatic change pattern obtained from *CCC* GCM prediction and (b) it also, to some extent, buffers the great variations among grid points. The disadvantage is that it ignores any East-West climate change pattern.

Table 5-2. Climatic change in the study area (Scenario 2)

Geographic Location	January Tem. Increase	April Tem. Increase	July Tem. Increase	October Tem. Increase	Annual Precip. Increase
46.39°N	8.9°C	11.7°C	3.6°C	3.1°C	34.4
50.10°N	7.1°C	7.4°C	4.5°C	3.0°C	20.3
53.81°N	4.6°C	2.4°C	4.4°C	2.1°C	22.8
57.52°N	3.3°C	1.2°C	2.6°C	1.6°C	111.6
61.23°N	5.0°C	1.8°C	4.0°C	1.5°C	110.6

***Note:** The data for the latitude of 46.39N come from four grid points which are not shown in Map 5-1. Tem. stands for temperature. Precip. stands for precipitation. The unit for precipitation change is millimetre.*

(3) Scenario 3. We assume the *CCC* GCM predicts future climate. We

use the climatic change data predicted by the CCC GCM for the six grid points in the research area and other 14 grids around the research area (*Table 5-3*). Future climatic data for all the research areas can be obtained by interpolation based on these 20 GCM grid points. The advantage is that the climate change pattern is well-preserved but the disadvantage is that the large variation of climatic change among grid points seems questionable.

Table 5-3. Climatic change in the study area (Scenario 3)

Grid Geographic Location	January Tem. Increase	April Tem. Increase	July Tem. Increase	October Tem. Increase	Annual Precip. Increase
46.39°N,105.0°W	8.9°C	11.2°C	4.1°C	3.3°C	80.1
46.39°N,101.25°W	8.9°C	12.2°C	3.6°C	3.2°C	71.0
46.39°N,97.5°W	8.9°C	11.3°C	3.7°C	3.1°C	7.9
46.39°N,93.75°W	8.1°C	5.8°C	3.5°C	2.7°C	9.9
50.1°N,105.0°W	8.2°C	9.2°C	4.3°C	3.1°C	21.1
50.1°N,101.25°W	7.2°C	8.5°C	4.8°C	3.1°C	6.8
50.1°N,97.5°W	7.1°C	6.8°C	4.3°C	2.9°C	33.9
50.1°N,93.75°W	6.3°C	2.9°C	4.0°C	2.9°C	82.7
53.81°N,105°W	5.1°C	2.8°C	3.6°C	2.5°C	47.2
53.81°N,101.25°W	4.3°C	2.6°C	4.2°C	2.2°C	29.5
53.81°N,97.5°W	4.9°C	2.7°C	4.5°C	1.9°C	15.7
53.81°N,93.75°W	5.5°C	5.6°C	4.3°C	1.8°C	-5.6
57.52°N,105.0°W	3.3°C	0.6°C	2.7°C	1.9°C	107.2
57.52°N,101.25°W	3.1°C	0.9°C	2.9°C	1.6°C	111.4
57.52°N,97.5°W	3.5°C	1.4°C	3.3°C	1.5°C	118.7
57.52°N,93.75°W	6.2°C	3.7°C	3.6°C	1.8°C	111.9
61.23°N,105.0°W	3.3°C	2.1°C	2.8°C	1.3°C	58.5
61.23°N,101.25°W	3.8°C	1.6°C	3.0°C	1.4°C	116.5
61.23°N,97.5°W	5.0°C	2.1°C	3.4°C	1.6°C	106.8
61.23°N,93.75°W	6.2°C	1.9°C	5.6°C	1.7°C	118.6

Note: The data for the latitude of 46.39N come from four grid points which are not shown in Map 5-1. Tem. stands for temperature and Precip. stands for precipitation. The unit for precipitation change is millimetre.

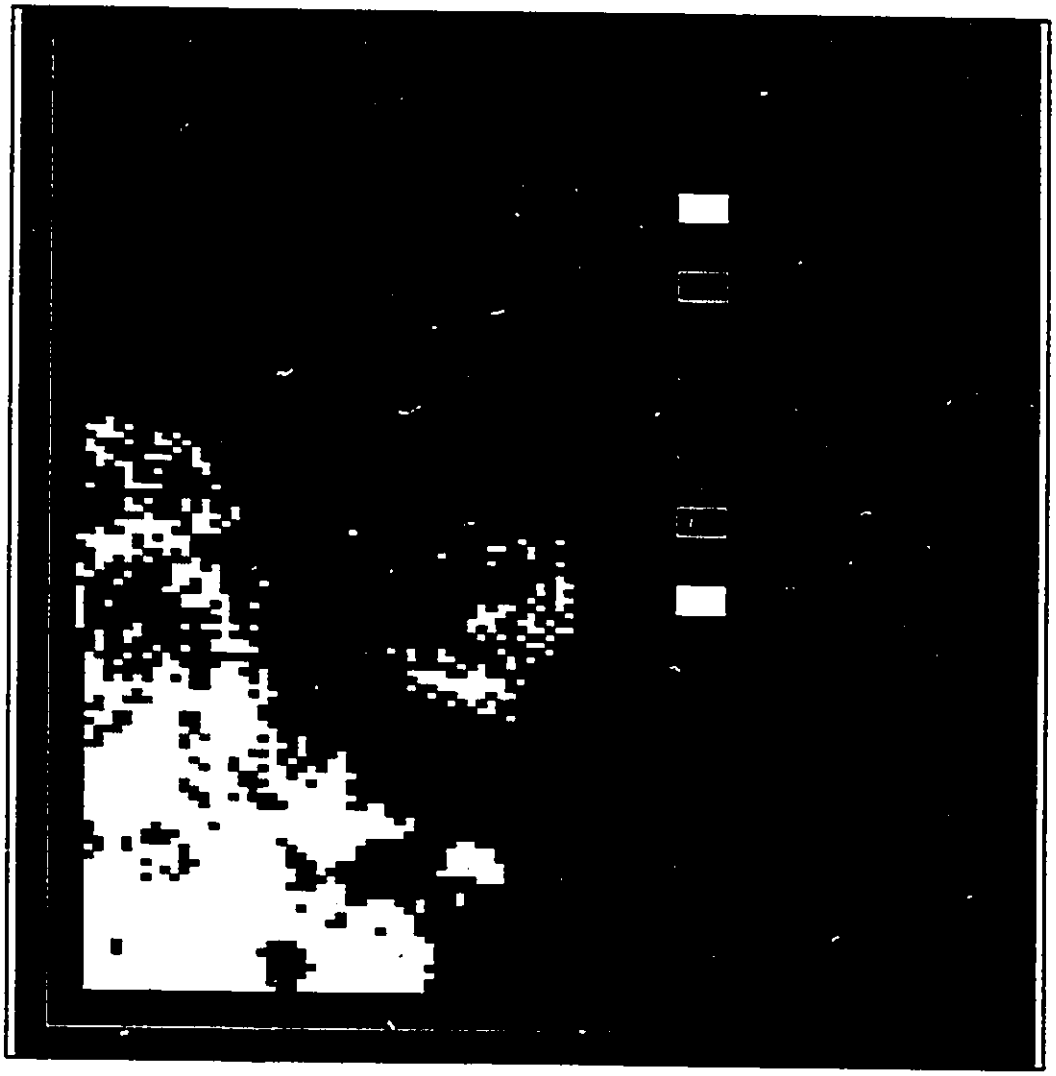
5.2 Prediction of forest distribution based on climate scenarios

Based on the predictions of *KBFM*, global warming will have a great impact on the forest distribution in the research area under all climate scenarios described above. The predictions of forest area change are given in *Table 5-4* and the future forest distribution after global warming are shown in *Map 5-2, Map 5-3, and Map 5-4* for scenario 1, scenario 2 and scenario 3 respectively

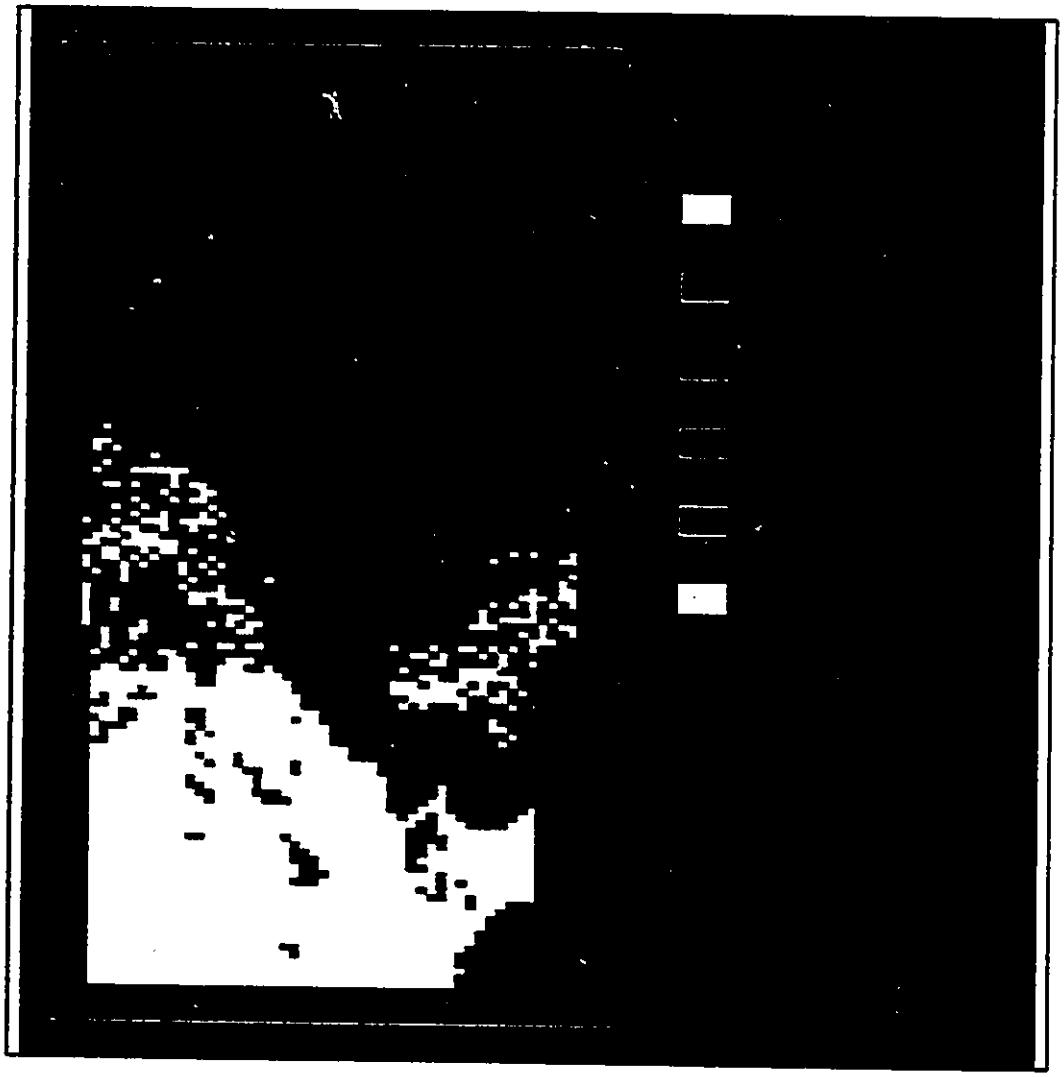
Table 5-4. Change of the area of the forest types for three climate scenarios

Forest Types	Current forest Cell No.	Scenario 1		Scenario 2		Scenario 3	
		Cell No.	Increase	Cell No.	Increase	Cell No.	Increase
Boreal Forest	2425	2954	21.8%	3111	28.3%	3122	28.7%
Broad leaf Forest	893	1158	29.7%	706	-20.9%	839	-6.0%
Transi-tion Zone	114	0	-100%	0	-100%	0	-100%
Tundra	1597	0	-100%	0	-100%	0	-100%
Prairie	467	1384	196.4%	1679	259.5%	1535	228.7%

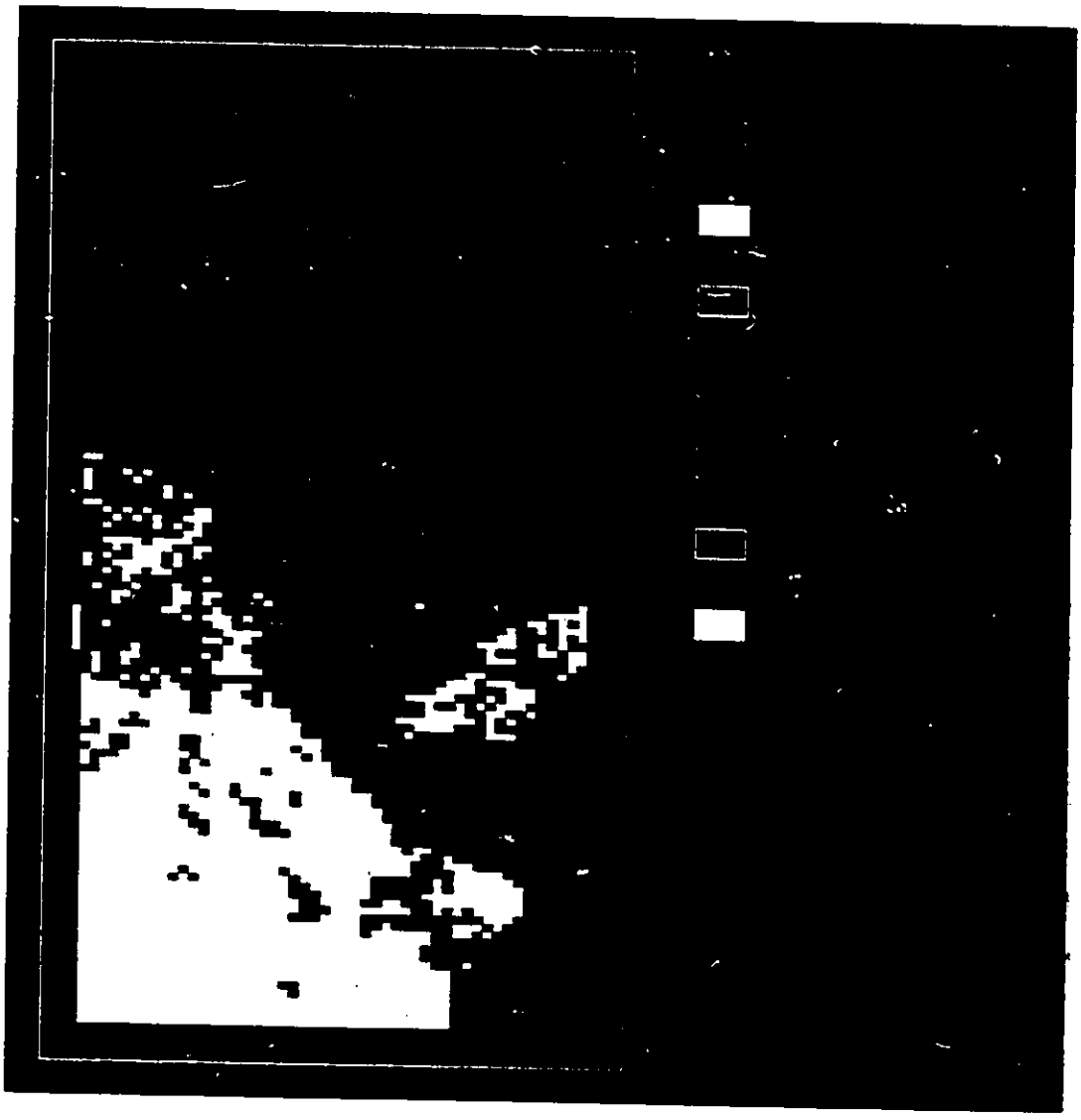
Map 5-2. Predictive forest distribution after global warming (Scenario 1)



Map 5-3. The predictive forest distribution after global warming (Scenario 2).



Map 5-4 The predictive forest distribution after global warming (Scenario 3).



(1) Under all the three climatic scenarios, the area of boreal forest will expand by about 21.8% to 28.7%. The area of prairie will also expand up to 196.4% to 259.5%. The area change of broad-leaf forest is different among the three climate scenarios. In scenario one, the area of broad leaf forest will increase by about 29.7% because of a high increase of annual precipitation. For scenario two and scenario three, its area will shrink 20.9% and 6.0% respectively.

(2) Under all the three climate scenarios, tundra and the transition zone between tundra and boreal forest will disappear in the research area after global warming. The area currently covered by tundra and the transition zone between tundra and boreal forest will be replaced by boreal forest because of the temperature increase.

(3) The northern border of boreal forest will move northward dramatically and it will go beyond the research area. The southern border also will move northward about 120 to 180 km.

(4) The southern part of the present-day boreal forest will be replaced by either broad-leaf forest or prairie depending on how the precipitation will change. If precipitation does not increase much, the southern part of boreal forest will be replaced by prairie as temperature increases. However, if precipitation

increases sufficiently, the southern part of boreal forest will be replaced by broad-leaf forest. The predictions of broad-leaf forest are different among the three climatic scenarios. In scenario one, the annual precipitation increases 52.2mm for the entire research area, and the southeast part of boreal forest will be replaced mainly by broad-leaf forest because of the relatively higher of annual precipitation. On the contrary, the southwest part of boreal forest is mainly replaced by prairie because the annual precipitation is lower. In scenario two, the southern part of present boreal forest will be mainly replaced by prairie because precipitation in the south of the research area will only increase by 20-23mm. Therefore, the environment will become relatively drier and the area of prairie will expand. Scenario three is intermediate between scenarios one and two.

The prediction of forest responses to climate change at a regional scale thus depends on the future climatic pattern predicted by GCMs. The reliability of the predicted future forest depends on two aspects: (a) the forest model itself; (b) the future climate data. So far, GCMs cannot provide satisfactory future climatic pattern for regional scale study. This will limit our ability to predict future forest distribution at regional scale.

Chapter 6. Discussion

From this study, we find that the quality of the prediction of forest response to climatic change at a regional scale is limited by the quality of future climatic prediction. At present, GCMs can not provide good quality future climatic data for regional scale studies because of the limitation of computational power. The improvement of the future climatic data will help the improvement of the precision of the prediction of forest response to global warming at regional scale.

Although GCMs cannot give satisfying future climatic prediction for a regional scale studies, people can still explore the potential impacts of global warming on forest at regional scale with a number of climatic scenarios and a good forest model. The *KBFM* is such a forest model.

It is interesting to compare the *KBFM's* predictions with other models' predictions. At present; however, there are some difficulties for the comparison because of following reasons:

(1) There is not other regional scale studies in the area of Manitoba;

(2) The use of different GCMs prediction in different studies makes the comparison difficult since the prediction of future forest or vegetation depends on the future climatic data.

The study done by Rizzo *et al. (1992)* is the most comparable to the *KBFM's* prediction among all the studies. As we discussed in Chapter 2, Rizzo *et al. (1992)* explored the sensitivity of Canada's ecosystem to climatic change. This study covered the province of Manitoba. However, the future climatic data they used were from *GISS* GCM prediction and the scale of the study was at a continental scale.

The prediction by Rizzo *et al. (1992)* is quite different from *KBFM's* prediction. In the province of Manitoba, Rizzo *et al. (1992)* predicted: (a) a much larger area of prairie than *KBFM's* prediction and it also predicted that prairie will occupy most area of Manitoba after global warming; (b) a disappearance of boreal forest and the area currently occupied by boreal forest will be replaced by prairie and the cool temperate forest (equivalent to the broad-leaf forest in *KBFM*); (c) a relative larger area of cool temperate forest. The common predictions between Rizzo *et al.'s* model and the *KBFM* are that both tundra and the transition zone between boreal forest and tundra will disappear in the province of Manitoba.

The great difference of the predictions between Rizzo *et al.'s* model and the *KBFM* is mainly caused by the uses of different GCMs' predictions. The *CCC* GCM's prediction used by *KBFM* predicted less temperature increase than the *GISS* GCM's prediction which was used by Rizzo *et al.'s* model (*Table 1-2*).

Comparing the predicted future vegetation distribution, we can see the advantage of the *KBFM* which gives much more detailed forest distribution influenced by soil texture pattern, precipitation and temperature distribution. However, Rizzo *et al.'s* model only gave a rough vegetation distribution and less detailed vegetation boundaries, which is less meaningful at a regional scale.

This *KBFM* is a new approach to regional scale modelling of forest response to climate change. This research shows it can be very well used for this research. It has two predictive functions: (1) It can predict the future forest distribution, and give a clear forest distribution pattern at a regional scale; (2) It can also predict the change of the area of forest types.

The *KBFM* is a different approach from gap models. However, it has some similarities with the empirical models. The *KBFM* and other empirical models assume that the present-day vegetation is equilibrium with the climate. They use the relationship between

present vegetation distribution and environmental variables, to predict future vegetation distribution. The future vegetation predicted by the *KBFM* and the empirical models are a kind of *"potential vegetation"* after global warming. This means that if global warming occurs in the coming 100 years, as the GCMs predicts, and then the climate stabilizes afterward, the future vegetation will eventually develop under the new climatic conditions given sufficient time for species migration and soil development.

The new aspects incorporated in this *KBFM* are:

(1) The *KBFM* uses *PEGASE,* an analytical method based on information theory, to find out the relationships between vegetation and environmental variables whereas other empirical models use spatial boundary matching methods and other statistical methods to characterize the relationships. The advantages using *PEGASE* are: (a) it can handle multiple environmental variables; (b) it can handle nominal, ordinal and numerical variables such that both climatic and non-climatic variables can be included; (c) it can effectively figure out the quality of the relationship between forest and environmental variables using the concept of entropy, negentropy, redundancy and probability; (d) the most important thing is that the relationships we obtained from *PEGASE* can be

computerized and form a knowledge base, which can be efficiently used for predicting future vegetation. However, the spatial boundary matching method used by other empirical models do not have these advantages. Moreover, other empirical models are hard to apply at a regional scale because the inclusion of both climatic and non-climatic variables makes the model construction complicated.

(2) The *KBFM* has been developed using a series of analytical processes including: (a) the selection of critical environmental variables; (b) the analysis of the impact of the scale on the selection of the critical environmental variables; (c) the flexibility of handling different levels of forest types generalization. These processes cannot be well done in other empirical models. However, these processes are helpful for the construction of a high quality model. They make it clear why we choose specific environmental variables, why we exclude others, and to what extent we lose the predictive ability when we exclude some environmental variables. Since we have many environmental variables which are potentially important to forest distribution at regional scale, we need these analytical processes.

(3) The *KBFM* uses a data base to store the information necessary for the prediction. In this model, the study area was divided into

approximately 6400 small cells (since there are some cells cover water body, the number of cells which are used for analysis are 5693), vegetation and forest information was organized cell by cell to form a data base. Using this data base and the knowledge base, we can conveniently and efficiently predict future forest for various climatic scenarios with the help of a computer.

Table 6-1. The comparison of the *KBFM* and other empirical models

Models	*KBFM*	Empirical Models
Method to get the relationship between vegetation and environmental variables	PEGASE	Boundary matching
The way to represent the relationships	Knowledge base	None
The way to store correlated data	Data base	Maps
The way to predict	Cell base prediction	Boundary base prediction
By-product of prediction	Changes of vegetation boundary and vegetation type areas	Changes of vegetation type areas
Vegetation predicted	Potential vegetation	Potential vegetation
Flexibility	Flexible to Predict various climatic scenarios	Difficult to predict various climatic scenarios

(4) The *KBFM* does not directly predict the change of the boundaries between forest types. The prediction unit is a cell. The change of area of a specific forest type is automatically obtained by mapping the cell system with the simulated vegetation types. More important, the forest boundaries show up automatically after all cells have been predicted.

From these analyses, it is clear that the *KBFM* is better than other empirical models to deal with forest response to global warming at a regional scale. In this application, the *KBFM* is used at a regional scale, but it does not mean that it can only be used at that scale. It can also be used at continental and global scales. When the *KBFM* is used at continental or global scales, it should use different environmental variables and a different size of cell.

The *KBFM* is not perfect at this point and it has room to be improved. First, the way *PEGASE* handles quantitative environmental variables is to convert them into qualitative data. For example, October temperature 1.00-1.99^{0}C is classified as TO3 in this study. People may also classify October temperature 1.50-2.49^{0}C as TO3. It means that some information carried by the quantitative environmental variable may be lost during this conversion. In other words, different classification of quantitative variables may possibly have slightly different outcomes of the critical

environmental variable selection and the final prediction. Secondly, the prediction of future forest by *KBFM* is based on the correlation of forest and environmental variables. The detailed mechanism is not very well explored in this study. It needs to be improved in the future.

Being a new and powerful model, the *KBFM* has a great potential to be improved and updated. For example, we can relate future forest distribution not only to the future environmental variables but also to the present forest distribution. This means that the present-day forest type of a cell can also be included in *KBFM* and performs an influence on its future forest type. It is also possible to divide the future 100 years into several intermediates (e.g. 20 years) and predict forest distribution changes progressively. Improving these two aspects will allow the *KBFM* to show the dynamic process of forest response to climatic change. Technically, there is no fundamental difficulty for these improvements, but it needs more work and more time.

References

Baudry, J. 1989: Interactions between agricultural and ecological systems at landscape level. *Agriculture, Ecosystems and Environment* **27**:119-130.

Bernabo, J.C. 1981: Quantitative estimates of temperature changes over the last 2700 years in Michigan based on pollen data. *Quaternary Research* **15**:143-159.

Boer, G.J. et al. 1991: Greenhouse gas induced climatic change simulated with the CCC Second Generation GCM. in Application of the Canadian CLimate Centre General Circulation Model Output for Regional Climate Impact Studies, Guidelines for Users(unpublished). Canadian Climate Centre, Atmosphere Environment Centre, Downsview, Ontario, Canada, M3h 5T4.

Bonan, G. B. 1990: The sensitivity of some high-latitude boreal forests to climatic parameters. *Climatic Change* **16**:9-29.

Botkin, D.B. *et al.* 1972: Some ecological consequences of a computer model of forest growth. *The Journal of Ecology* **60**:849-872.

Botkin, D.B. *et al.* 1992: "Projecting the effects of climate change on biological diversity in forests", in Peters, R.L and Lovejoy, T.E. (eds.), *Global Warming and Biological Diversity*, Yale University Press, New Heaven, **pp**277-296.

Bowes, M.D. *et al.* 1993: Impacts and responses to climate change in forests of the MINK region. *Climatic Change* **24**:63-82.

Box, E.O. 1981: Predicting Physiognomic Vegetation Types with Climate Variables. *Vegetatio* **45**:127-139.

Brubaker, L.B. 1986: Responses of tree populations to climatic change. *Vegetatio* **67**:119-130.

Burrough, P.A. 1986: Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press. Oxford.

Cohen, S.J. 1991: Introduction. in Application of the Canadian CLimate Centre General Circulation Model Output for Regional Climate Impact Studies, Guidelines for Users(*unpublished*). Canadian Climate Cnetre, Atmosphere Environment Centre, Downsview, Ontario, Canada, M3h 5T4.

Cook, E. *et al*. 1991: On Predicting the Response of Forests in Eastern North America to Future Climatic Change. *Climatic Change* 19:271-282.

Dale V.H. *et al*. 1994: Assessing impacts of climate change on forests: the state of biological modeling. *Climatic Change* **28**:65-90.

Davis, F.W. *et al*. 1990: Information Analysis of a spatial database for ecological land classification. *Photogrammetric Engineering And Remote Sensing*. **vol**.56, **No**.5 May 1990, **pp**.605-613.

Davis, M.B. 1989: Lags in vegetation response to greenhouse warming. *Climatic Change* **15**:75-82.

Davis, M.B. *et al*. 1985: Sensitivity of cool temperate forests and their fossil pollen record to rapid temperature change. *Quaternary Research* **23**:327-340.

Davis M.B. *et al*. 1986: Dispersal versus climate: expansion of Fagus and Tsugus into the Upper Great Lakes region, *Vegetatio* **67**:93-103.

Davis, F.W. 1990: Information Analysis of a Spatial Database for Ecological Land Classification. *Photogrammatic Engineering and Remote Sensing*. **vol.56, No.5, May 1990. pp**605-613.

Dickinson, R.E. 1983: A History of and Introduction to Climate Models. In A. Henderson-Sellers et al. ed. 1987, *A climate Modelling Primer*. John Wiley & Sons, Chichester . New York . Brisbane . Toronto . Singapore. **pp40-41.**

Dumanski, J. M. Phipps, and E. Huffman, 1987: A study of Relationships between Survey Data and Agricultural Land Use Using Information Theory. *Can. J. Soil Sci.* **67**:95-102.

Ecoregions Working Group, 1989: Ecoclimatic Regions of Canada, First Approximation, Ecoregions Working Group of the Canada Committee on Ecological Land Classification, Ecological Land Classification Series, No. 23, Sustainable Development Branch, Canadian Wildlife Service, Conservation and Protection. Environment Canada, Ottawa, Ontario. 119pp.

Emanuel, W.R. *et al.* 1985: Climatic change and the broad-scale distribution of terrestrial ecosystem complexes. *Climatic Change* **7**:29-43.

Eyre, F.H. 1980: Forest Cover Types of the United States and Canada. Socity of American Foresters.

Gates, D. 1990: Climate Change and The Response of Forest. *Int. J. of Remote Sensing,* 1990. **vol.11 No.7, pp**1095-1107.

Graham, R.L. *et al.* 1990: How increasing CO2 and climate Change Affect Forests. *BioScience* **vol.40 No.8 pp**575-587.

Grey, S.L. *et al.* 1989: Canada's Forest Inventory 1986 - Technical Supplement. Information Report PI-X-86, Petatwawa

National Forest Institute, Forest Canada.

Kelly, J. 1993: Artificial Intelligence-A Modern Myth. Ellis Horwood, New York . London . Toronto . sydney . Tohyo . Singapore. pp29.

Hansen, J., I. *et al.* 1991: Regional Greenhouse Climate Effects. in M.E Schlesinger (editor), *Greenhouse-Gas-Induced Climate Change: A Critical Appraisal of Simulations and Observations.* Amsterdam, Elsevier Science Publishers B.V.

Hansen, J., I. *et al.* 1988: Global Climate Changes as Forecast by Goddard Institute for Space Studies Three-Dimensional Model. *Journal of Geophysical Research*, **vol.93, No.D8** P9341-9364, August 20, 1988.

Hanson J.S. et al. 1989: Spatial Constraints on The Response of Forest Communities to Climate Change. In George P. Malanson, ed. *Natural Area Facing Climate Change*. The Hague, Elsevier Science Publishers B.V.

Henderson-Sellers, A. 1991: Developing an Interactive Biosphere for Climate Models. *Vegetatio* **91**:149-166.

Holdridge, L.R. 1947: Determination of World Plant Formations from Simple Climatic Data. *Science* **105**:367-368.

Land Resource Research Centre, Research Branch, Agriculture Canada, 1986: Soil Landscapes of Canada, Scale 1:1 000 000 Ottawa, Ontario, Canada.

Lenihan, J.M. 1993: Ecological response surfaces for North American boreal tree species and their use in forest classification. *Journal of Vegetation Science* **4**:667-680.

Malanson, G.P. 1993: Comment on modeling ecologicl response to

climatic change. *Climatic Change* February 1993 pp95-109.

Mendelsohn, R. *et al.* 1994: Framework for integrated assesments of global warming impacts. *Climatic Change* **28**:15-40.

Manitoba Mineral Resources Division, 1981: Surrficial Geological Map of Manitoba. Scale 1:1 000 000 Map 81-1.

Overpeck, J.T. *et al.* 1990: Climate-induced changes in forest disturbance and vegetation. *Nature* **343**:51-53.

Pastor, J. *et al.* 1988: Response of northern forest to CO2-induced climate change. *Nature* **334**:55-58.

Pennington, W. 1986: Lags in adjustment of vegetation to climate caused by the pase of soil development:Evidence from Britain. *Vegetatio* **67**:105-118.

Phipps, M. 1981a: Entropy and Community Pattern Analysis. *J. Theor. Biol. (1981)* **93**:253-273.

Phipps, M. 1981b: Information Theory and Landscape Analysis. In *Proc. Int. Congr. Neth. Soc. Landscape Ecol.*, Veldhoven, 1981. Pudoc, Wageningen.

Prentice, C.I. 1992: A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* **19**:117-134.

Prentice, C.I. 1986: Vegetation responses to past climatic variation. *Vegetatio* **67**:131-140.

Rizzo, W. *et al.* 1992: Assessing the sensitivity of Canada's Ecosystems to climatic change. *Climatic Change* **21**:37-55.

Roberts, L. 1989: How fast can trees migrate? *Science* **243**:735-737.

Robock, A. *et al.* 1993: Use of General Circulation Model Output. *Climatic Change* April 1993 pp300-335.

Rotmans, 1990: IMAGE: *An Integrated Model to Assess the Greenhouse Effect*. Kluwer Academic Publishers, Boston.

Savory, S.E. 1988: *Artificial Intelligence and Expert Systems*. John Wiley & Sons, Chichester . New York . Brisbane . Toronto. Singapore. pp236.

Shugart, H.H. *et al*. 1977: Development of an Appalachian Deciduous Forest Succession Model and Its Application to Assessment of the Impacts of Chestnut Blight. *Journal of Environmental Management* 5:161-170.

Shugart, H.H. and D.C. West 1979: Size and Pattern of Simulated Forest Stands. *Forest Science* 25:120-122.

Solomon, A.M. *et al*. 1981: Simulating the role of climate change and species immegration in forest succession. in West D.C., Shugart, H.H., and Botkin, D.B.(eds.), *Forest Succession*, New York, SPringer-Verlag, pp154-177.

Solomon, A.M. 1986: Transcent response fo forests to CO2-induced climate change:simulation modeling experiments in eastern Northe America. *Oecologia* 68:567-579.

Steinhauser, F. 1979: Climatic Atlas of North and Central America-I. WMO UNESCO Cartographia. 1979.

Urban, D.L. *et al*. 1993: Potential response of Pacific Northwestern forests to climatic change: Effects of stand age and initial composition. *Climatic Change* 23:247-266.

Webb, T. 1981: The past 11,000 years of vegetational change in Eastern North America. *BioScience* 31:501-506.

Webb, T. 1986: Is vegetation in equilibrium with climate? How to interpret late-Quaternary pollen data. *Vegetatio* 67:75-91.

Weir, T.R. 1983: Atlas of Manitoba. Surveys and Mapping Branch Department of Naturl Resources, Province of Manitoba, Winnipeg, Canada.

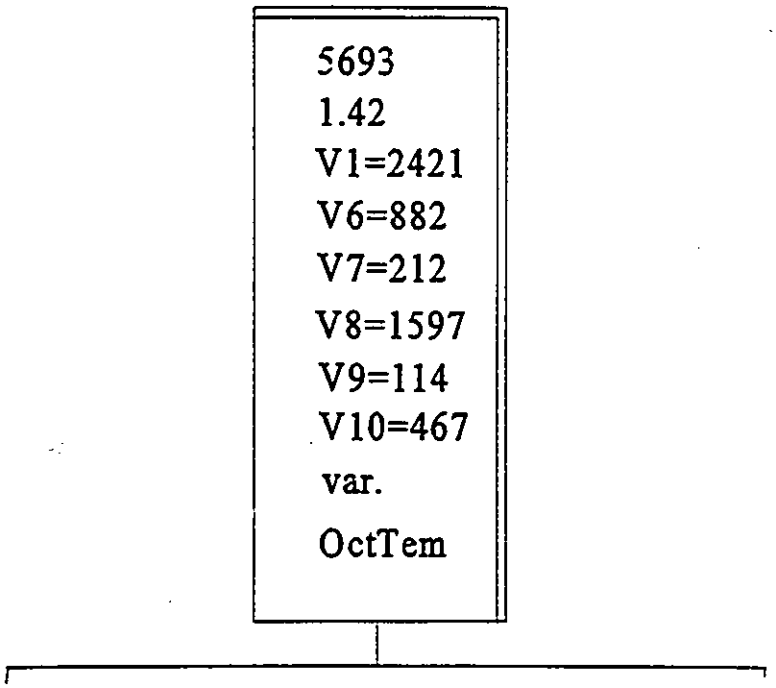
Appendix A

The tree-like structure created by the stepwise divisive process of *PEGASE* is given here for the analysis of obtaining the knowledge base. The total tree structure is given in several pages because it is impossible to draw this big tree in one page. In each page, a set or a subset is divided by a variable into a number of subsets or terminal subsets. The position of this set or subset in the tree is given as the level. For example, level 2-1 means that the set or subset is at the second level of the tree and it is the first subset at level 2.

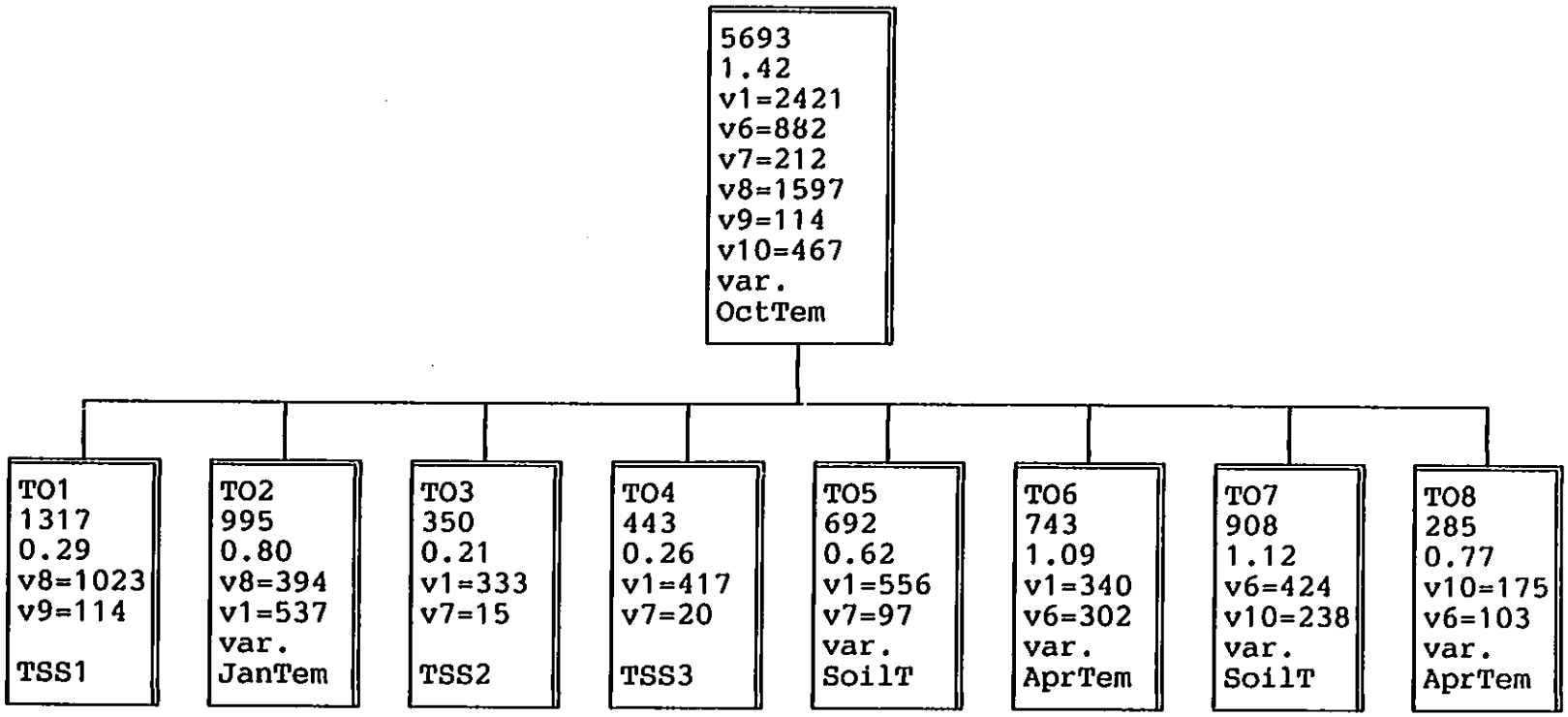
For each set (tree root), subset (non-leaf node) or terminal subset (leaf node, also called *TSS*), the information given includes:

1. the number of cells in the set. It is 5693 cell in the example below;
2. the initial entropy, which is 1.42 in the example below;
3. the number of cells for each forest types. In the example below, V1, V6, V7, V8, V9, and V10 represent specific forest types; and the number on the right side of the equation is the number of cells for the specific forest type. In this example, we have V1=2421, V6=882 , V7=212, V8=1597, V9=114, and V10=467;
4. the environmental variable which divides this set into subset. In this example, it is OctTem, which is October temperature.

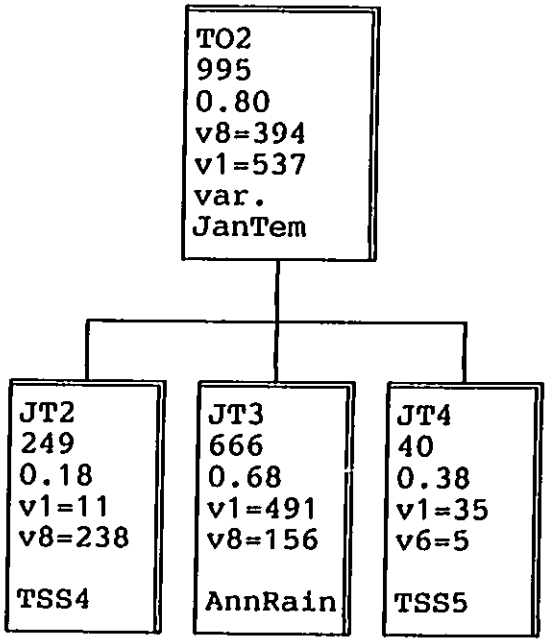
Here is the example:



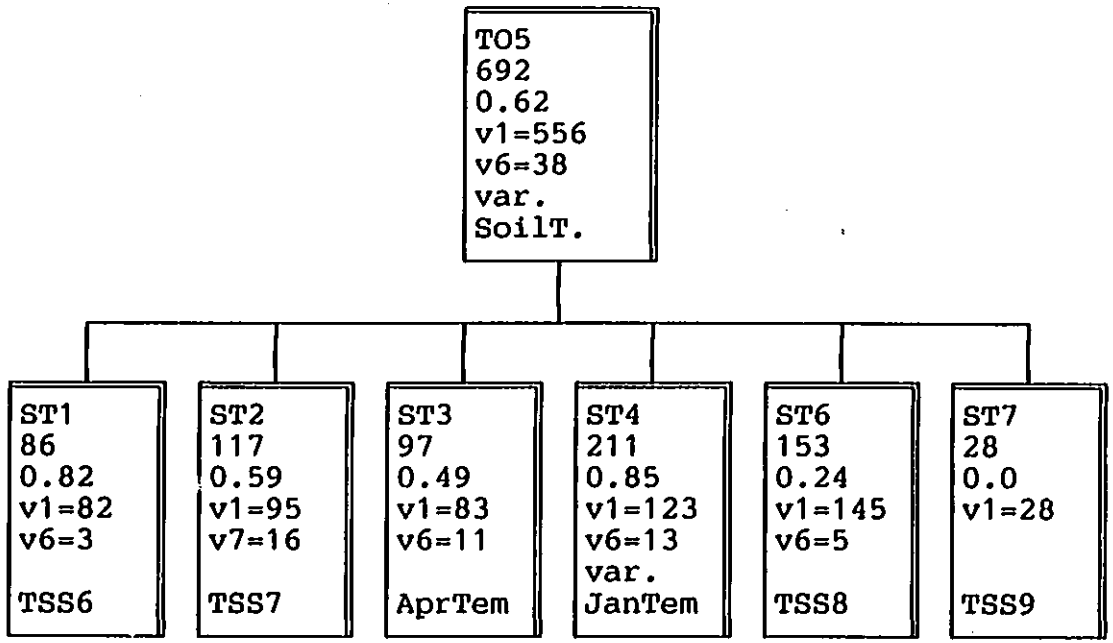
Divisive Processes of PEGASE
Level One (1-1)



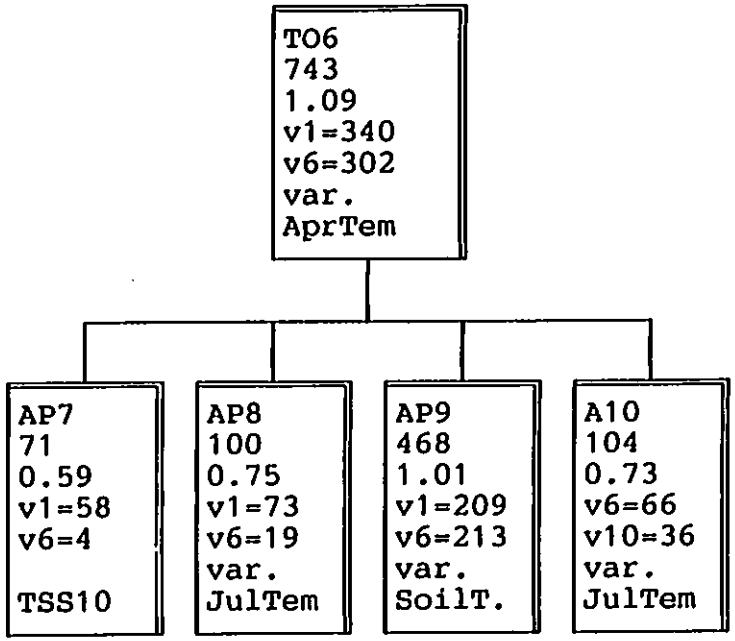
Divisive Processes of PEGASE
Level Two (5-1)



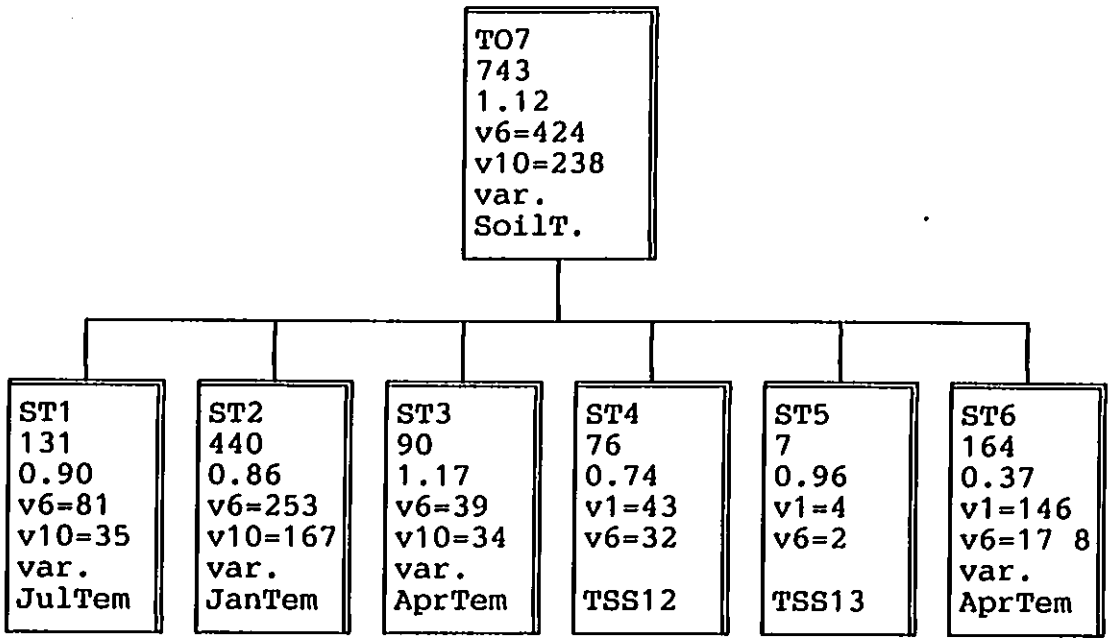
Divisive Processes of PEGASE
Level Two (5-2)



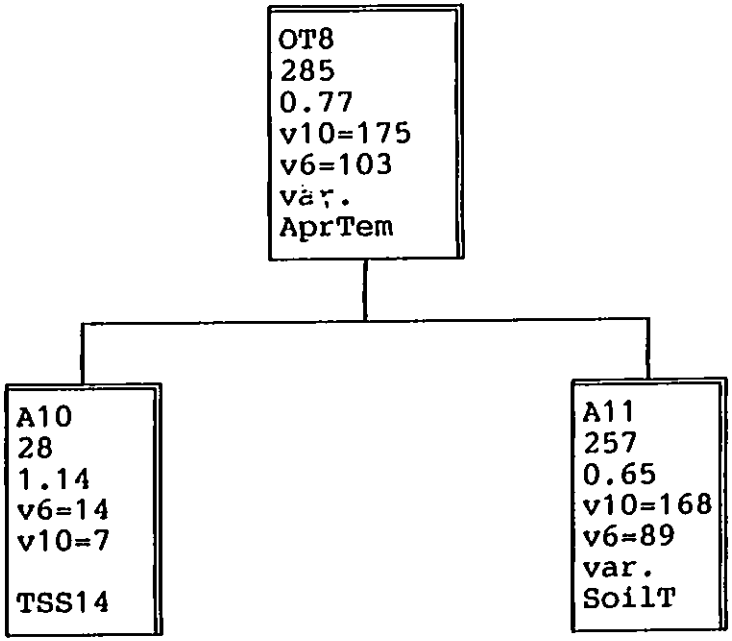
Divisive Processes of PEGASE
Level Two (5-3)



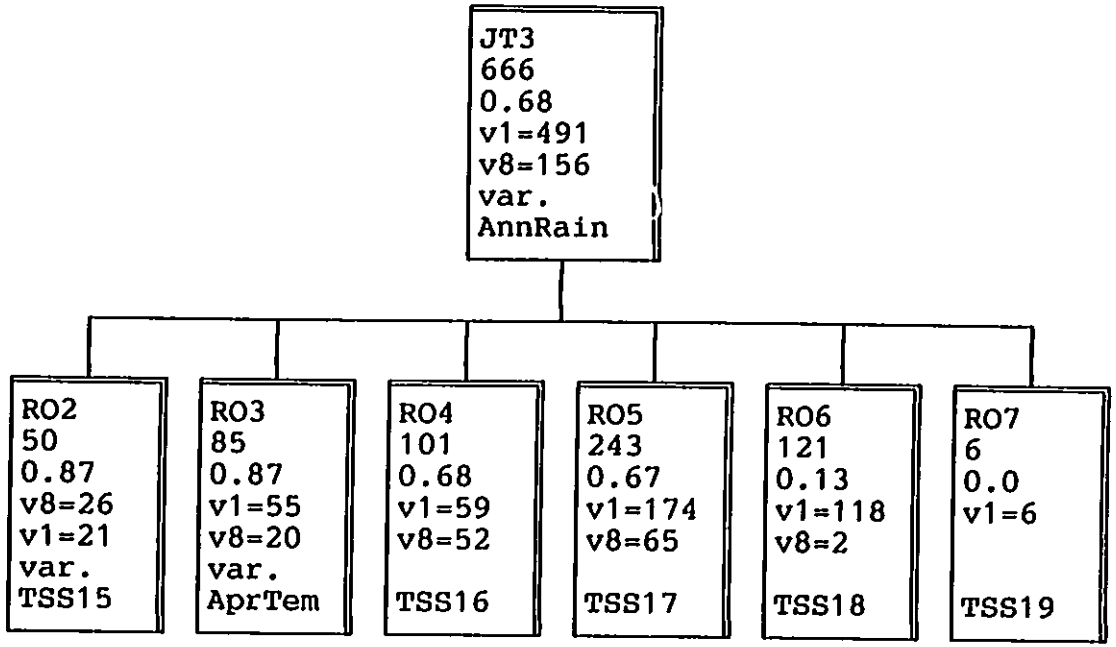
Divisive Processes of PEGASE
Level Two (5-4)



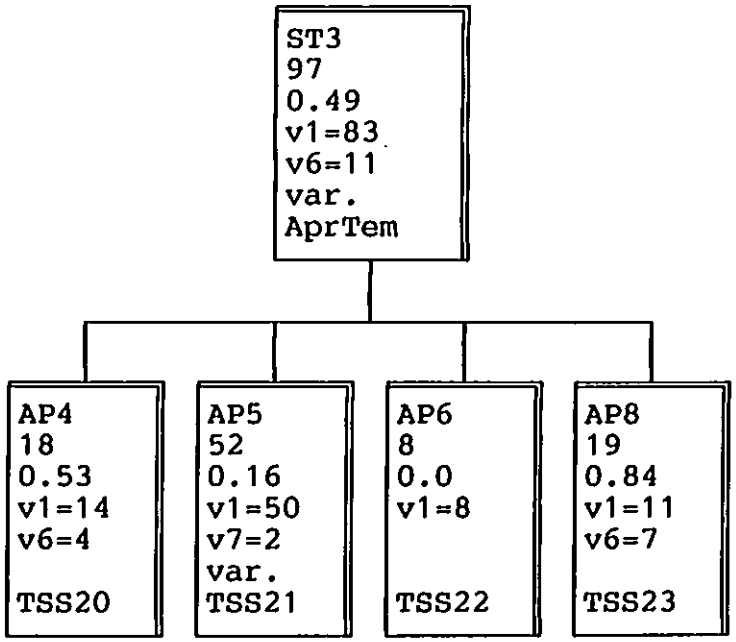
Divisive Processes of PEGASE
Level Two (5-5)



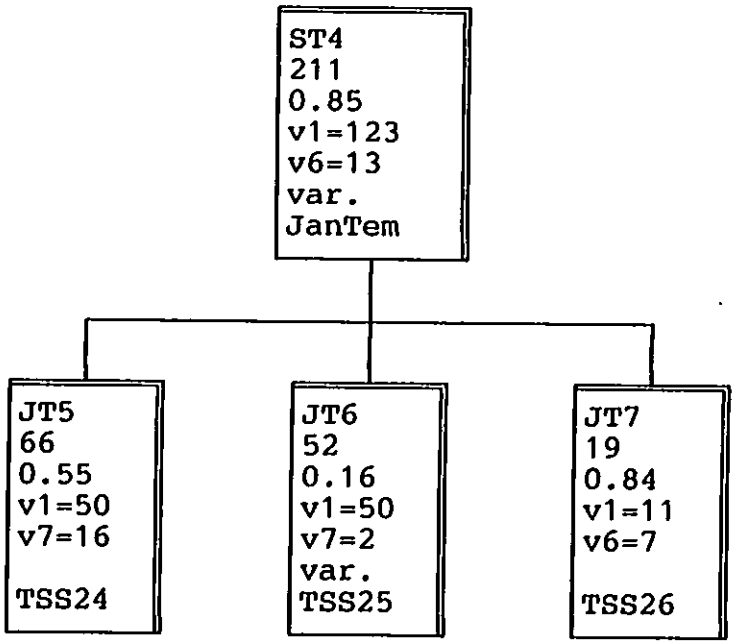
Divisive Processes of PEGASE
Level Three (10-1)



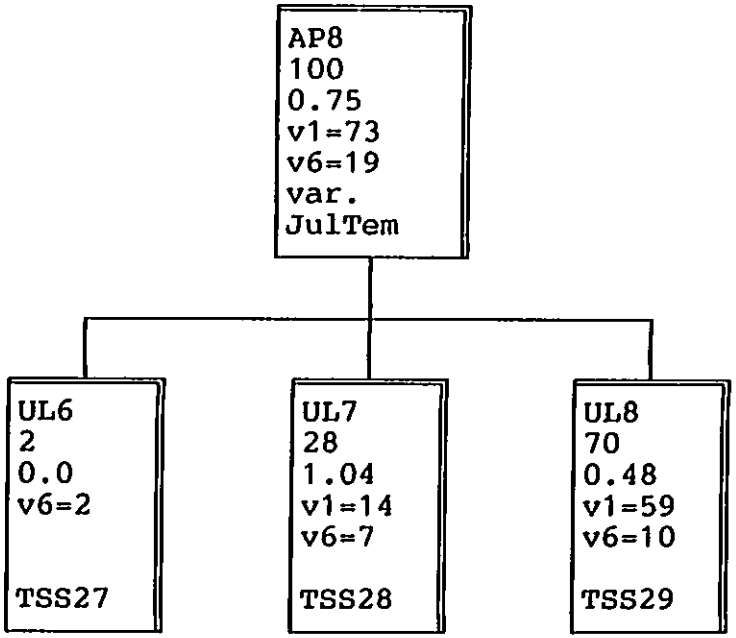
Divisive Processes of PEGASE
Level Three (10-2)



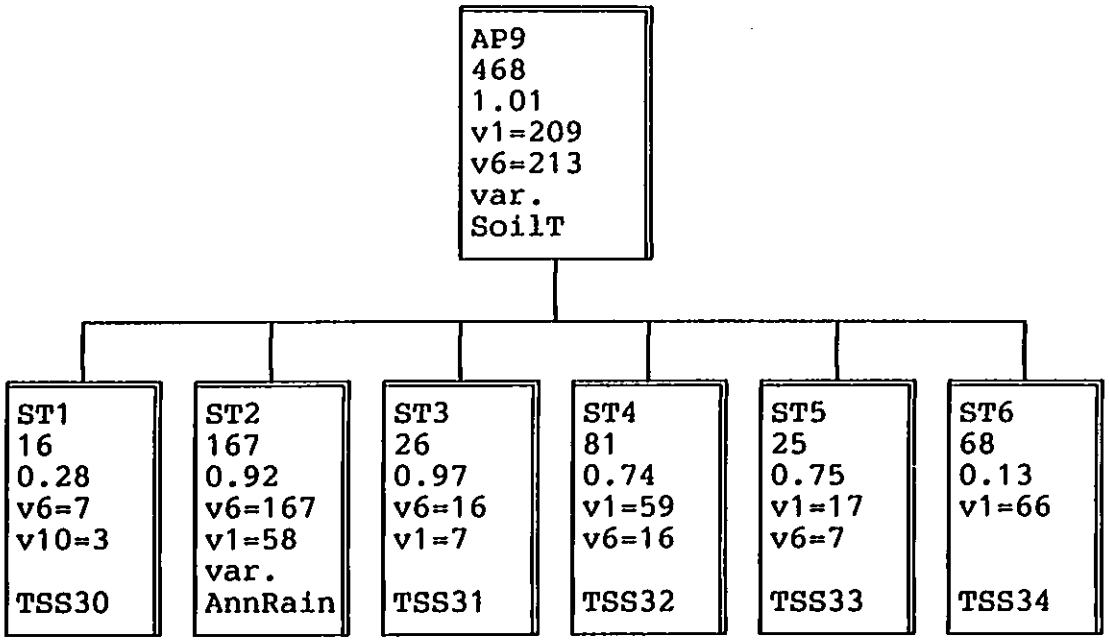
Divisive Processes of PEGASE
Level Three (10-3)



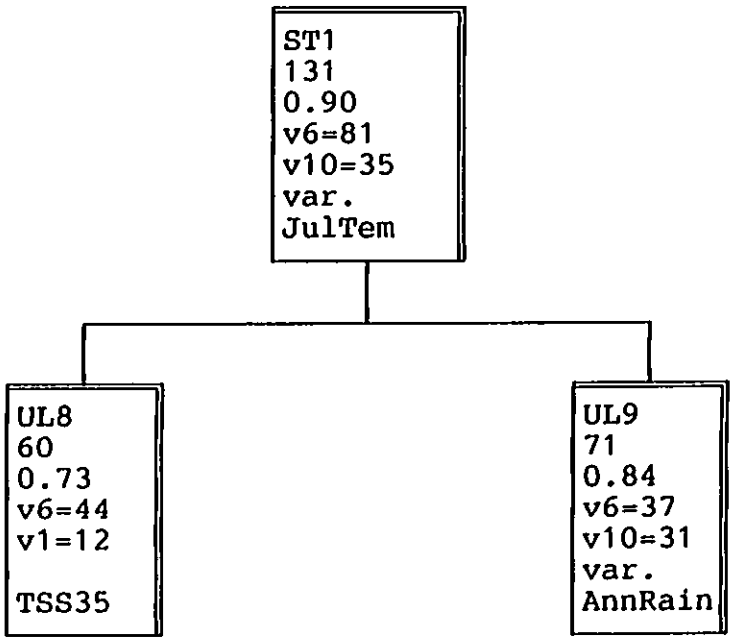
Divisive Processes of PEGASE
Level Three (10-4)



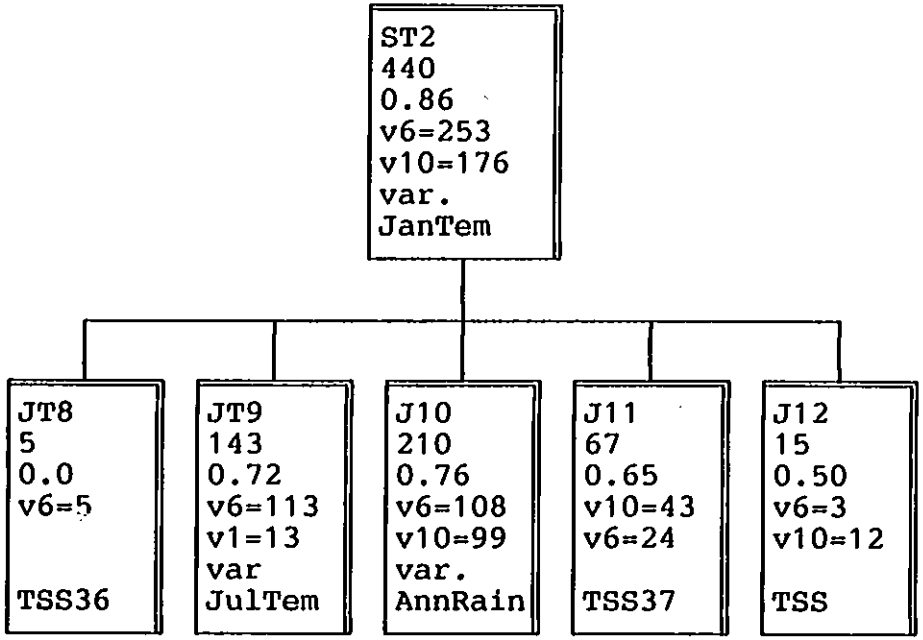
Divisive Processes of PEGASE
Level Three (10-5)



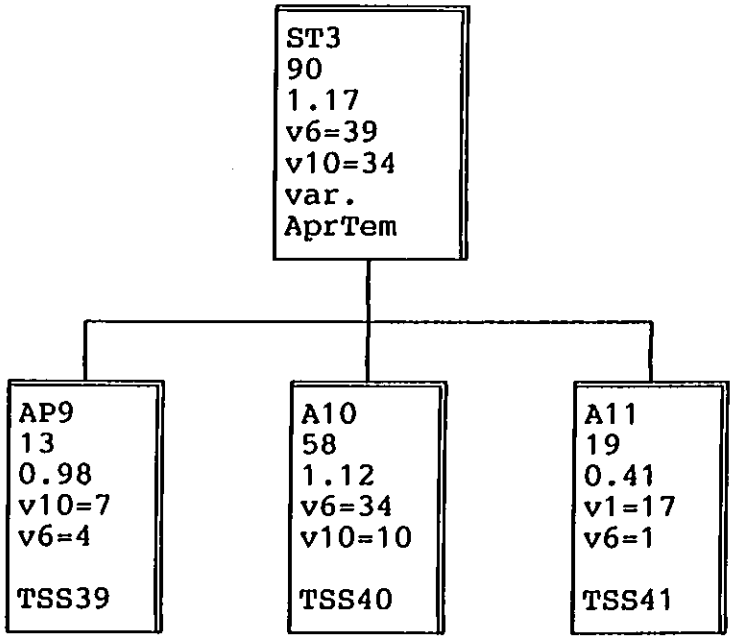
Divisive Processes of PEGASE
Level Three (10-6)



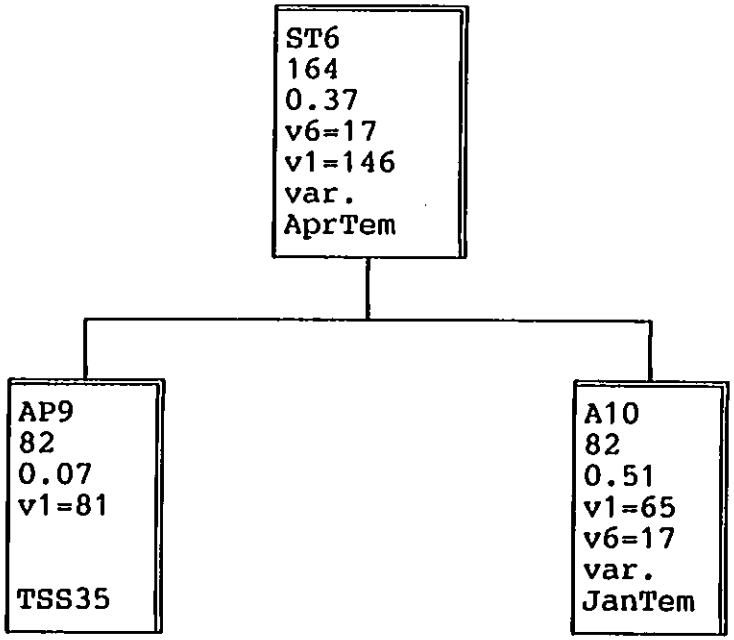
Divisive Processes of PEGASE
Level Three (10-7)



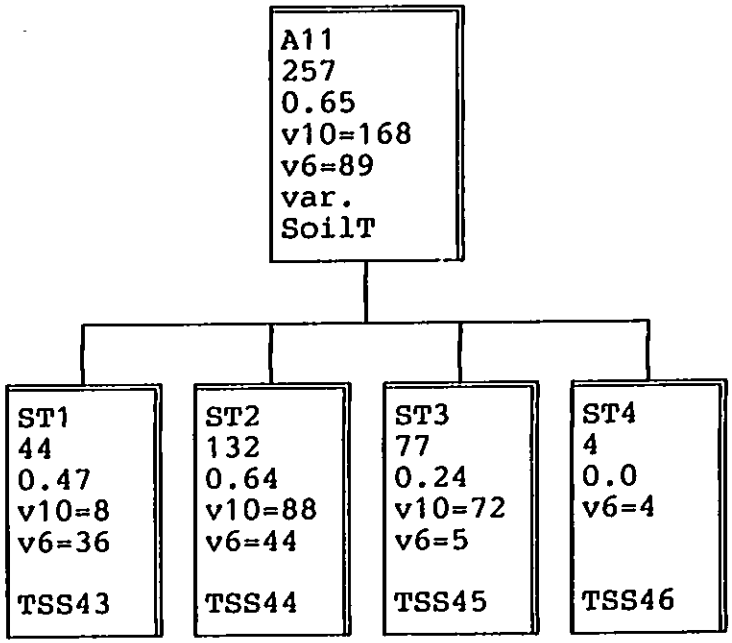
Divisive Processes of PEGASE
Level Three (10-8)



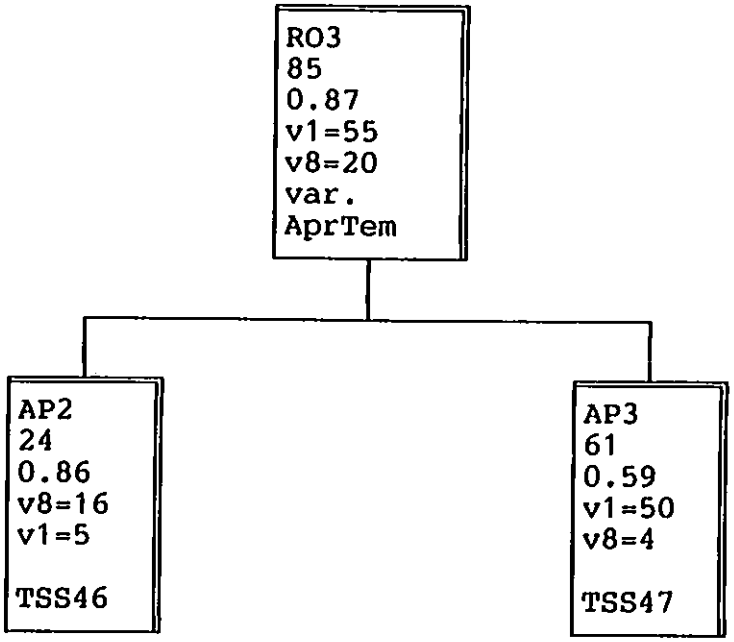
Divisive Processes of PEGASE
Level Three (10-9)



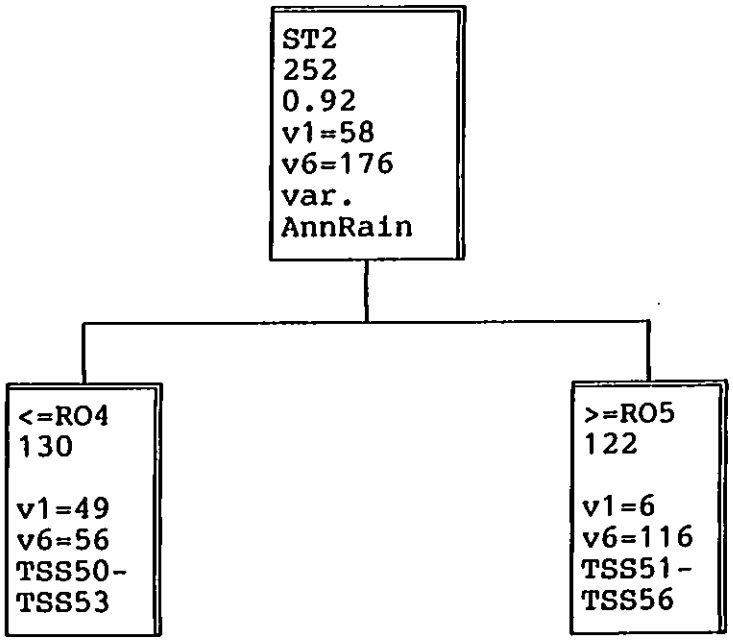
Divisive Processes of PEGASE
Level Three (10-10)



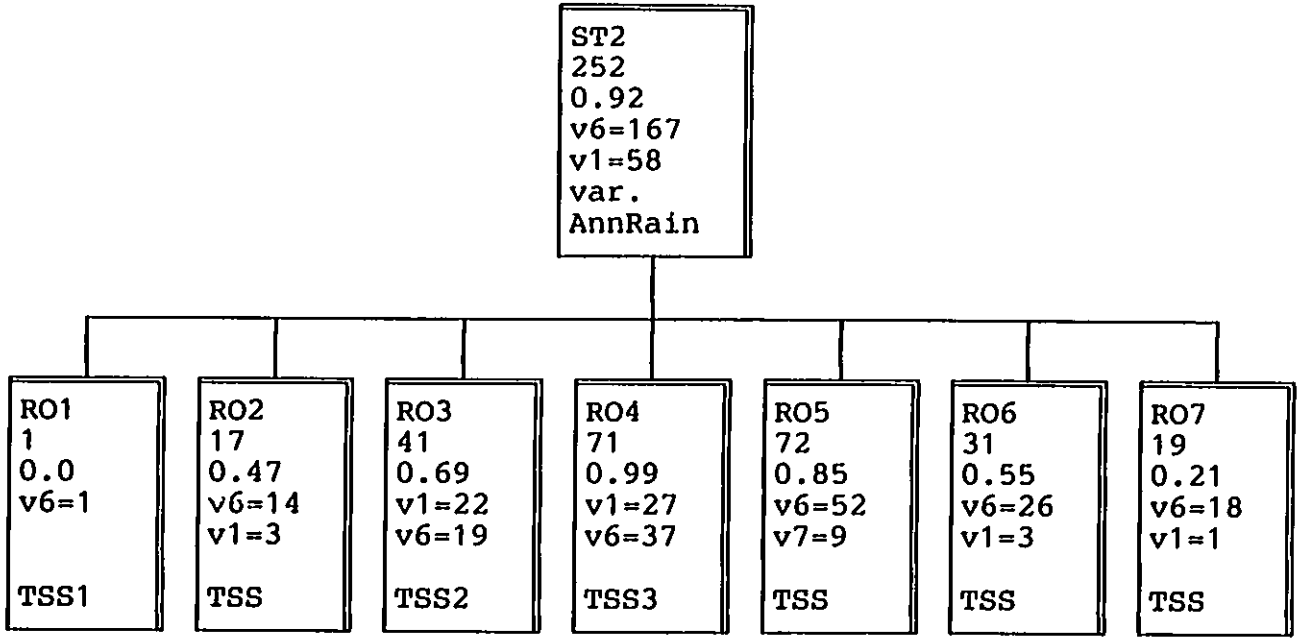
Divisive Processes of PEGASE
Level Four (7-1)



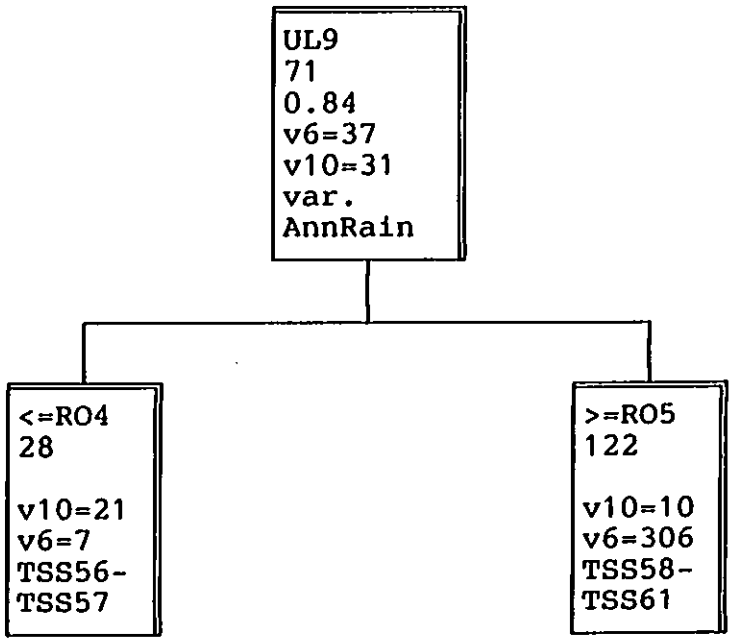
Divisive Processes of PEGASE
Level Four(7-2)



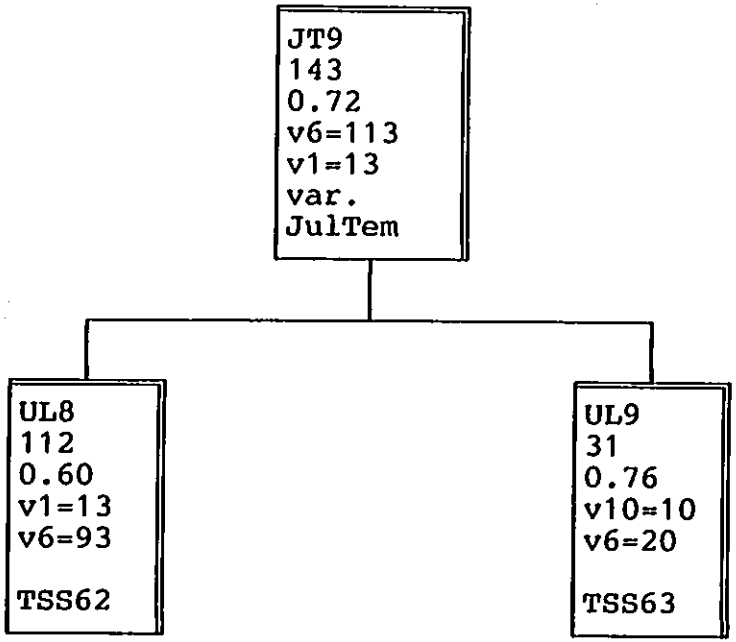
Divisive Processes of PEGASE
Level Four (7-3)



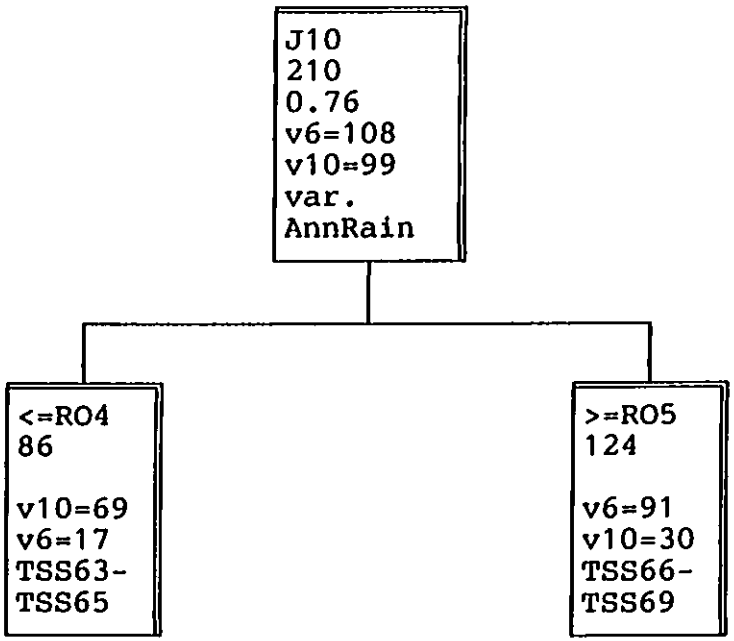
Divisive Processes of PEGASE
Level Four (7-4)



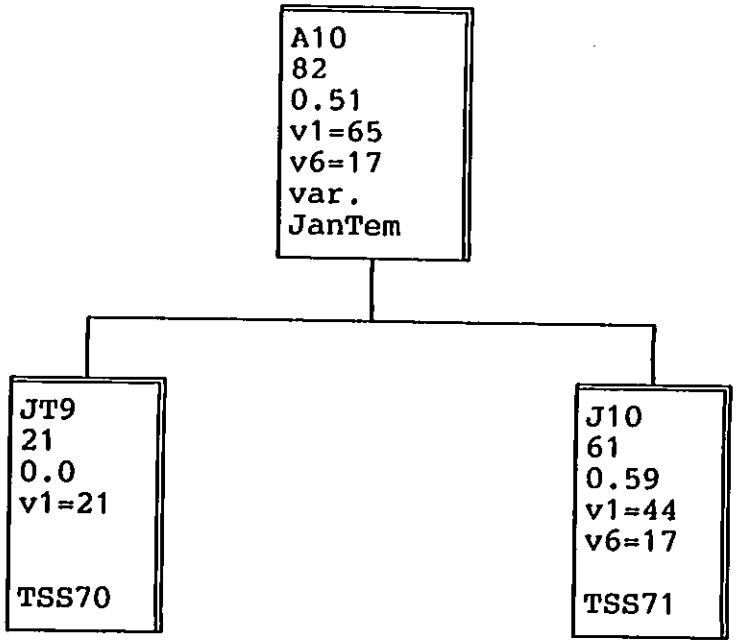
Divisive Processes of PEGASE
Level Four (7-5)



Divisive Processes of PEGASE
Level Four (7-6)



Divisive Processes of PEGASE
Level Four (7-7)



Appendix B

Here is the file which contains the climatic data. The units for temperature and precipitation (including rainfall in summer and snowfall) are degree centigrade and millimetre respectively. The data are organized cell by cell and each cell has a record which is a row of data. The data are also organized in such a way that data in the same column represent the same climatic variables The meanings of each column is explained in following table.

Column No.	Data	Column No.	Data
1	Spatial Location Code	6	October Temperature
2	Annual Temperature	7	Annual Precipitation
3	January Temperature	8	Rainfall in Summer
4	April Temperature	9	Growing Degree Days
5	July Temperature	10	Snowfall

This is a huge data file that contains about 6,000 records; therefore, we cannot give all the data here. We just give a very small part of them so that the reader can have some ideas about the data file.

Here is the data file:

101 -4.2 -27.4 -4.1 15.7 0.1 460.9 251.4 998.4 171.8
102 -4.2 -27.4 -4.1 15.7 0.1 460.9 251.4 998.3 171.8
103 -4.2 -27.4 -4.1 15.7 0.1 460.9 251.4 998.4 171.8
104 -4.2 -27.4 -4.1 15.7 0.1 460.9 251.4 998.4 171.8
105 -4.2 -27.4 -4.1 15.7 0.1 461.0 251.5 998.5 171.8
106 -4.2 -27.4 -4.1 15.7 0.1 461.1 251.6 952.8 171.9
107 -4.6 -27.4 -4.8 15.3 -0.1 454.9 244.9 951.1 174.4
108 -4.6 -27.4 -4.8 15.2 -0.1 454.8 244.7 949.2 174.6
109 -4.6 -27.4 -4.8 15.2 -0.1 454.7 244.5 947.0 174.7
110 -4.6 -27.4 -4.8 15.2 -0.1 454.6 244.3 944.7 174.9
111 -4.6 -27.4 -4.9 15.2 -0.1 454.4 244.1 942.1 175.0
112 -4.6 -27.4 -4.9 15.1 -0.1 454.2 243.8 939.3 175.2
113 -4.6 -27.4 -5.0 15.1 -0.1 454.0 243.5 936.2 175.4
114 -4.7 -27.4 -5.0 15.1 -0.1 453.8 243.2 932.8 175.6
...
4523 -3.5 -26.2 -3.3 15.9 0.6 520.4 284.7 1026.0 209.4
4524 -3.6 -26.3 -3.4 15.8 0.5 525.3 287.0 1019.7 214.9
4525 -3.7 -26.4 -3.5 15.8 0.4 530.2 289.4 1014.4 219.4

4526 -3.7 -26.4 -3.5 15.7 0.4 534.4 291.4 1012.8 224.0
5043 -3.9 -26.4 -4.7 15.8 0.8 450.2 245.4 993.0 178.6
...
6527 -1.6 -24.4 -2.4 18.0 2.9 440.9 231.6 1337.2 163.1
6528 -1.6 -24.4 -2.4 18.0 2.9 441.5 231.6 1335.1 163.2
7351 -2.0 -24.7 -3.1 17.2 3.2 561.0 268.7 1251.3 321.0
7401 -0.6 -22.7 0.4 17.6 3.5 472.0 256.4 1394.9 139.9
...
8432 0.2 -21.7 0.7 18.1 4.5 505.3 264.6 1481.5 151.3
8433 -0.0 -21.9 0.4 18.1 4.4 511.2 269.1 1466.8 157.4
...
9406 1.1 -20.3 2.0 18.4 5.1 484.0 256.5 1570.2 147.1
9407 1.1 -20.2 2.0 18.5 5.2 482.9 255.7 1580.1 146.3
...
11049 1.6 -19.7 2.5 19.0 5.7 544.1 284.7 1648.4 133.1

12647 1.9 -18.9 3.0 18.7 5.7 566.9 305.9 1650.0 121.9
12648 1.8 -19.0 2.9 18.8 5.7 568.5 306.3 1639.3 123.5
12649 1.8 -19.0 2.9 18.8 5.7 568.0 305.0 1643.5 124.8
12650 1.8 -19.1 2.8 18.9 5.8 567.5 302.6 1643.1 126.7
12651 1.7 -19.2 2.8 18.8 5.7 567.0 301.7 1648.5 127.3

Appendix C

The second file of the *KBFM*'s data base is a data file used for *PEGASE* analysis. This data file is composed by two parts: the header and the data.

The header part contains the information which is read by *PEGASE* program. This information includes: (1) the data file name (e.g. MANI.) ; (2) the number of cells (e.g. 6100); (3) the number of column (e.g. 51) and the number of rows (e.g. 125) for the cell system; (4) the number of dependent variables used for analysis (e.g. 12); (5) the default maximum cell number for stopping the division (e.g. 30); (6) the default minimum entropy for stopping the analysis (e.g. 0.30); (7) and the priority sequence of dependent variables which is used when two dependent variables reduce the same amount of entropy. In this case, the *PEGASE* program needs to know which variable to choose (e.g. 2 3 4 5 10 12 6 8 7 9 11 1 14 13 15). The second part of the header gives the number of classes of each variable and the corresponding codes. In short, the header gives a default information for the *PEGASE*. When running *PEGASE*, the user can follow the menu to input or change the default information if he/she wants to.

The second part of the file is the data. The data are organized cell by cell and each cell has a record which is a row of data. The data are also organized in such a way that data in the same column represent the same variables The meanings of each column is explained in following table.

Column No.	Data	Column No.	Data
1	Spatial Location Code	9	July Temperature Code
2	Vegetation Type Code	10	October Temperature Code
3	Geological Deposit Class Code	11	Annual Precipitation
4	Soil Type Code	12	Rainfall in Summer Code
5	Soil Texture Code	13	Growing Degree Days Code
6	Annual Temperature Code	14	Topography Code
7	January Temperature Code	15	Snowfall Code
8	April Temperature Code	16	Permafrost Code

Note: This data file contains information of soil type and permafrost, but there are not used in the analysis.

This is a huge data file, which contains about 6,000 records. Therefore, we cannot give all the data here. We just give a very small part of them so that the reader can have some idea about the data file.

MANI.
6100 51 125 12 1
20 2 2 30 0.30
2 3 4 5 10 12 6 8 7 9 11 1 14 13 15
1 10 Vege.
1 VE1
2 VE2
3 VE3
4 VE4
5 VE5
6 VE6
7 VE7
8 VE8
9 VE9
10 V10
2 12 Geo.depos.
1 GE1
2 GE2
3 GE3
4 GE4
5 GE5
6 GE6
7 GE7
8 GE8
9 GE9
10 G10
11 G11
12 G12
3 7 Soil.depos.
1 ST1
2 ST2
3 ST3
4 ST4
5 ST5
6 ST6
7 ST7
4 13 Soil.type
1 SO1
2 SO2
3 SO3
4 SO4
5 SO5
6 SO6
7 SO7
8 SO8

9 SO9
10 S10
11 S11
12 S12
13 S13
5 12 Tem.annu.
1 TA1
2 TA2
3 TA3
4 TA4
5 TA5
6 TA6
7 TA7
8 TA8
9 TA9
10 T10
11 T11
12 T12
6 12 Tem.jan.
1 JT1
2 JT2
3 JT3
4 JT4
5 JT5
6 JT6
7 JT7
8 JT8
9 JT9
10 J10
11 J11
12 J12
7 11 Tem.apr.
1 AP1
2 AP2
3 AP3
4 AP4
5 AP5
6 AP6
7 AP7
8 AP8
9 AP9
10 A10
11 A11
8 10 Tem.july
1 UL1

2 UL2
3 UL3
4 UL4
5 UL5
6 UL6
7 UL7
8 UL8
9 UL9
10 U10
9 8 Tem.oct.
1 TO1
2 TO2
3 TO3
4 TO4
5 TO5
6 TO6
7 TO7
8 TO8
10 10 Rain_annu.
1 R01
2 R02
3 R03
4 R04
5 R05
6 R06
7 R07
8 R08
9 R09
10 R10
11 6 Rain_grow
1 RG1
2 RG2
3 RG3
4 RG4
5 RG5
6 RG6
12 12 Degr.day
1 D01
2 D02
3 D03
4 D04
5 D05
6 D06
7 D07
8 D08

9 D09
10 D10
11 D11
12 D12
13 5 Landform
1 LF1
2 LF2
3 LF3
4 LF4
5 LF5
14 9 Snowfall
1 SN1
2 SN2
3 SN3
4 SN4
5 SN5
6 SN6
7 SN7
8 SN8
9 SN9
15 4 Permafrost
1 PM1
2 PM2
3 PM3
4 PM4
0
0
103 8 11 1 7 4 2 3 5 2 4 3 3 3 5 3
104 8 11 1 7 4 2 3 5 2 4 3 3 3 5 3
105 8 11 1 7 4 2 3 5 2 4 3 3 3 5 3
106 8 11 1 7 4 2 3 5 2 4 3 3 3 5 3
...
...
12643 5 1 1 3 10 11 11 8 7 9 6 10 3 2 1
12644 5 1 1 3 10 11 11 8 7 9 6 10 3 2 1
12645 5 1 4 12 10 11 11 8 7 9 6 10 3 2 1
12646 6 1 4 12 10 11 10 8 7 9 6 10 3 3 1