CANADIAN THESES ON MICROFICHE

I.S.B.N.

THESES CANADIENNES SUR MICROFICHE

National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

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

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

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

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

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

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

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

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

A STOCHASTIC SYSTEMS MODEL OF GROUNDWATER LEVEL FLUCTUATIONS

BY

Tamas Hamory

A thesis presented to the
School of Graduate Studies and Research
of the University of Ottawa
in partial fulfillment of the requirements
for the degree of Master of Applied Science (Civil Engineering)

Ottawa, Ontario, 1981

ABSTRACT

Groundwater level fluctuations (log transformed) are considered as the response of a watershed fluvial system to riverflow (log transformed) in an input-output dynamic model formulation of a groundwater regime. The model is hypothesized as a lumped linear system, with another superimposed linear system representing the errors (e.g., in sampling, discretizing etc.) with white noise as input.

Frequency and time domain methods of time series techniques were employed in the analysis of two watersheds in eastern Ontario. The results have shown that a parsimonious linear input-output model is adequate to represent the five-daily nonstationary groundwater level and riverflow series. A second order dynamic model (with five parameters) was found adequate to describe the groundwater level fluctuation process. This model explains 92 to 95 percent of total variance.

The remaining noise (error) component accounting for 5 to 8 percent of variance in the five-daily groundwater level fluctuation was found to be a white noise series.

Unit step and frequency response functions were found to provide further information on modeling and interpreting the watershed fluvial characteristics. The memory content of the groundwater level - riverflow system, as indicated by the frequency response function, was found to be insignificant.

ACKNOWLEDGEMENTS

The author is indebted to all those who contributed information and advice during this project, especially to:

Dr. Kazimierz Adamowski, for the help in combining theory and practice, and for the review and criticism of early drafts of this work;

Elisabeth Sullivan for the help with computer subroutines;

Krisztina Nagy, of CANMET Library, for providing hard-ly accessible literature material;

The Agrometeorology Research and Service, Land Research Institute, Canada Department of Agriculture and the Inland Waters Directorate, Water Resources Branch, Water Survey of Canada for making available the data.

The research was financially supported by the National Research Council of Canada through a bursary and grant No. A7467.

LIST OF FIGURES

Figure 1. 1	Input-output view of groundwater hydrology . . .	3
Figure 3. 1	Location of selected watersheds	24
Figure 4. 1	Black box representation of groundwater level fluctuations	33
Figure 4. 2	Lumped linear representation of the ground- water level fluctuation process	44
Figure 4. 3	Lumped linear representation of the groundwater level fluctuation process as hypothesized	52
Figure 6. 1	Power spectra of non-stationary five-daily series (Castor watershed)	86
Figure 6. 2	Power spectra of non-stationary five-daily series (Carp 1 watershed)	87
Figure 6. 3	Power spectra of non-stationary five-daily series (Carp 2 watershed)	88
Figure 6. 4	Cross-correlogram (Castor watershed)	99
Figure 6. 5	Cross-correlogram (Carp 1 watershed)	100
Figure 6. 6	Cross-correlogram (Carp 2 watershed)	101
Figure 6. 7	Phase spectrum (Castor watershed)	105
Figure 6. 8	Phase spectrum (Carp 1 watershed)	106
Figure 6. 9	Phase spectrum (Carp 2 watershed)	107
Figure 6.10	Gain spectrum (Castor watershed)	112
Figure 6.11	Gain spectrum (Carp 1 watershed)	113
Figure 6.12	Gain spectrum (Carp 2 watershed)	114
Figure 6.13	Coherence spectrum (Castor watershed)	118
Figure 6.14	Coherence spectrum (Carp 1 watershed)	119
Figure 6.15	Coherence spectrum (Carp 2 watershed)	120
Figure 6.16	Power spectrum of the noise component (Castor watershed)	124
Figure 6.17	Power spectrum of the noise component (Carp 1 watershed)	125
Figure 6.18	Power spectrum of the noise component (Carp 2 watershed)	126
Figure 6.19	Cummulative periodogram (Castor watershed)	134
Figure 6.20	Cummulative periodogram (Carp 1 watershed)	135
Figure 6.21	Cummulative periodogram (Carp 2 watershed)	136
Figure 6.22	Step response function for five-daily groundwater level fluctuation system	145
Figure 6.23	Gain function for five-daily groundwater level fluctuation system	150
Figure 6.24	Phase function for five-daily groundwater level fluctuation system	151

LIST OF TABLES

Table 3. 1 Features of sampling stations in selected watersheds	25
Table 4. 1 Results of multiple linear regression analysis (log transformed variables)	47
Table 6. 1 Results of hypothesis testing - five-daily-groundwater levels	90
Table 6. 2 Results of hypothesis testing - five-daily riverflow	91
Table 6. 3 Percent variance explained by harmonics - five-daily-groundwater levels	93
Table 6. 4 Percent variance explained by harmonics - five-daily riverflow series	94
Table 6. 5 Values of harmonic coefficients - five-daily groundwater levels	95
Table 6. 6 Values of harmonic coefficients - five-daily riverflow series	96
Table 6. 7 Values of cross correlation coefficients at zero lag and the variance explained	103
Table 6. 8 Results of test of randomness for the phase function	108
Table 6. 9 Results of Mann-Whitney U-test of identical population for phase function	109
Table 6.10 Results of test of randomness for the gain function	115
Table 6.11 Results of Mann-Whitney U-test of identical population for gain function	116
Table 6.12 Coherence values at annual cycle for non-stationary and stationary series	121
Table 6.13 Mann-Whitney U-test of identical population for noise series	127
Table 6.14 Cross-correlation check for residuals	131
Table 6.15 Auto-correlation check for residuals	132
Table 6.16 Testing of residuals for whiteness (Portmanteau test)	133
Table 6.17 Parameter estimates of accepted model orders	138
Table 6.18 Analysis of variance	139
Table 6.19 Variances explained by model orders	152

CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
LIST OF FIGURES	v
LIST OF TABLES	vi

<u>Chapter</u>	<u>page</u>
I. INTRODUCTION	1
Motivation	1
Groundwater level fluctuations in general	2
Mathematical models	5
II. LITERATURE REVIEW	10
Groundwater level modeling	10
Time series approach	13
Applications of time series analysis in hydrology	17
Objectives	21
III. RESEARCH DATA	23
Watersheds selected	23
Hydrogeology	26
Hydrology	26
Geology	27
Data selection	28
IV. THEORETICAL DEVELOPMENTS	31
Model hypothesis	31
Mathematical representation of the process	32
Error analysis	43
Multiple linear regression analysis	45
Hypothesized input-output model	49

V.	METHODS OF ANALYSIS	56
	Introduction	56
	Input-output analysis	57
	Input-output analysis in the frequency domain	60
	Computational procedure	65
	Periodic component	69
	Linear system identification	71
	Harmonic analysis	73
	Model diagnostic checking	77
	Cross correlation check	78
	Auto correlation check	79
	Testing the power spectrum	79
	Periodogram testing method	80
	Portmanteau test	82
VI.	RESULTS AND DISCUSSION	83
	Introduction	83
	Model construction	84
	Stochastic component deduction	84
	Cross spectral and cross correlation analysis	97
	Cross correlogram	98
	Phase spectra	104
	Gain spectra	111
	Coherency spectra	117
	Power spectra of the noise component	123
	Model fitting for five-daily groundwater and riverflow processes	128
	Frequency versus time domain analysis	128
	Model fitting in terms of time domain techniques	129
	Remarks on the performance of model hypothesis	140
	Discussion on groundwater level fluctuation process	141
	Introduction	141
	Unit step response function	141
	Frequency response function	147
VII.	CONCLUSION	155
VIII.	SUGGESTIONS FOR FUTURE STUDIES	158
	BIBLIOGRAPHY	160

Chapter I

INTRODUCTION

1.1 MOTIVATION

The simulation of the watertable in a given aquifer is a problem of extreme importance to hydrology. For optimal groundwater management purposes it is necessary to be able to simulate with reasonable accuracy the responses of the aquifer system. Therefore many related models have been developed. Such models can be used to study the response of a groundwater system to various inputs. However, great variability of geological formations (aquifers) and hydrologic processes affecting groundwater levels create a very complex hydrologic response.

Starting from the basic fact that these hydrologic processes are time dependent and contain a significant amount of uncertainty, it seems plausible to consider them as random variables. Random variables, then, are best described by stochastic means therefore the need arises to develop a

stochastic model for simulation purposes. The objective of this study therefore is to develop such a stochastic model by employing time series analysis techniques.

1.2 GROUNDWATER LEVEL FLUCTUATIONS IN GENERAL

A groundwater level depends on many interrelated factors, the important ones being precipitation, riverflow, and is also influenced by other hydrological and meteorological processes (see Figure 1.1). The primary source of water supply is from precipitation, which must pass through many systems (i.e. surface retention, snow pack, etc.). However, rainfall is not an accurate indicator of groundwater changes: it is estimated [13]¹ that in the study area approximately 99% of rainfall either runs off or evaporates immediately, only the remainder infiltrates the topsoil and permeates downwards until it reaches a zone saturated with water.

Other storage changes are produced by variations of flow in surface waters. A rise or decline of water level in surface waters will cause variations in groundwater levels. The level of a watertable may also exhibit changes due to inflow from surface waters during flood periods.

¹ Numerals in brackets refer to corresponding items in the Bibliography.

FIGURE 1.1: INPUT OUTPUT VIEW OF GROUNDWATER HYDROLOGY

○ h_1, x_1, y_1 location of well

Natural discharges may be in the form of evaporation and/or transpiration to the atmosphere, surface discharges such as springs, or seepage to deep lying groundwater zones. Evaporation and transpiration are highly correlated with temperature. Also, the atmospheric pressure is related to groundwater level fluctuations.

The movement of groundwater in both vertical and horizontal directions are controlled by the geologic environment.

Superimposed on the natural hydrologic regime are man's various activities. These influences include artificial discharge (e.g. pumping) and artificial recharge (e.g. irrigation). Infiltration, evapotranspiration and overland flow are also affected by land use and cultivation.

It is clear therefore that the great heterogeneity of soil types and the very irregular pattern of precipitation in both time and space creates a very complex hydrologic regime to be modeled.

Hydrologists have been undertaking studies aimed directly or indirectly at providing a model suitable for groundwater level simulation. Because simulation entails a mathematical abstraction of real world systems, some degree of misrepresentation of system behaviour can occur. It has to be recognized that the levels of accuracy of the simula-

tion depend on the accuracy with which the model can account for the real process. The more detailed the model the more accurate the results are likely to be. This in turn necessitates greater data demands and time requirements for developing and refining the model.

Groundwater models are characterized by large dimensionality and a large number of variables which are in practice very difficult to obtain. Therefore, for practical purposes models are of necessity greatly simplified. In spite of this, for the analysis of complex water resources systems comprised of many interacting components, such as the groundwater level fluctuation process, computer simulation often proves to be the only feasible tool.

A brief, but general overview is presented of various mathematical models in the following section.

1.3 MATHEMATICAL MODELS

One of the major aims of hydrology is the mathematical description of the response of a hydrologic system to a series of events during a selected time period. Simulation models are designed so as to contain within their parameters all of the information concerning the system and its response characteristics.

There are two major approaches to simulation modeling: optimization methods and nonoptimization methods. The former techniques are concerned with the selection of the best set of conditions to satisfy a set of objectives for a given set of constraints. Nonoptimization methods in mathematical modeling consist of the technique where the response of a system is analyzed for particular sets of data.

These models may be used, among others, either for simulation ("description" of actual process), or synthesis ("generation" of data). In this study, only the simulation problem will be considered further. Simulation modeling consists of deterministic and statistical techniques.

The deterministic simulation techniques are categorized into empirical and conceptual methods. The former describes the hydrologic process based on the assumption that the input and output observations do not contain errors, and a unique transformation from input to output exists. This type of model, being governed by known physical laws, does not include probabilistic interpretation of the results.

Conceptual models rely heavily on theory to interpret phenomena rather than represent the physical process. Since the information (data) content in this case is not complete, the equation describing the process is only an approximate one, therefore such a deterministic conceptual model is not

very realistic. In order to account for uncertainties in the data, model, and parameters, many statistical models have been introduced.

These statistical methods include regression and correlation analyses, probabilistic analysis, stochastic methods, and systems analysis. Long records of hydrological variables can be treated as time series in the statistical sense, and as such, could be employed in simulation problems.

Hydrological simulation models, whether stochastic or deterministic, may be either linear or non-linear [30]. A model is linear if the principle of superposition holds: that is, given that $Y(t)$, $y(t)$ are the outputs corresponding to inputs $X(t)$, $x(t)$, a model is linear if the output, corresponding to input $X(t)+x(t)$, is $Y(t)+y(t)$.

Stochastic linear models are constructed from random events and allow for a 'memory' effect so that each observation is dependent to some degree upon its predecessors in time. Therefore, statistical properties of observed time series provide a means in attempting to identify the stochastic process. This identification approach consists of an a priori model postulation and acceptance or rejection of the model in terms of statistical tests.

The stochastic methods of time series analysis with a conceptual representation in the form of an input-output system will be considered further in this thesis as it relates to the simulation problem.

The input-output stochastic models were mainly developed for rainfall-runoff relations. However, not much research was reported in the literature using these methods in groundwater hydrology. In the past, deterministic methods were mainly used. These methods were extended to very complex models [24],[23] but their data requirements are impossible to obtain in practice. Therefore, the systems methods of input-output analysis might be useful, since they do not require such an extensive amount of data. Furthermore, the systems methods incorporate the stochastic nature of the variables affecting the groundwater level fluctuation process, and as such are more realistic.

It can then be concluded, that due to the random nature of the observed groundwater level time series, it would be advantageous to employ stochastic models in the analysis.

Groundwater level modeling based on the input-output concept mentioned earlier will be employed in this study which transfers inputs of precipitation, riverflow, evapotranspiration, etc. information into groundwater level as output. In this approach the inputs are deterministic-sto-

chastic ones to a probabilistic fluvial system with ground-water level as a deterministic-stochastic output.

Chapter II

LITERATURE REVIEW

2.1 GROUNDWATER LEVEL MODELING

In its earliest days, hydrologic modelling dealt with the most visible aspects of the hydrologic cycle, namely with rainfall and runoff. These models were simple: reflecting knowledge at that time on the one hand, and the necessity of hand computation limited the complexity on the other. As understanding was gained, the developing complexity forced scientists to create and use more complex models. With complexity came the separation of major components of the hydrologic cycle; these being infiltration, evapotranspiration, aquifer response, and surface streamflow routing. The emerging specialists developed empirical formulas to approximate theoretical physical laws for each of these components.

Early work in aquifer response goes back as far as Henry Darcy (1803-1858), but the main developments have been in

the past 30 years. In the middle of the twentieth century, the need for more sophisticated watershed management techniques was recognized, therefore the methods of well hydraulics, field designs, as well as regional groundwater flow studies were developed. These studies were pioneered by Hubbert [32] who showed that groundwater flow can be represented by a three-dimensional potential field. The model was based on a solution of Laplace's equation, which in turn was first improved by Toth [58], who classified the groundwater flow into local, intermediate, and regional flow systems, then by Freeze [23] who introduced a steady state model for groundwater flow systems. For unsteady flows, however, the latter model is not sufficient; nonlinear partial differential equations govern groundwater flow in this case. These equations can not be solved directly except upon replacing the differential equations by a finite difference equivalent which can then be solved by repeated arithmetic calculations, as was demonstrated by Freeze [24].

This and similar numerical analysis procedures have been applied to a variety of groundwater problems. Included are relaxation techniques for flow net analysis [60], for problems relating to seepage [57] and with the introduction of computers, possibilities in groundwater basin modelling have considerably widened (see for example Skibitzke [53], and Remson et al. [46]).

The application of computers, and numerical methods just described, had profoundly changed groundwater hydrological investigations. Since the applicable mathematical model for groundwater flow systems is usually written in finite difference form, digital computers proved to be versatile tools because of their speed and large capacity. These favourable characteristics enabled hydrologists to extend mathematical methods to groundwater studies. To analyse groundwater aquifers several techniques have been employed, among those analytical [59], empirical [39], and numerical [24] methods while utilizing the rapid development of digital computers.

In the analytical study [59] it was concluded that the regulating effect of creeks draining the groundwater flow system determines the zone of fluctuation. This fluctuation is likely to be smaller in discharge areas than in recharge areas. Daily groundwater losses may include water table draining to adjacent surface waters and evapotranspiration. This latter one was found to be directly related to the depth of the water table [39]. Freeze [24] developed a mathematical model by assuming that a system is in dynamic equilibrium, i.e. the water table is in an almost steady state condition. These, and other [7], [56], by now classical investigations of groundwater aquifers are most often

confined to the study of the aquifer's deterministic properties. Recently, Smith et al. [54] using Monte Carlo experiments extended these models to include probabilistic interpretation.

As it was mentioned in the preceeding chapter, all hydrologic variables effecting groundwater levels are random processes. Given that hydrologic processes are random or stochastic by nature, it is hardly surprising that many hydrologists are prolific users of statistical methods and time series or stochastic analysis. The use of time series analysis is common in many other fields, but before exploring applications in groundwater hydrology, several statistical concepts need to be explained, therefore the following section is devoted to this purpose.

2.2 TIME SERIES APPROACH

A time series is a set of even time-spaced observations on some given natural phenomenon. If the series contain some unpredictable elements the variable making up the series is considered a stochastic variable [41]. A time series is called stationary if it has finite variance and all covariances depend only on the difference between time points. In the case of non-stationarity, these very time points are of

extreme importance; the series then are classified as evolutionary, or stochastic [38].

A stationary series may contain completely predictable deterministic components such as trends, jumps, and periodic components. These components can be interpreted as linearly superimposed on a time series generated by some linear stochastic process [29].

Trends are gradual increases or decreases of variables which happened in the past and will continue for a reasonable future period. These can be removed either by regression analysis or by the moving average method.

Sudden increases or decreases in variables in the process that produces the variable are considered as jumps. These slippages are to be appropriately removed according to the situation in which they have occurred.

The basic property of a periodic variable is that it repeats itself at regular time intervals with unmistakable connections to astronomical events. The variation is likely to be sinusoidal of some sort (higher harmonics are also to be regarded as deterministic components) and is conveniently removed by fitting a Fourier series with the appropriate number of harmonics to the stationary series [1].

Upon removal of the above mentioned deterministic components the remaining series consists of only the random el-

ement. A stationary stochastic process generates a time series which show fluctuations and this has led to the concept of the frequency spectrum density which explains the manner in which the total variance of a stochastic variable is distributed in various frequencies. The notion, called white noise spectrum [28], states that the variance is expected to be evenly distributed over all possible frequencies. Since spectrum density functions are computed from time series, it is possible to obtain a condensed picture of the statistical properties of a time series.

A model, called ARIMA (autoregressive-integrated-moving-average) process, has been developed by Box and Jenkins [10] and is capable of describing a wide range of behaviour of time series observed in hydrology.

It belongs to a lumped stationary dynamic linear system, a system that is said to be dynamic in the sense that it receives certain quantitative inputs and accordingly acts to produce certain quantitative outputs [17]. The mathematical relation between input and output is provided uniquely by the impulse response function of the system. Once this function is known, the output can be determined for any specified input for a given linear system. This function may be determined through analytical or numerical means if the internal structure of the system is accounted for.

Generally the internal system can not be mathematically described accurately and therefore observed sets of input and output data are used to identify the response function. The identification procedure includes cross-spectral analysis and model diagnostic checking. The latter is for verifying the adequacy of the model developed by cross-spectral procedures.

Therefore a further objective of this study is to give a statistical description of the internal structure of the system through response functions generated by observed input and output variables.

As was distinguished before, there are two aspects in modelling, namely simulation and synthesis (generation). Since the statistical and modelling aspects are of concern in this thesis, thus generation of equally likely sequences are not considered further in this study.

The stochastically simulated series of groundwater levels may be used to estimate the response of the aquifer to various hypothetical inputs of streamflow, which is in turn useful in the management of an aquifer.

2.3 APPLICATIONS OF TIME SERIES ANALYSIS IN HYDROLOGY

The use of time series analysis is common in many other fields, as well as in water resources. Numerous methods of time series analysis have been performed on various subsystems of the hydrologic system in order to determine their stochastic properties. Examples of application include open channel flow [14], studies of streamflow [22],[44]; precipitation [4]; monthly streamflow and precipitation analysis [49]; monitoring of water quality by using time independent data [55]; and water use studies [50].

The foundation of studies dealing with time series analysis in groundwater hydrology, to the best knowledge of the writer, has been laid by Eriksson [20]. This study departs from the classical investigations of groundwater aquifers which most often stress their deterministic properties. In this study, local river stages, air temperature, and precipitation was used in conjunction with groundwater levels from a glaciofluvial deposit. The investigation showed that strong persistence exists in all records except in records of precipitation. A stochastic model, based on the continuity equation, was developed to give estimates of the mean annual infiltration of the deposit, however no impulse response function of the system was presented. Therefore, the

model in its present state is not capable of simulating the groundwater level fluctuation process.

Eriksson, nevertheless, provided impetus for others to extend his work. In a stochastic analysis of groundwater level time series [37], a univariate model was presented. The study followed techniques of structural decomposition of series, and the resulting stochastic series was examined for dependence using Markov models of lag 1,2,3 including the periodicity in the autocorrelation coefficients. When the dependence effects were removed, the resulting process was tested for independence. Thus the original nonstationary process is reduced to a second order stationary and independent stochastic component process. The final independent stochastic residuals were analyzed to find their probability distribution functions. The results indicated that the model may be rightfully used for other groundwater level time series, and suggested the inclusion of precipitation process for a more detailed description of the system.

The above models were further developed by Rao et al. [45] in a landmark paper where the previous work was extended to bivariate processes of precipitation and groundwater levels. The study illustrated the construction and validation of the model and the results indicated that the groundwater level series can be successfully modelled by stochas-

tic models. A fourth-order autoregressive model was found adequate to describe the groundwater system. The model fitted to the research data revealed that 77% of the variation in the stochastic component of groundwater levels was explained. The authors caution that perhaps a strict casuality can not be completely inferred from statistical considerations alone. The model developed can be used to evaluate the time lag between precipitation and changes in groundwater levels and in evaluating the groundwater resources in a region.

The next step in formulating a multivariate model is to develop and validate, through residual testing, a model to evaluate the effect of various inputs of riverflow on water levels in adjacent wells through input-output analysis. This is another of the objectives of this thesis.

Several studies in the hydrologic literature have been performed to determine the impulse response function of linear systems via input-output analysis. Multiple frequency response analysis provided one way to establish a relationship between inputs and outputs of a multiple input-single output system set up for water quality monitoring [33]. Cross-correlation and cross-spectral analysis were employed in the analysis of rainfall and runoff to explain the relationship in this single input-single output system [9].

Similar techniques were used to identify the cause-effect interrelationships between runoff and sediment yield [48], again a single input-single output system was considered.

In subsurface hydrology there have also been some advances in specifying impulse response functions of groundwater flow systems.

Parametric linear programming [42] was shown capable of providing a stable and physically realizable response function. Transformation of the data into another domain provided one way to determine the response function. An approach based on fast Fourier transforms has been performed [21] using above procedure. The method of linear algebraic equations by matrix inversion [40] also gave satisfactory solution for response function. Similarly, various optimization techniques for coordinate functions, such as Laguerre functions [5], Walsh functions [19], proved to be another possible way to determine the response functions. Others suggested different approaches to obtain positive response function [16], to determine a stable solution by applying the average of response functions resulting from several events [40], to use the least squares criterion [19] in estimating the response function.

These methods allow the imposition of upper and lower bounds on the individual discrete response function, and

hence to smooth continuously the result along the computational process.

From the literature review it is evident that the response functions were studied mainly by deterministic models. Studies of Eriksson [20] and Rao et al. [45], used time series analysis, but did not formulate the problem for the multivariate case. Therefore, this thesis will extend these works into the multivariate case.

It, however, should be pointed out that the developed model in this study will not be compared directly with any other and different models existing in the literature. The main purpose of this thesis is to hypothesize and develop a systems model and lay the mathematical foundation for such a model. The practicality and relative usefulness of such a model will not be explored in this study.

2.4 OBJECTIVES

The objectives of this study are to use time series and systems analysis to:

1. Postulate and investigate the structure of a multivariate lumped linear dynamic model of groundwater level fluctuations;

2. Detect and remove the deterministic components in the system;
3. Investigate the structure of the stochastic component;
4. Develop a parsimonious lumped linear input-output model for groundwater level fluctuation;
5. Interpret the fluvial characteristics of the selected watersheds based on unit step and frequency response functions.

A handwritten mark, possibly a signature or a stylized letter 'L', located in the bottom right corner of the page.

Chapter III

RESEARCH DATA

3.1 WATERSHEDS SELECTED

For the numerical analysis two watersheds were used which are located in eastern Ontario (see Figure 3.1). The Castor River watershed lies east-southeast of Ottawa with approximately 782 km² total surface area drained by the Castor at the confluence of South Nation River. The second watershed is of the Carp River with a drainage area of approximately 332 km².

The watersheds have been gauged for groundwater level observation since 1972, other hydrological and meteorological observation stations are also operating nearby since the early twentieth century. The locations of the observation wells, riverflow, precipitation, temperature measuring stations are shown on the map (Figure 3.1) where the outline of each watershed is shown by dashed lines. Additional features of these watersheds relevant to this study are shown in Table 4.1.

FIGURE 3.1' LOCATION OF SELECTED WATERSHEDS

TABLE 3.1 FEATURES OF SAMPLING STATIONS IN SELECTED
WATERSHEDS

Watershed- Sampling station		Location	Period of records available for analysis
Castor	Metcalfe (well)	75°29'W 45°13'N	1972.11. - 1979.04.
	Russell (flow)	75°24'W 45°15'N	1970.01. - 1979.06.
	Int'l Airport (precipitation)	75°40'W 45°20'N	1970.01. - 1979.06.
	Int'l Airport (evapotranspiration)	75°40'W 45°20'N	1970.01. - 1979.06.
	Int'l Airport (temperature)	75°40'W 45°20'N	1972.11. - 1979.04.
	Int'l Airport (hourly sunshine)	75°40'W 45°20'N	1972.11. - 1979.04.
Carp	Carp 1 (well)	76°02'W 45°22'N	1972.11. - 1979.04.
	Carp 2 (well)	76°02'W 45°22'N	1972.11. - 1979.04.
	Kinborn (flow)	76°13'W 45°24'N	1971.01. - 1979.06.
	Woodlawn (precipitation)	76°09'W 45°28'N	1970.01. - 1979.06.
	Cent. Exp. Farm (evapotranspiration)	75°45'W 45°23'N	1970.01. - 1979.04.
	Cent. Exp. Farm (temperature)	75°45'W 45°23'N	1972.11. - 1979.04.
	Cent. Exp. Farm (hourly sunshine)	75°45'W 45°23'N	1972.11. - 1979.04.

3.2 HYDROGEOLOGY

3.2.1 Hydrology

The average annual temperature at Ottawa International Airport is approximately 5.5 C° and the average annual precipitation is approximately 876 mm. For about four months snow remains on the ground. Precipitation is exceeded by the potential evapotranspiration during four months of the year.

Examination of groundwater levels shows a yearly fluctuation pattern which is considered typical in most shallow aquifers generally found in the study area. Groundwater recharge begins at the end of March and during early April. This coincides with the large-scale flooding at the same time period. The groundwater levels usually reach their highest level between May and June. Throughout the summer groundwater levels gradually fall, reaching their lowest level around the end of September. During the fall there is some recharge before deep frost penetration occurs. Groundwater levels remain fairly static throughout the period of continuous frost.

Streamflows are generally unevenly distributed from year to year and during each year. Since there are no natural lakes, few swampy areas, and shallow till soil with poor drainage characteristics, flooding is the most prominent feature of the study area and is an annual occurrence during spring [36]. Lowest flows usually occur between May and October. During the period of deep frost on the ground surface, streamflows are retarded by continuous surface ice for about four months of the year.

3.2.2 Geology

The rate at which the water moves below the ground surface depends on the permeability of the successive layers of geologic formations. Sand and rock layers permit greater mobility than clays, silts, or glacial tills would. The characteristic topographic feature in the watersheds under study is the low relief ranging between 45.7 m and 122 m above mean sea-level.

The bedrock below the overburden consist mainly [12] of limestone of the Ottawa Formation with the occasional shale of Ordovician age. Surficial overburden deposits consist primarily of marine clays, silts, and glacial deposits of

till. The overburden deposits are generally thin (less than 15 m) and, because of their characteristics, are poorly permeable. This restricted permeability prevents any substantial percolation of water to deeper watertable zones. In contrast, silty sand and shallow water sand formations, which are moderately to highly permeable by nature, occur along the river banks [27].

The clay loam soils which cover much of the watersheds require extensive tile and surface drainage works because of their poor permeability. These drainage methods are likely to contribute to a higher rate of runoff and therefore to spring floods resulting from snowmelt. However, flooding is also (primarily) caused by low streambed gradients and small channel capacities.

3.3 DATA SELECTION

For the numerical analysis, the groundwater level data selected was collected by the Agrometeorology Research and Service, Land Resource Research Institute, Canada Department of Agriculture. From the records kept by this organization eight wells were selected for possible analysis. The data for five wells could not be used because a portion of the

record was missing thus, the remaining three wells were subjected to analysis. All the three wells had a period of record stretching from 1972 to 1979.

The frequency of the depth-to-water measurements for the selected wells in the Castor River and Carp River was five days of the month: 3, 8, 13, 18, 23, and EMO (i.e. the end-of-month) reading. This system results in a time series with a constant five day time interval.

The corresponding observations of riverflow rate of adjacent rivers (Castor and Carp), of precipitation, of temperature, of evapotranspiration, and of bright sunshine were also available.

Riverflow data was available from the Inland Waters Directorate, Water Resources Branch, Water Survey of Canada on punched cards for the Carp River (near Kinborn, Ontario) and Castor River (at Russell, Ontario).

Temperature data originates from the Atmospheric Environment Service, Environment Canada for Ottawa International Airport, and Agrometeorology Section, Research Branch, Central Experimental Farm, Canada Department of Agriculture.

Precipitation data was abstracted from "Monthly Meteorological Summaries", a monthly publication of Atmospheric Environment Service, Environment Canada for the following stations: Woodlawn, Ontario (Carp River watershed), Ottawa International Airport (Castor River watershed).

Daily meteorological observations at the Central Experimental Farm, Canada Department of Agriculture and at the Ottawa International Airport, Atmospheric Environment Service, Environment Canada were the basis for evapotranspiration and bright sunshine data.

It was possible to analyze observations for the common years between 1972-1979 for all variables. Thus, groundwater level and riverflow data represent levels/rates at these times, whereas temperature represents averages for these periods, and precipitation, evapotranspiration, and bright sunshine the collected amount for each interval.

All data were converted to metric units before being used for analytical purposes: flow was converted to m^3/s , precipitation and evapotranspiration to mm, and temperature to $^{\circ}\text{C}$.

Chapter IV

THEORETICAL DEVELOPMENTS

4.1 MODEL HYPOTHESIS

To represent the groundwater level fluctuation process it can be assumed that the watershed is a 'black box' or lumped system. The inputs to this system can be all factors which may trigger the process within the system yielding some output. Inputs such as precipitation, riverflow, evapotranspiration, temperature, and bright sunshine are the causative factors in the system, whereas the groundwater levels are outputs. The physical processes operating inside the system are not considered (i.e. "black box" formulation) and all properties of a watershed are reflected in a set of system parameters. These properties include topography, soil characteristics, land use, permeability, etc. In order to transform inputs into output, system parameters may be used to set up the characteristic function of the system.

A typical input-black box-output configuration is shown in Figure 4.1.

4.2 MATHEMATICAL REPRESENTATION OF THE PROCESS

A hypothesized mathematical relationship between inputs and output is as follows:

$$G_t = f(Q_t, Q_{t-1}, Q_{t-2}, \dots, Q_{t-m}, P_t, P_{t-1}, P_{t-2}, \dots, P_{t-r}, T_t, T_{t-1}, T_{t-2}, \dots, T_{t-f}, E_t, E_{t-1}, E_{t-2}, \dots, E_{t-n}, A_t, A_{t-1}, A_{t-2}, \dots, A_{t-s}) \quad \dots 4.1$$

where G_t is the groundwater level (cm) at the present time interval t ; $Q_t, Q_{t-1}, Q_{t-2}, \dots, Q_{t-m}$ are riverflow rates (m^3/s) at the present time interval t and preceding times $t-1, t-2, \dots, t-m$; $P_t, P_{t-1}, P_{t-2}, \dots, P_{t-r}$ are precipitation depths (mm) at the present time interval t and preceding times $t-1, t-2, \dots, t-r$; $T_t, T_{t-1}, T_{t-2}, \dots, T_{t-f}$ are temperature values ($^{\circ}C$) at the present time interval t and preceding times $t-1, t-2, \dots, t-f$; $E_t, E_{t-1}, E_{t-2}, \dots, E_{t-n}$ are evapotranspiration quantities (mm) at the present time interval t and at preceding times $t-1, t-2, \dots, t-n$; $A_t, A_{t-1}, A_{t-2}, \dots, A_{t-s}$ are bright sunshine amounts (hrs) at present time interval t , and preceding times $t-1,$

FIGURE 4.1 BLACK BOX REPRESENTATION OF GROUNDWATER LEVEL FLUCTUATION PROCESS

$t-2, \dots, t-s$. The time intervals, depending on the analysis, can be any discrete time intervals such as months, days, etc. For shorter notation the right hand side of Eq. 4.1 will be denoted by $f(*)$.

In the physical sense it can be argued that if the riverflow rate approaches a constant rate in the long run, such that $Q_t = Q_{t-1} = Q_{t-m} = Q_0$ then it follows that the requirement for this condition is such that $P_t \rightarrow P_0, T_t \rightarrow T_0, E_t \rightarrow E_0, A_t \rightarrow A_0, G_t \rightarrow G_0$ where $Q_0, P_0, T_0, E_0, A_0, G_0$ are steady state values of riverflow, precipitation, temperature, evapotranspiration, bright sunshine, and groundwater level quantities [51]. Therefore the use of Taylor-series expansion of multiple variables is applied to the steady state values of $Q_t, Q_{t-1}, Q_{t-2}, \dots, Q_{t-m}, P_t, P_{t-1}, P_{t-2}, \dots, P_{t-r}, T_t, T_{t-1}, T_{t-2}, \dots, T_{t-f}, E_t, E_{t-1}, E_{t-2}, \dots, E_{t-n}, A_t, A_{t-1}, A_{t-2}, \dots, A_{t-s}$, i.e. about Q_0, P_0, T_0, E_0 , and A_0 . Thus, the Taylor-series expansion applied to Eq. 4.1 is expressed in the following form:

$$\begin{aligned}
G_t = \sum_{k=0}^{\bar{p}} \left\{ \frac{1}{k!} \left[(Q_t - Q_0) \frac{\partial}{\partial Q_t} + (Q_{t-1} - Q_0) \frac{\partial}{\partial Q_{t-1}} + \right. \right. \\
+ (Q_{t-2} - Q_0) \frac{\partial}{\partial Q_{t-2}} + \dots + (Q_{t-m} - Q_0) \frac{\partial}{\partial Q_{t-m}} + \\
+ (P_t - P_0) \frac{\partial}{\partial P_t} + (P_{t-1} - P_0) \frac{\partial}{\partial P_{t-1}} + \\
+ (P_{t-2} - P_0) \frac{\partial}{\partial P_{t-2}} + \dots + (P_{t-r} - P_0) \frac{\partial}{\partial P_{t-r}} + \\
+ (T_t - T_0) \frac{\partial}{\partial T_t} + (T_{t-1} - T_0) \frac{\partial}{\partial T_{t-1}} + \\
+ (T_{t-2} - T_0) \frac{\partial}{\partial T_{t-2}} + \dots + (T_{t-f} - T_0) \frac{\partial}{\partial T_{t-f}} + \\
+ (E_t - E_0) \frac{\partial}{\partial E_t} + (E_{t-1} - E_0) \frac{\partial}{\partial E_{t-1}} + \\
+ (E_{t-2} - E_0) \frac{\partial}{\partial E_{t-2}} + \dots + (E_{t-n} - E_0) \frac{\partial}{\partial E_{t-n}} + \\
+ (A_t - A_0) \frac{\partial}{\partial A_t} + (A_{t-1} - A_0) \frac{\partial}{\partial A_{t-1}} + \\
+ (A_{t-2} - A_0) \frac{\partial}{\partial A_{t-2}} + \dots + (A_{t-s} - A_0) \frac{\partial}{\partial A_{t-s}} \left. \right] \cdot \\
\left. f(*) \right|_{Q_0, P_0, T_0, E_0, A_0} \Bigg\} = \sum_{k=0}^{\bar{p}} \phi_k
\end{aligned}$$

.....4.2

where ϕ_k is the quantity inside { };

$$f(*) \Big|_{Q_0, P_0, T_0, E_0, A_0}$$

are the partial derivatives of $f(*)$ evaluated at Q_0, P_0, T_0, E_0, A_0 , k is a dummy variable, and $\bar{p}$ is the order up to

which the expansion approximates the real process. The contributions for the first three k values of each term are listed below. The remaining terms in the Taylor-series expansion contribute negligibly towards the total when compared to the first three terms.

$$k = 0 ; \phi_0 = f_1(Q_0, P_0, T_0, E_0, A_0) = a_{00}$$

$$\begin{aligned} k = 1 ; \phi_1 = & (Q_t - Q_0) a_{10} + (Q_{t-1} - Q_0) a_{11} + \\ & + \dots + (Q_{t-m} - Q_0) a_{1m} + \\ & + (P_t - P_0) b_{10} + (P_{t-1} - P_0) b_{11} + \\ & + \dots + (P_{t-r} - P_0) b_{1r} + \\ & + (T_t - T_0) c_{10} + (T_{t-1} - T_0) c_{11} + \\ & + \dots + (T_{t-f} - T_0) c_{1f} + \\ & + (E_t - E_0) d_{10} + (E_{t-1} - E_0) d_{11} + \\ & + \dots + (E_{t-n} - E_0) d_{1n} + \\ & + (A_t - A_0) e_{10} + (A_{t-1} - A_0) e_{11} + \\ & + \dots + (A_{t-s} - A_0) e_{1s} \end{aligned}$$

where

$$\begin{aligned} a_{10} &= \left. \frac{\partial f(*)}{\partial Q_t} \right|_{Q_0, P_0, T_0, E_0, A_0} ; \\ &\vdots \\ a_{1m} &= \left. \frac{\partial f(*)}{\partial Q_{t-m}} \right|_{Q_0, P_0, T_0, E_0, A_0} \end{aligned}$$

$$b_{10} = \frac{\partial f(*)}{\partial P_t} \bigg|_{Q_0, P_0, T_0, E_0, A_0};$$

$$b_{1r} = \frac{\partial f(*)}{\partial P_{t-r}} \bigg|_{Q_0, P_0, T_0, E_0, A_0},$$

$$c_{10} = \frac{\partial f(*)}{\partial T_t} \bigg|_{Q_0, P_0, T_0, E_0, A_0};$$

$$c_{1f} = \frac{\partial f(*)}{\partial T_{t-f}} \bigg|_{Q_0, P_0, T_0, E_0, A_0},$$

$$d_{10} = \frac{\partial f(*)}{\partial E_t} \bigg|_{Q_0, P_0, T_0, E_0, A_0};$$

$$d_{1n} = \frac{\partial f(*)}{\partial E_{t-n}} \bigg|_{Q_0, P_0, T_0, E_0, A_0},$$

$$e_{10} = \frac{\partial f(*)}{\partial A_t} \bigg|_{Q_0, P_0, T_0, E_0, A_0};$$

$$e_{1s} = \frac{\partial f(*)}{\partial A_{t-s}} \bigg|_{Q_0, P_0, T_0, E_0, A_0},$$

$$\begin{aligned}
k = 2 ; \phi_2 = & \frac{1}{2} (Q_t - Q_0)^2 a_{20} + (Q_{t-1} - Q_0)^2 a_{21} + \\
& + \dots + (Q_{t-m} - Q_0)^2 a_{2m} + \\
& + \frac{1}{2} (P_t - P_0)^2 b_{20} + (P_{t-1} - P_0)^2 b_{21} + \\
& + \dots + (P_{t-r} - P_0)^2 b_{2r} + \\
& + \frac{1}{2} (T_t - T_0)^2 c_{20} + (T_{t-1} - T_0)^2 c_{21} + \\
& + \dots + (T_{t-f} - T_0)^2 c_{2f} + \\
& + \frac{1}{2} (E_t - E_0)^2 d_{20} + (E_{t-1} - E_0)^2 d_{21} + \\
& + \dots + (E_{t-n} - E_0)^2 d_{2n} + \\
& + \frac{1}{2} (A_t - A_0)^2 e_{20} + (A_{t-1} - A_0)^2 e_{21} + \\
& + \dots + (A_{t-s} - A_0)^2 e_{2s} + \\
& + \text{cross product terms}
\end{aligned}$$

where

$$a_{20} = \frac{\partial^2 f(*)}{\partial Q^2_t} ;$$

$$a_{21} = \frac{\partial^2 f(*)}{\partial Q^2_{t-1}} ;$$

...

$$a_{2m} = \frac{\partial^2 f(*)}{\partial Q^2_{t-m}}$$

$$b_{20} = \frac{\partial^2 f(*)}{\partial P^2_t} ;$$

$$b_{21} = \frac{\partial^2 f(*)}{\partial P^2_{t-1}} ;$$

...

$$b_{2r} = \frac{\partial^2 f(*)}{\partial P^2_{t-r}}$$

$$c_{20} = \frac{\partial^2 f(*)}{\partial T^2_t} ;$$

$$c_{21} = \frac{\partial^2 f(*)}{\partial T^2_{t-1}} ;$$

$$\vdots$$

$$c_{2f} = \frac{\partial^2 f(*)}{\partial T^2_{t-f}}$$

$$d_{20} = \frac{\partial^2 f(*)}{\partial E^2_t} ;$$

$$d_{21} = \frac{\partial^2 f(*)}{\partial E^2_{t-1}} ;$$

$$\vdots$$

$$d_{2n} = \frac{\partial^2 f(*)}{\partial E^2_{t-n}}$$

$$e_{20} = \frac{\partial^2 f(*)}{\partial A^2_t} ;$$

$$e_{21} = \frac{\partial^2 f(*)}{\partial A^2_{t-1}} ;$$

$$\vdots$$

$$e_{2s} = \frac{\partial^2 f(*)}{\partial A^2_{t-s}}$$

Upon substitution and rearrangements, Eq. 4.2 becomes:

$$\begin{aligned}
G_t = & a_{00} + \sum_{j=0}^m a_{1j} Q_{t-j} + \sum_{j=0}^r b_{1j} P_{t-j} + \\
& + \sum_{j=0}^f c_{1j} T_{t-j} + \sum_{j=0}^n d_{1j} E_{t-j} + \sum_{j=0}^s e_{1j} A_{t-j} + \\
& + \frac{1}{2} \sum_{j=0}^m a_{2j} Q^2_{t-j} + \frac{1}{2} \sum_{j=0}^r b_{2j} P^2_{t-j} + \\
& + \frac{1}{2} \sum_{j=0}^f c_{2j} T^2_{t-j} + \frac{1}{2} \sum_{j=0}^n d_{2j} E^2_{t-j} + \\
& + \frac{1}{2} \sum_{j=0}^s e_{2j} A^2_{t-j} + \text{cross product terms} + \\
& + \text{cubic terms and higher order terms}
\end{aligned}$$

....4.3

Examination of Eq. 4.3 reveals the following:

The coefficients a_{00} , a_{ij} , b_{ij} , ... etc. are dependent on the steady state values of Q_0 , P_0 , T_0 , E_0 , A_0 . Time variability in the values of these coefficients might be expected, then infinite sets of these values could exist. On the other hand, Eq. 4.3 could describe the behaviour of a lumped non-linear time invariant dynamic system if the assumption that these values are constants is assumed, as it will be for further analysis.

Another assumption is made that a non-linear relationship exists between G_t , Q_t , P_t , T_t , E_t , and A_t and it requires the inclusion of the second and higher order terms in Eq. 4.3.

It can be seen that a large number of functions exist which could be expanded in the form of Eq. 4.4. One of these functions, to represent the system behaviour, is given below:

$$G_t = J Q_t^{V_0} Q_{t-1}^{V_1} \dots Q_{t-m}^{V_m} P_t^{F_0} P_{t-1}^{F_1} \dots P_{t-r}^{F_r}$$

$$T_t^{R_0} T_{t-1}^{R_1} \dots T_{t-f}^{R_f} E_t^{D_0} E_{t-1}^{D_1} \dots E_{t-n}^{D_n}$$

$$A_t^{K_0} A_{t-1}^{K_1} \dots A_{t-s}^{K_s}$$

....4.4

in which J , V_i ($i=0, \dots, m$), F_i ($i=0, \dots, r$), R_i ($i=0, \dots, f$), D_i ($i=0, \dots, n$), K_i ($i=0, \dots, s$) are constants. It can also be seen that the model shown above is easily linearizable through logarithmic transformation, which in turn suggests that the non-linear approach can be shelved. Therefore, applying logarithmic transformation on Eq. 4.4 the model is linearized but retains its non-linear character.

Development of a dynamic relationship between inputs and output or modeling the system behaviour can be achieved by conducting system analysis on the linearized model. The model therefore can be written in terms of the logarithmic transformed variables as follows:

$$\begin{aligned}
\ln G_t = & \ln J + V_0 \ln Q_t + V_1 \ln Q_{t-1} + \dots + V_m \ln Q_{t-m} + \\
& + F_0 \ln P_t + F_1 \ln P_{t-1} + \dots + F_r \ln P_{t-r} + \\
& + R_0 \ln T_t + R_1 \ln T_{t-1} + \dots + R_f \ln T_{t-f} + \\
& + D_0 \ln E_t + D_1 \ln E_{t-1} + \dots + D_n \ln E_{t-n} + \\
& + K_0 \ln A_t + K_1 \ln A_{t-1} + \dots + K_s \ln A_{t-s}
\end{aligned}$$

....4.5

Simplification of Eq. 4.5 by using $\ln G_t = g_t$, $\ln J = j_0$, $\ln Q_t = q_t$, $\ln P_t = p_t$, $\ln T_t = t_t$, $\ln E_t = e_t$, $\ln A_t = a_t$ yields

$$\begin{aligned}
g_t = & j_0 + V_0 q_t + V_1 q_{t-1} + \dots + V_m q_{t-m} + \\
& + F_0 p_t + F_1 p_{t-1} + \dots + F_r p_{t-r} + \\
& + R_0 t_t + R_1 t_{t-1} + \dots + R_f t_{t-f} + \\
& + D_0 e_t + D_1 e_{t-1} + \dots + D_n e_{t-n} + \\
& + K_0 a_t + K_1 a_{t-1} + \dots + K_s a_{t-s}
\end{aligned}$$

....4.6

Thus Eq. 4.6 represents a discrete lumped linear time invariant dynamic model of the groundwater level fluctuation process, provided the constants V_i ($i=0, \dots, m$), F_i ($i=0, \dots, r$), R_i ($i=0, \dots, f$), D_i ($i=0, \dots, n$), K_i ($i=0, \dots, s$) and j_0 , are assumed to be independent of time.

4.3 ERROR ANALYSIS

The model should take into consideration the stochastic nature of the process and hence data and modeling errors as well. Errors in the observed data for input and output information include sampling errors associated with the non-homogeneous periodic sampling of groundwater levels, errors introduced by smoothing the sampled values to obtain groundwater level hydrographs; errors are also present in evapotranspiration, precipitation, temperature, riverflow etc. (i.e. input) data.

Since the above data are used to fit model parameters, the errors are thus transferred to the parameter values and an error component is introduced in the simulated output. The model is naturally only an approximation of the process, therefore by being an approximation introduces further errors.

The output values are contaminated by all of the above errors, which are referred to as the white noise series. Therefore to simulate the effect of this noise, it is conceivable to assume the errors as inputs to a linear system from which output is a noise component [10]. A diagram of the groundwater level fluctuation process in the presence of the noise term (N_t) is shown in Figure 4.2, where W_t , repre-

FIGURE 4.2 LUMPED LINEAR REPRESENTATION OF THE GROUNDWATER
LEVEL FLUCTUATION PROCESS

sents a white noise series. This W_t series consists of random values drawn from a fixed distribution. This distribution is said to be normal with zero mean and variance $\sigma_{W_t}^2$. Accordingly, Eq. 4.6 changes to:

$$\begin{aligned}
 g_t = & j_0 + V_0 q_t + V_1 q_{t-1} + \dots + V_m q_{t-m} + \\
 & + F_0 p_t + F_1 p_{t-1} + \dots + F_r p_{t-r} + \\
 & + R_0 t + R_1 t_{t-1} + \dots + R_f t_{t-f} + \\
 & + D_0 e + D_1 e_{t-1} + \dots + D_n e_{t-n} + \\
 & + K_0 a_t + K_1 a_{t-1} + \dots + K_s a_{t-s} + N_t
 \end{aligned}
 \tag{4.7}$$

where N_t is a white noise - error - series (having mean zero, homogeneous variance $\sigma_{N_t}^2$) and the effect of errors mentioned previously are assumed to be incorporated in this term. Since the W_t series is not correlated with the input series, it could be auto-correlated and can be modeled by a linear process from the ARIMA family [10].

4.4 MULTIPLE LINEAR REGRESSION ANALYSIS

It was hypothesized in Section 4.2, that the logarithms of groundwater levels are to be linearly related to the logarithms of riverflow, of temperature, of evapotranspiration, of precipitation, and of bright sunshine.

Since there is a relatively large number of independent variables, i.e. precipitation, riverflow, temperature, evapotranspiration, bright sunshine, it is necessary to estimate their relative contributions in influencing the groundwater level fluctuation process. Thus, a multiple linear regression was run between the five-daily values of the independent variables and five-daily groundwater levels - as dependent variable - at simultaneous five days (December through November). As it was stated in Section 4.3, riverflow rates obtained at the gauging stations selected for each watershed have five-daily values, therefore temperature values are averaged over five day periods. Accumulated amounts of precipitation, bright sunshine, and evapotranspiration data are justified in view of published studies, such as in [20] where it was shown that collected amounts for each interval could be used for stochastic analysis and yielded satisfactory results.

The results of the multiple linear regression are shown in Table 4.1. The results suggest that the riverflow rates contribute a higher percentage towards explaining the variance in the five-daily groundwater levels than do the other independent variables.

Therefore the riverflow rates can be considered the main causative factor in the groundwater level fluctuation

TABLE 4.1 RESULTS OF MULTIPLE LINEAR REGRESSION ANALYSIS
(log transformed variables)

r_{12} = correlation coefficient between [1] and [2]
 r_{13} = correlation coefficient between [1] and [3]
 r_{14} = correlation coefficient between [1] and [4]
 r_{15} = correlation coefficient between [1] and [5]
 r_{16} = correlation coefficient between [1] and [6]

		Castor	Watershed Carp 1	Carp 2
Simple linear correlations	r_{12}	- 0.8156	- 0.7244	- 0.7258
	r_{13}	- 0.0018	- 0.0110	- 0.0013
	r_{14}	0.1489	- 0.0066	- 0.0563
	r_{15}	0.4282	0.1421	0.0788
	r_{16}	0.4578	0.2056	0.1375
% variance due to	2	66.52	52.47	52.68
	3 (given 2)	0.13	0.01	0.01
	4 (given 2,3)	0.54	1.23	0.83
	5 (given 2,3,4)	2.42	0.01	0.02
	6 (given 2,3,4,5)	0.11	0.32	3.10

where

[1] - groundwater level
 [2] - riverflow
 [3] - precipitation
 [4] - bright sunshine
 [5] - temperature
 [6] - evapotranspiration

process. It is understood that the precipitation is the primary source of groundwater, which may percolate the soil directly to the groundwater table or may enter rivers and percolate from these streams to the groundwater. Since the groundwater has the lowest priority on the water from precipitation; interception, depression storage, and soil moisture must be satisfied before any large amount of water can percolate to the groundwater table. Therefore water from rivers is the main source of recharge for the groundwater in geologically favourable terrains, as is the case in this study. However, precipitation might be the main factor in other conditions [45].

The response of the groundwater is dependent on storage characteristics of the watershed [31,18]. These are usually time dependent phenomena [3]: the groundwater level series may lag behind the riverflow series.

It is suggested by the results of the multiple linear regression, (Table 4.1) that the riverflow is transferred to the aquifer and the groundwater level fluctuation process is shaped by riverflow irrespective of precipitation effects. Such results were expected for studied watersheds as was suggested in Chapter 4.

4.5 HYPOTHESIZED INPUT-OUTPUT MODEL

In view of the results of the multiple linear regression, the model represented by Eq. 4.7 can be reduced into (i.e. single input-output system) the following form:

$$g_t = j_0 + V_0 q_t + V_1 q_{t-1} + \dots + V_m q_{t-m} + N_t \quad \dots 4.8$$

where, as before $g_t = \ln G_t$; $q_t = \ln Q_t$; $q_{t-1} = \ln Q_{t-1}$ (Q_t = riverflow, G_t = groundwater levels), and so on. By employing a delay parameter b , Eq. 4.8 can be written as

$$g_t = j_0 + V_0 q_{t-b} + V_1 q_{t-b-1} + \dots + V_m q_{t-b-m} + N_t \quad \dots 4.9$$

Introducing the backshift operator B [10], then the following changes can be made:

$$\begin{aligned} q_{t-1} &= Bq_t \\ q_{t-b} &= B^b q_t \\ q_{t-m} &= B^m q_t \\ q_{t-b-m} &= B^b q_{t-m} = B^{b+m} q_t \end{aligned}$$

By substituting above terms in Eq. 4.9, gives

$$g_t = j_0 + (V_0 + V_1 B + V_2 B^2 + \dots + V_m B^m) q_{t-b} + N_t \quad \dots 4.10$$

which can be simplified to yield

$$g_t = j_0 + V(B) q_{t-b} + N_t \quad \dots 4.11$$

where $V(B) = (V_0 + V_1 B + V_2 B^2 + \dots + V_m B^m)$ is called the impulse response function of the system.

It is to be noted, that since the terms below are equivalent

$$V(B) q_{t-b} = V(B) B^b q_t$$

therefore when series g_t , q_t are standardized for their means or are expressed in terms of the deviations from their respective means $\bar{g}$, $\bar{q}$ (j_0 terms cancel out) Eq. 4.11 becomes:

$$(g_t - \bar{g}) = V(B) B^b (q_t - \bar{q}) + N_t \quad \dots 4.12$$

Using $(q_t - \bar{q}) = y_t$ and $(q_t - \bar{q}) = x_t$, Eq. 4.12 can be written as:

$$y_t = V(B)B^b x_t + N_t$$

or

$$y_t = V(B)x_{t-b} + N_t$$

....4.13

Two systems are represented by the model described in Eq. 4.13, namely a deterministic linear dynamic system with x_t as input, and y_t as output, and a stochastic linear dynamic system with white noise W_t as input and noise N_t as output (Figure 4.3).

The deterministic input-output model in terms of a more parsimonious form can be written as

$$\begin{aligned} y_t - \delta_1 y_{t-1} - \dots - \delta_p y_{t-p} &= \\ &= \omega_0 x_{t-b} - \omega_1 x_{t-b-1} - \dots - \omega_n x_{t-b-n} \end{aligned}$$

....4.14

or

$$\delta(B)y_t = \omega(B)x_{t-b} = \omega(B)B^b x_t$$

....4.15

in which,

FIGURE 4.3 LUMPED LINEAR REPRESENTATION OF THE GROUNDWATER
LEVEL FLUCTUATION PROCESS AS HYPOTHESIZED

$$\delta(B) = (1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_p B^p)$$

$$\omega(B) = (\omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_n B^n)$$

i.e. $\delta(B)$ is a polynomial of degree p and $\omega(B)$ is a polynomial of degree n , such that

$$\frac{\omega(B)B^b}{\delta(B)} = V(B)$$

....4.16

Eq. 4.15 in terms of g_t and q_t takes the form

$$\delta(B)g_t = \theta_0 + \omega(B)q_{t-b} + N_t$$

....4.17

where θ_0 is a constant function of $\bar{g}$, $\bar{q}$, $\delta(B)$, $\omega(B)$.

In most hydrologic studies the problem of non-commensurate units on the variables has been handled by standardizing them through a linear transformation. Such an approach takes into account the non-linear effects due to varying levels of input, and presumes that the seasonal fluctuation in the lag one correlation coefficient can be ignored. Therefore, if the input q_t and output g_t series are standardized (for both means and variances) then

$$X_t = \frac{q_t - \bar{q}}{\sigma_q}$$

$$Y_t = \frac{q_t - \bar{q}}{\sigma_g}$$

where σ_q and σ_g are the standard deviations of input and output series respectively. In this case, Eq. 4.12 through 4.16 would run parallel for the X_t and Y_t series with parameters

$$\begin{aligned} &V_0, V_1, V_2, \dots, V_m \\ &\delta_1, \delta_2, \delta_3, \dots, \delta_p \\ &\omega_0, \omega_1, \omega_2, \dots, \omega_n \end{aligned}$$

which will have different values. The operators for X_t and Y_t series will assume the following relationships:

$$\begin{aligned} \overline{V(B)} &= \frac{\sigma_q}{\sigma_g} V(B) \\ \overline{\delta(B)} &= \sigma_g \delta(B) \\ \overline{\omega(B)} &= \sigma_q \omega(B) \end{aligned}$$

Then,

$$\overline{V(B)} = \frac{\overline{\omega(B)}}{\overline{\delta(B)}}$$

....4.18

Substituting and rearranging in Eq. 4.18

$$\overline{V(B)} = \frac{\frac{\sigma_q}{\sigma_g} (B)}{\frac{\sigma_g}{\sigma_g} (B)} = \left[\frac{\sigma_q}{\sigma_g} \omega(B) \right] \frac{1}{\delta(B)}$$

will yield the following:

$$\overline{V}(B) = \omega(B) \left[\frac{1}{\frac{\sigma_q}{\sigma} \sigma(B)} \right]$$

....4.19

where $V(B)$, $\delta(B)$, $\omega(B)$ are operators for the standardized (for their means) series, q_t and g_t , while $\overline{V}(B)$, $\overline{\delta}(B)$, and $\overline{\omega}(B)$ are the operators for the q_t and g_t standardized (for both means and variances) series.

The hypothesized model for groundwater level fluctuation process on a five-daily basis is therefore given in Eq. 4.17; the computational version is described in Eq. 4.14. For practical purposes, physical systems can most likely be modeled (by using multiple linear regression) with orders p and $n = 2$. A diagram of the hypothesized groundwater level fluctuation process is shown in Figure 4.3.

The techniques for the process identification, parameter estimation, and model hypothesis application to the groundwater level fluctuation process are discussed in the subsequent chapters.

Chapter V

METHODS OF ANALYSIS

5.1 INTRODUCTION

The tools available for describing the behaviour of linear processes are the correlation analysis and the spectral analysis. Correlation functions are useful in analyzing the time series in the time domain. Usually the periodicities contained in the data can best be explained by a frequency domain analysis, that is by spectral analysis. The spectral density function determines the explained variance of the time series contributed to various frequencies, and its mathematical development suggest a close relationship between the correlation function and the spectral density function. Some difficulties may arise in interpreting the correlation function, if neighbouring lags exhibit high correlations and the spectral density function does not show this kind of dependency at a discrete set of frequencies. In the statistical sense this latter property is helpful indeed

in interpreting spectral estimates, and seems inferior to the correlation function. Because both functions are mathematically equivalent, each contributes valuable information (although via different approaches) on the stochastic process in question, and should be regarded as complementaries in the system identification and analysis processes.

Correlation and spectral analyses were subjected to many investigations in which the question often raised was on the necessity of stationarizing. Authors, such as [52,32], do favour stationarizing, however, numerous others, such as [2,35,47] do not. This clearly shows that there is no one approach which should be followed in all situations. Therefore it was decided that both ways be pursued in developing a successful parsimonious model of systems behaviour in this study.

5.2 INPUT-OUTPUT ANALYSIS

The riverflow as input (q_t) and groundwater level as output (g_t) series are assumed to be trend and jump free and also to contain periodic (nonstationary) and stochastic (stationary) components. The following linear additive model therefore is assumed:

$$q_t = \bar{q}_t + \alpha_t \quad \dots 5.1$$

$$g_t = \bar{g}_t + \beta_t \quad \dots 5.2$$

where $\bar{q}_t$ is the deterministic periodic component of the input series, $\bar{g}_t$ is of the output series. The stochastic components α_t and β_t (for input and output series respectively) are viewed as statistically independent of deterministic components.

Equations 5.1 and 5.2 can be rewritten accordingly to the assumption that linear linkage models are functioning between the stochastic and the deterministic pairs. Thus the equations

$$\bar{g}_t = V_2(B) \bar{q}_t \quad \dots 5.3$$

$$\beta_t = V_3(B) \alpha_t + e_t \quad \dots 5.4$$

represent two systems with system impulse response functions $V_2(B)$, $V_3(B)$, and in each one a delay parameter is included. In the model given by Eq. 5.3, the input is $\bar{q}_t$, whereas the output is $\bar{g}_t$. However, Eq. 5.4 contains the noise, e_t , together with α_t , as input, and β_t , as output. If the noise

term is statistically independent of α_t , and hence of q_t , it could be autocorrelated and could be represented by an ARIMA [10] type model.

However, after substituting Eq. 5.3 and 5.4 in Eq. 5.2, and assuming that there is no significant difference between the nonstationary and stationary input-output pairs, then in this case $V_2(B) = V_3(B)$, and one may obtain:

$$g_t = V_2(B) (\bar{q}_t + \alpha_t) + e_t \quad \dots 5.5$$

or because of $\bar{q}_t + \alpha_t = q_t$, and $V_2(B) = V(B)$, one may arrive at:

$$g_t = V(B)q_t + e_t \quad \dots 5.6$$

which is a system with nonstationary input (q_t) and output (g_t) series. Also included are the noise term, e_t , and the system impulse response function containing a delay parameter. The system described above was also obtained in Eq. 4.13, but through a different approach. The proposed model above is possibly inadequate in some circumstances, such as when feedback is present between the output and the input, however, it is assumed in this study that no feedback occurs.

5.3 INPUT-OUTPUT ANALYSIS IN THE FREQUENCY DOMAIN

Using Fourier transformation, one may obtain the equivalent system in the frequency domain. It is as follows:

$$q(f) = \overline{q}(f) + \alpha(f) \quad \dots 5.7$$

$$g(f) = \overline{g}(f) + \beta(f) \quad \dots 5.8$$

$$\overline{g}(f) = H_2(f)\overline{q}(f) \quad \dots 5.9$$

$$\beta(f) = H_3(f)\alpha(f) + e(f) \quad \dots 5.10$$

then

$$\overline{g}(f) = H_2(f)\overline{q}(f) + H_3(f)\alpha(f) + e(f) \quad \dots 5.11$$

and similarly as before, if $H_2(f) = H_3(f) = H(f)$, then Eq. 5.11 becomes

$$g(f) = H(f)q(f) + e(f) \quad \dots 5.12$$

Fourier transforms of q_t , $\overline{q}_t$, g_t , $\overline{g}_t$, α_t , β_t , e_t , and of $v_2(B)$, $v_3(B)$, $v(B)$ take the form of $q(f)$, $\overline{q}(f)$, $g(f)$, $\overline{g}(f)$,

$e(f)$, $\alpha(f)$, $\beta(f)$, $H_2(f)$, $H_3(f)$, $H(f)$ in which the latter three components are the frequency response functions. These functions' polar forms involves the gain and phase functions, as expressed below

$$H(f) = G(f) e^{j\phi(f)} \quad \dots 5.13$$

$$H_2(f) = G_2(f) e^{j\phi_2(f)} \quad \dots 5.14$$

$$H_3(f) = G_3(f) e^{j\phi_3(f)} \quad \dots 5.15$$

In Eqs. 5.13 through 5.15 $G(f)$, $G_2(f)$, $G_3(f)$ represent the gain functions, while $\phi(f)$, $\phi_2(f)$, $\phi_3(f)$ stand for the phase functions.

The gain function measures the ultimate amplification of a unit step change after the system has been allowed to settle down to its new steady state value, and is interpreted as a regression coefficient of one series upon the other one at a given frequency. The phase function gives the time delay needed for changes in one series to be reflected by similar changes in the other series.

Since the covariance functions $\gamma_{11}(k)$ and $\gamma_{12}(k)$ are known for all values of lag k , the Wiener-Hopf integral equation yields the impulse response functions. The in-

tegral equation replaced by an equivalent summation as shown in Eq. 5.16.

$$\gamma_{12}(k) = \sum_{i=0}^{\infty} V_i \gamma_{11}(k-i) \quad \dots 5.16$$

where V_i ($i=1,2,\dots$) are the impulse response functions, $\gamma_{11}(k)$ is the auto-covariance of the input (q_t) process, $\gamma_{12}(k)$ is the cross-covariance between input and output (g_t) series. However, correlation functions $\rho_{11}(k)$ and $\rho_{12}(k)$ could be employed in place of covariance functions once both input and output series are standardized for their respective means and variances.

Transforming the cross covariance function in Eq. 5.16 gives:

$$P_{12}(f) = H(f)P_{11}(f) \quad \dots 5.17$$

Therefore, the frequency response function may be obtained from

$$H(f) = \frac{P_{12}(f)}{P_{11}(f)} \quad \dots 5.18$$

where $P_{11}(f)$ is the auto power spectrum, whereas $P_{12}(f)$ stands for the cross power spectrum. Rewriting Eq. 5.18 in polar form yields:

$$H(f) = \frac{A_{12}(f)}{P_{11}(f)} e^{jF(f)} \quad \dots 5.19$$

where $A_{12}(f)$ is the amplitude and $F(f)$ is the phase functions. It is clearly seen that Eq. 5.19 embodies the gain and phase functions of the linear system, such that

$$G(f) = \frac{A_{12}(f)}{P_{11}(f)}$$

and

$$F(f) = \phi(f).$$

On pursuing further the aspects of autocovariance of the output, the result can be expressed as

$$P_{22}(f) = G^2(f) P_{11}(f) + P_{zz}(f) \quad \dots 5.20$$

which includes the power spectrum of the noise component. A term known as squared coherency (κ^2) between the input and output at frequency (f) reflects the level of noise, such that, if the output power spectrum is identical to the noise

power spectrum, the squared coherency amount to zero. This can be expressed as:

$$P_{ZZ}(f) = P_{22}(f) (1 - \kappa^2(f)) \quad \dots 5.21$$

where

$$\kappa^2(f) = \frac{A_{12}^2(f)}{P_{11}(f) P_{ZZ}(f)} \quad \dots 5.22$$

Comparison of Eqs. 5.20 and 5.22 shows that $P_{22}(f)$ can be eliminated, therefore

$$\kappa^2(f) = \frac{1}{1 + P_{ZZ}(f)/G^2(f)P_{11}(f)} \quad \dots 5.23$$

where $P_{ZZ}(f)$, is as before, the power spectrum of the noise term.

Eqs. 5.18 through 5.23 provide a means to determine the gain, phase, coherency, and noise spectra for each system represented by Eqs. 5.9, 5.10, 5.12. The individual autocovariance $\gamma(k)$, and autocorrelation $\rho(k)$, functions are estimated by the sample autocovariance estimators $c_{11}(k)$ for inputs and $c_{22}(k)$ for outputs, and by the sample correlograms $r_{11}(k)$ for inputs and $r_{22}(k)$ for outputs respective-

ly. Similarly, the cross covariance functions $\gamma_{12}(k)$ are estimated by sample estimators $c_{12}(k)$, the cross correlation functions $\rho_{12}(k)$ and $\rho_{21}(k)$, for both positive and negative, are estimated by the sample cross correlograms $r_{12}(k)$. Smoothed sample spectral estimators for the power spectrums $P'_{11}(f)$, $P'_{22}(f)$, $P'_{ZZ}(f)$; amplitude function, $A'_{12}(f)$; phase function, $F'(f)$; gain function, $G'(f)$; and coherence function, $\kappa'(f)$, are needed for the power spectra, amplitude, phase, gain, and coherence functions. Documentation of properties and estimation techniques are in abundance in hydrologic literature, the most standard being [12,34,59].

5.4 COMPUTATIONAL PROCEDURE

The analysis in Section 5.2 points out that when either the impulse response or frequency response functions are not different from each other significantly, then the nonstationary input-output series could be used throughout the systems analysis procedure. On the other hand, when the difference between the impulse response or frequency response functions is significant, then separate linkage models are needed for the deterministic and stochastic input-output pairs.

The course of the analysis is as follows. Periodic and stochastic series are to be formed through decomposition of the original nonstationary input-output series. Several ways can be chosen to achieve this and they are to be discussed later in this chapter. The nonstationary and newly formed stochastic input-output series are then subjected to spectral analysis, to determine whether the frequency response functions for nonstationary and stochastic input-output series are identical or different from each other. Spectral analysis is a convenient method to extract useful information on system behaviour without sophisticated parameter estimation.

Estimates of the power spectrum of the noise term, for both the nonstationary input-output pairs and stochastic pairs, are also obtained through the spectral analyses. These power spectrum estimates should be identical, if the frequency response functions were found to be identical through the above spectral analyses, which incidentally would also reveal information on whether or not the noise component is a white noise. Should the power spectrum of the noise component fluctuate about a constant value, it is a white noise term; otherwise the noise term may belong to the Box and Jenkins' [10] ARIMA family.

For both nonstationary and stochastic input-output series gain, phase, coherence, and noise spectra should be computed and inspected whether any significant difference occurs between each other. Significant differences suggest separate models for deterministic and stochastic pairs, in-difference would imply that the use of nonstationary input-output series are suitable. Numerous tests can be applied in testing hypotheses that the above spectra are coalescent for both nonstationary and stochastic input-output pairs.

Gain and phase spectra may reveal some aspects on the memory content and delay parameter of the input-output process. Time domain analysis should be carried out to ascertain the nature of the delay parameter, impulse response function, and the orders of input-output operators. If trends are encountered in gain and phase functions, it can be expected that the system has memory and should be provided for in the system identification. It is known that the basic information in time domain analysis is contained in the auto- and crosscovariance and correlation functions.

If the analyses of the gain, noise, phase spectra indicate that separate models are necessary to describe the process properly, then the delay parameter is to be determined for both models separately from the cross correlogram. The use of multiple linear regression analyses for both deter-

ministic and stochastic input-output pairs may help to determine the required models. Harmonic analysis for deterministic pairs and other methods proposed for stochastic pairs are further ways of estimating the essential parameters of the respective systems [10].

If, however, a single model could be produced by using the nonstationary input-output series, then the delay parameter could be obtained from the cross correlogram. As it was demonstrated [8], the delay parameter is the individual lag value at which the cross correlation function peaks in the cross correlogram. Orders of operators of input-output series can be determined through a linear least squares algorithm used in multiple regression analyses.

Transfer function models are used to evaluate the noise component. The noise term in turn is subjected to auto correlation analysis in order to find estimates of parameters associated with ARIMA models. If the noise component is identified as white noise, then a linear least squares technique would provide sufficient estimates of parameters for the ARIMA model chosen.

The model structure for input-output and noise terms thus could be identified, parameters estimated, model adequacy checked. This latter verification consists of testing the residuals obtained from fitting the data to the model.

Tests, such as auto- and cross correlation checks are regarded as relevant to show any discrepancies of the model in question. As long as no satisfactory characteristics of residuals attained, model reconstruction is necessary.

5.4.1 Periodic component

There are several options in use today to remove the periodic component in monthly, daily (or in any other selected time interval) hydrologic series. In the following, a brief description of few available methods are presented, where μ_t represents the mean, and σ_t denotes the standard deviation of the variable at any subscripted time period t .

A time series could be standardized by using the simple formula:

$$z_t = \frac{x_t - \mu_t}{\sigma_t} \quad \dots 5.24$$

where z_t is the standardized series with zero mean and unit variance, x_t is the original series, μ_t and σ_t are as described before. The standardized series still retains the random and nonrandom properties of the original series.

It was also determined [62] that all the important periodic components could be removed by subtracting the computed mean values for each time period from the original series, such that in case of monthly series: $T = 1, 2, \dots, 12$; provided that σ_t remains unchanged for all time periods. Thus,

$$z_t = x_t - \bar{x}_t$$

.....5.25

removes the periodicity.

A nonstationary process could be transformed to a stationary process by introducing a seasonal differencing operator [10]. Such a differencing is performed in an important class of models, called the autoregressive-integrated-moving average (ARIMA) process. In this very powerful models, for example, a monthly series requires 12 monthly differencing, whereas for a daily sequence 365 days differencing is necessary.

Natural conditions of a watershed are changing with time, causing nonstationary time series. However, these variations might be negligible over a short period of time. Also, it is known that a mean value of an observed time series is not constant but changes in time. So does the stan-

dard deviation, but the coefficient of variation remains constant, which implies that different parts of the time series behave alike. Therefore the periodicity can be removed with the transformation

$$z_t = \frac{x_t}{\sigma_t} \quad \dots 5.26$$

where the z_t series is stationary stochastic such that the mean is equal to μ_t / σ_t and the variance is unity.

The hydrologic variables (e.g. μ_t , σ_t , etc.) exhibit very pronounced regular fluctuations, which periodicities can be mathematically described by a series of sine and cosine functions. Such a harmonic analysis results in a model which may involve up to six harmonics in monthly, and up to twelve harmonics in daily series. After twelve harmonics, however, significant peaks are not common occurrences [61].

5.4.2 Linear system identification

As it was postulated in Section 4.5, if a model could be developed by considering nonstationary input (q_t) and output (g_t) series, then multiple linear regression provides one way of determining the orders (p, n) associated with the operators of input and output series. The system can be represented as (see Eq. 4.14)

$$\begin{aligned}
 g_t = & \delta_1 g_{t-1} + \delta_2 g_{t-2} + \dots + \delta_p g_{t-p} + \omega_0 g_{t-b} + \\
 & + \omega_1 g_{t-b-1} + \omega_2 g_{t-b-2} + \dots + \omega_n g_{t-b-n} + e_t
 \end{aligned}
 \quad \dots 5.27$$

where b is the delay parameter and e_t is the noise component. A linear least squares algorithm is used in the multiple linear regression analysis for parameter estimation in Eq. 5.27 for $\delta_i (i = 1, 2, \dots, p)$ and $\omega_i (i = 0, 1, 2, \dots, n)$ [18]. The value of the delay parameter can be determined from the cross correlogram; as it was stated before, the lag at which the cross correlation function peaks is the numerical value of the delay parameter.

It is also necessary to point out, that the noise series should be uncross-correlated with the input series in order to ensure the validity of Eq..5.27. The noise series, however, might be modeled by available methods [31] and might also be autocorrelated.

The orders of operators for input-output series can be initially estimated by procedures outlined in [10]. Residuals obtained through multiple linear regression analysis with orders $p=0$, $n=0$ are the estimates of the noise component. A cross correlation analysis may shed light on whether or not the input and noise series are cross correlated at present and past lags. If they are uncross-correlated, then the postulated model with $p=0$, $n=0$ is adequate and the noise

term can be modeled. If the noise and input series are found to be cross correlated, higher orders should be proposed until the condition of uncross-correlation is satisfied. Since at each step residuals are computed, and crosscorrelation analysis conducted, it is likely that a step will be encountered where no cross correlation exists between input and noise series. Thus, the assumed orders at that step will represent proper orders of p and n .

Most physical systems, however, can be modeled with p and $n=2$ [10]. Therefore, some other model structure should be considered if the noise and input series show cross correlation tendencies at values higher than p and $n=2$.

5.4.3 Harmonic analysis

There is the possibility, as it was noted in Section 5.4, that two separate models are needed to describe the process in question. The original series then should be decomposed into deterministic periodic pairs and stochastic stationary pairs. Hence the two separate models. The delay parameter can be determined from the cross correlogram for each system, and harmonic analysis may furnish answers in modeling the deterministic periodic pairs.

Suppose, that $\overline{q}_t$ and $\overline{g}_t$ are the periodic components in the input-output series respectively. For finite, even, and equally time spaced series, Fourier series expansion is viable such that

$$\overline{a}_t = a_0 + \sum_{k=1}^m (A_n \cos f_0 kt + B_n \sin f_0 kt) \quad \dots 5.28$$

in which A_n and B_n are the Fourier coefficients. However, Fourier series containing only cosine series with phase angle can be represented as

$$\overline{q}_t = a_0 + 2 \sum_{k=1}^m |a_k| \cos(f_0 kt + \psi_k) \quad \dots 5.29$$

and

$$\overline{g}_t = b_0 + 2 \sum_{k=1}^m |b_k| \cos(f_0 kt + \phi_k) \quad \dots 5.30$$

where m is the number of harmonics, f_0 is the fundamental frequency, $|a_k|$ and $|b_k|$ are absolute values of amplitudes of the k -th harmonic; ψ_k is the input phase of the k -th harmonic, ϕ_k is the output phase of the k -th harmonic; a_0 and b_0 are the mean of the input and output series respectively. It is also worth noting that

$$f_0 = \frac{2\pi}{N}$$

and that the procedures for estimating above characteristics are summed up by [59,60].

The response of the system to a sinusoidal input of the type

$$\overline{q}_t = |a_k| \cos(f_0 kt + \psi_k)$$

can be represented by

$$\overline{g}_t = |a_k| |\epsilon_k| \cos(f_0 kt + \psi_k + \theta_k)$$

It is to be noted that the output will have the same harmonic frequency but will include some gain, and corresponding phase factors. Therefore, $\overline{g}_t$ can be related to $\overline{q}_t$ directly by

$$\overline{g}_t = \epsilon_0 a_0 + 2 \sum_{k=1}^m |\epsilon_k| |a_k| \cos(f_0 kt + \psi_k + \phi_k) \quad \dots 5.31$$

where $|\epsilon_k|$ represents the gain and θ_k stands for the phase shift. Comparison of Eqs. 5.30 and 5.31 reveals that

$$\epsilon_k = \frac{b_k}{a_k} \quad \dots 5.32$$

also

$$\epsilon_0 = \frac{b_0}{a_0}$$

.....5.33

and

$$\theta_k = \phi_k - \psi_k$$

.....5.34

where $k=1,2,\dots,n$.

However, if the input and output series are compensated for means, then in Eq. 5.33 $a_0 = 0$, $b_0 = 0$ and consequently $\epsilon_0 = 0$. Therefore, the estimates of amplitudes, phase angles, and fundamental frequency can determine the gain, and phase shift at each harmonic frequency.

The periodic components $\overline{q_t}$, $\overline{g_t}$ thus can be removed by fitting Fourier series with a finite number of harmonics to the nonstationary input-output series q_t and g_t . The periodogram analysis would provide a means to estimate the number of harmonics involved in the harmonic analysis. In practice, however, the number of harmonics used in the analysis is determined by

$$m_{\max} = \frac{N}{2}$$

where N is the number of data points in the series.

5.5 MODEL DIAGNOSTIC CHECKING

It is necessary to apply diagnostic tests after model specification and estimation of its parameters to see whether there is any serious inadequacy in the postulated model. The model consists of two parts, namely the transfer function model

$$\delta(B)g_t = \omega(B)q_{t-b}$$

....5.35

and the noise model

$$\phi(B)(1-B)^d e_t = \theta(B) a_t$$

....5.36

The noise model has an autoregressive operator of order p , represented by $\phi(B)$; a moving average operator of order q , represented by $\theta(B)$; and a differencing operator of order d , represented by $(1-B)^d$, in Eq. 5.36.

Useful types of diagnostic checks include autocorrelation and cross correlation techniques to assess the statistical significance of departures from zero of both auto- and cross-correlations. This would provide for detecting lack of fitness.

One method employs the residuals obtained by subtracting estimated values from observed ones. If the residuals, obtained from fitting a model, correspond to the characteristics of a white noise process, then it is indicated by near zero autocorrelations among the residuals. Therefore the inadequacy of the noise model is indicated by autocorrelation of residuals, which are essentially estimates of the white noise series.

If the transfer function model was inadequately modeled, it would be shown by crosscorrelation check. If, on the other hand, the noise model is inadequate the autocorrelation check would indicate this fact.

5.5.1 Cross correlation check

The inadequacy of the transfer function model is indicated by cross correlation between inputs and residuals. This is the test to be used when checking the cross correlation between the inputs and white noise series if a multiple linear regression algorithm was used to identify the transfer function model.

The cross correlation can be regarded as normally distributed with zero mean for two uncross-correlated series. The standard error associated with the cross correlation is given below:

$$E_{r_{1,2}} = \frac{1}{\sqrt{n-k}}$$

....5.37

where n and k are the sample size and lag respectively. Confidence limits for the cross correlation are determined according to Eq. 5.38.

$$CL_{95\%} = 0 \pm 1.96 \frac{1}{\sqrt{n-k}}$$

....5.38

5.5.2 Auto correlation check

The following are the brief descriptions of auto correlation tests that can be applied on the white noise series.

5.5.2.1 Testing the power spectrum

Periodicity in residuals may be detected by frequency domain checks. The power spectrum of the residuals are independent of frequency, if the residuals are belong to a pure white noise process. In this case the variance of the residuals would fall in the frequency range described by $-0.5 < f \leq 0.5$, where f is the frequency in cycles per unit time. Fourier transformation of the sample auto covariance function would provide estimates for the power spectrum.

Confidence limits for the power spectrum at a given level of confidence is determined accordingly to the degrees of freedom required for the chi-squared distribution. It is to be noted that the degrees of freedom are highly dependent upon the shape and bandwidth of the smoothing window used to obtain smooth sample power spectral estimates.

Direct estimation of power spectrum is also possible from the sample values of the series. The following section details associated procedures.

5.5.2.2 Periodogram testing method

From the sample values of the series the power spectrum can directly be estimated. This is known as cumulative periodogram method, where the periodogram $P(f_r)$ of the white noise series $a_t(i)$, is given as:

$$P(f_r) = \frac{2}{N} \left[\left(\sum_{i=1}^N a_t(i) \cos 2\pi f_r i \right)^2 + \left(\sum_{i=1}^N a_t(i) \sin 2\pi f_r i \right)^2 \right]$$

.....5.39

in, which N is the sample size, r is a dummy variable ($r=0,1,2,\dots,k$). Depending on the sample size, k could assume the value of $(N-1)/2$ if N is odd, or $(N-2)/2$ if N is even.

Finally, f_r is defined as

$$f_r = \frac{r}{N}$$

....5.40

The normalised periodogram, according to [10] takes the following form:

$$P_n(f_j) = \frac{\sum_{r=1}^j P(f_r)}{N\sigma_{a_t}^2}$$

....5.41

where $j=0,1,2,\dots,N/2$. Any non-randomness can be detected by the estimate of the variance of the white noise series. This test is also helpful in detecting sinusoidal trend components in the residual series. The white noise series is presented as an approximate straight line from (0.0,0.0) to (0.5,1.0) when plotting the normalized periodogram against frequency, which is called the 'white noise line'.

The confidence limits can be determined by choosing a band $\pm c/\sqrt{k}$ around the 'white noise line' joining (0.0,0.0) to (0.5,1.0). The value of c according to normal tables at 95% level of significance is ± 1.36 .

5.5.2.3 Portmanteau test

In the Portmanteau test, the whiteness of the residual series is tested according to Eq. 5.42,

$$d_1(a_t) = N \sum_{i=1}^m \left[\frac{r_i(a_t)}{r_0(a_t)} \right]^2 \quad \dots 5.42$$

where m is a dummy variable usually equal to one-tenth of the length of the series (N), $r_i(a_t)$ is the autocorrelation of the a series at lag i , $r_0(a_t)$ is the lag zero autocorrelation of the a_t series.

The Portmanteau statistic is calculated as

$$g_t(a_t) = d_1(a_t) - \chi^2_{m,0.05} \quad \dots 5.43$$

where the random variable $d_1(a_t)$ is chi-square distributed with m degrees of freedom. The value of $\chi^2_{m,0.05}$ can be determined from standard chi-square tables.

Chapter VI

RESULTS AND DISCUSSION

6.1 INTRODUCTION

The available riverflow and groundwater level data for the selected watersheds in the Ottawa Region (see page 24) were used in the numerical analysis and subsequent application of model hypothesis. Transforming of the input (riverflow) and output (groundwater level) data was in accordance with the hypothesized linear linkage model. All computations were accomplished by using accessible subroutines from the computational libraries on an IBM 360/65 computer available at the Computing Operations Centre, University of Ottawa, Ottawa, Ontario, Canada.

6.2 MODEL CONSTRUCTION

6.2.1 Stochastic component deduction

The significant periodicities present in a periodic stochastic series can be conveniently detected by performing power spectral analysis. Accordingly, the transformed river-flow and groundwater level data were subjected to spectral analysis using BMD (Biomedical Computer Programs) subroutine BMD02T. Good frequency resolution for smoothed sample spectral estimates were obtained by using lags of 45. The number of lags were determined by $l=N/10$, where N is the number of five days used for the analysis.

The data was standardized (i.e. the transformed series had mean zero and variance unity) before they have been subjected to spectral analysis and power spectral computations. Because of standardization, the power spectral estimates became spectral density estimates. The advantage of standardization is clearly seen since the spectral density estimates provide first hand knowledge of periodicities present in the data. By examining the plots of power spectrum, shown in Figures 6.1, 6.2, 6.3, it became evident that peaks occurred at annual, semiannual, and quarterly cycles (i.e. at 0.0137,

0.0274, and 0.0411 cycles/five-days respectively), however, the annual cycle dominated the whole spectrum in all three cases. It seemed natural then to analyse these periodicities with harmonic analysis of five-daily means and standard deviations for riverflow and groundwater level series.

It was noted in Chapter 5, that when the coefficient of variation is independent of the position of time (in non-log transformed series) then the periodicity is shown in the five-daily means' periodic behaviour. Furthermore, the homogeneity of the variance in the log transformed series is also ensured by logarithmic transformation, if above independence is upheld.

It is necessary therefore to check how constant the coefficient of variation is. A hypothesis of the relation between the coefficient of variation and means of five-daily means is shown below:

$$\sigma_t = c_v \mu_t^h$$

.....6.1

where c_v is the coefficient of variation, σ_t are the standard deviations, μ_t are the means of five-daily hydrologic series, and h is an exponent of the mean. The test consists of checking whether this exponent is significantly different than 1.0. A t-test procedure [62] was used to test the hypothesis of $h=1$.

FIGURE 6.1 POWER SPECTRA OF NON-STATIONARY FIVE-DAILY SERIES
(CASTOR WATERSHED)

FIGURE 6.2 POWER SPECTRA OF NON-STATIONARY FIVE-DAILY SERIES
(CARP 1 WATERSHED)

FIGURE 6.3 POWER SPECTRA OF NON-STATIONARY FIVE-DAILY SERIES
(CARP 2 WATERSHED)

The relationship of Eq. 6.1 in linear form is presented below:

$$\ln \sigma_t = \ln c_v + h \ln \mu_t$$

....6.2

Therefore the test essentially becomes of a check on the significance of linear correlation between the log transformed standard deviations and means. An appropriately documented test [26], which is based on the normal distribution (z-statistic) was used. The test involves a functional estimate of the correlation coefficient r_{12} . Results from above tests are tabulated in Tables 6.1 and 6.2.

In light of these results, the hypothesis that h being insignificantly different than 1.0 was accepted for all watersheds. The hypothesis of linear correlation between the log transformed standard deviations and means is accepted for all watersheds, the coefficients being significant at the 10% level of significance. Therefore, harmonic analysis of the five-daily input-output series would be advantageous in deducing the stationary stochastic component.

Fisher's G-test [62] was used to test the significance of the harmonics. Results of this test revealed in general, that the first three harmonics contain the majority of the explained variance (see Tables 6.3 and 6.4). The corre-

TABLE 6.1 RESULTS OF HYPOTHESIS TESTING - FIVE-DAILY GROUND-WATER LEVELS

Watershed	Coefficient of correlation r_{12}	Statistic "z"	Hypothesis Ho
Castor	0.5835	2.0034	accepted
Carp 1	0.5770	1.9738	accepted
Carp 2	0.5698	1.9416	accepted

Watershed	Value of h	Statistic "t"	Hypothesis Ho
Castor	0.9201	- 0.3527	accepted
Carp 1	1.3031	- 3.1354	accepted
Carp 2	0.9524	- 0.3927	accepted

Ho : (i) linear correlation exist (ii) $h = 1$

H1 : (i) linear correlation does not exist (ii) $h \neq 1$

Hypotheses were tested at the 10% level of significance.
At this level the test statistics are:

$$z_{0.01} = \pm 1.645$$

$$t_{0.01, 10} = \pm 3.169$$

TABLE 6.2 RESULTS OF HYPOTHESIS TESTING - FIVE-DAILY RIVER-FLOW

Watershed	Coefficient of correlation r_{12}	Statistic "z"	Hypothesis H_0
Castor	0.9822	7.0705	accepted
Carp 1	0.9720	6.3819	accepted
Carp 2	0.9720	6.3819	accepted

Watershed	Value of h	Statistic "t"	Hypothesis H_0
Castor	1.1883	2.6212	accepted
Carp 1	0.9721	- 0.3749	accepted
Carp 2	0.9721	- 0.3749	accepted

H_0 : (i) linear correlation exist (ii) $h = 1$

H_1 : (i) linear correlation does not exist (ii) $h \neq 1$

Hypotheses were tested at the 10% level of significance.
At this level the test statistics are:

$$z_{0.01} = \pm 1.645$$

$$t_{0.01,10} = \pm 3.169$$

sponding harmonic coefficients are also shown in Tables 6.5 and 6.6. As can be seen from the results, the harmonic model explains up to 18% of the variation in the five-daily groundwater level and up to 32% of the variation in the riverflow series. This not only shows that the riverflow series composition is built up by a greater percent of stochastic component than the groundwater level series, but that both series are composed of a large percentage of stochastic component not accounted for by the harmonic analysis.

It can also be stated that if a model is needed at this point, then it is possible to develop one to describe the stochastic component of the five-daily groundwater level fluctuation process. That is, if separate models (for deterministic and for stochastic input-output pairs) are required to describe the process under study, which will be determined later on.

TABLE 6.3 PERCENT VARIANCE EXPLAINED BY HARMONICS - FIVE-DAILY GROUNDWATER LEVELS

Watershed	Percent variance explained by						Total
	I	II	III	IV	V	VI	
	harmonics						
Castor	14.44	0.58	2.35	0.01	0.21	0.29	17.89
Fisher's G	yes	no	yes	no	no	no	
Carp 1	7.80	5.82	0.70	0.14	0.06	0.07	14.59
Fisher's G	yes	yes	no	no	no	no	
Carp 2	8.05	4.65	1.20	0.04	0.08	0.12	14.13
Fisher's G	yes	yes	yes	no	no	no	

The statistic used was the ratio of

$$G = \frac{\frac{A^2 + B^2}{\max} \max}{\sum_{j=1}^m \frac{(A_n^2 + B_n^2)}{n}}$$

where A_n and B_n are Fourier coefficients, $A_{\max}$ and $B_{\max}$ are the largest values of the coefficients in Eq. 5.28. If G is greater than the critical value the harmonic associated with the Fourier coefficients is significant, otherwise it is not.

Critical value for Fisher's G test is: 0.17137.

TABLE 6.4 PERCENT VARIANCE EXPLAINED BY HARMONICS - FIVE-DAILY RIVERFLOW SERIES

Watershed	Percent variance explained by						Total
	I	II	III	IV	V	VI	
	harmonics						
Castor	25.77	3.33	2.25	0.09	0.49	0.05	31.98
Fisher's G	yes	yes	yes	no	no	no	
Carp 1	26.62	3.11	1.26	0.07	0.10	0.09	31.25
Fisher's G	yes	yes	yes	no	no	no	
Carp 2	26.62	3.11	1.26	0.07	0.10	0.09	31.25
Fisher's G	yes	yes	yes	no	no	no	

The statistic used was the ratio of

$$G = \frac{A_{\max}^2 + B_{\max}^2}{\sum_{j=1}^m (A_n^2 + B_n^2)}$$

where A_n and B_n are Fourier coefficients, $A_{\max}$ and $B_{\max}$ are the largest values of the coefficients in Eq. 5.28. If G is greater than the critical value the harmonic associated with the Fourier coefficients is significant, otherwise it is not.

Critical value for Fisher's G test is: 0.17137.

TABLE 6.5 VALUES OF HARMONIC COEFFICIENTS - FIVE-DAILY GROUNDWATER LEVEL SERIES

Watershed	Ist harmonic		IInd harmonic		IIIrd harmonic	
	a_1	b_1	a_2	b_2	a_3	b_3
Castor	0.2928	0.2423	0.0681	-0.0343	-0.0929	-0.1220
Carp 1	-0.1266	0.2488	0.2386	-0.0356	-0.0540	-0.0641
Carp 2	-0.1895	0.2111	0.2131	-0.0335	-0.0416	-0.1011

Watershed	IVth harmonic		Vth harmonic		VIth harmonic	
	a_4	b_4	a_5	b_5	a_6	b_6
Castor	-0.0068	-0.0095	0.0450	0.0093	0.0184	-0.0502
Carp 1	-0.0367	-0.0066	0.0084	0.0227	-0.0062	-0.0267
Carp 2	-0.0187	0.0052	0.0204	0.0186	0.0023	-0.0351

The harmonic model used for the periodic series is

$$g_t = a_0 + \sum_{i=1}^6 (a_i \cos \frac{2\pi}{12} t + b_i \sin \frac{2\pi}{12} t)$$

but since the series are standardized (i.e. mean zero, variance unity)

$$a_0 = 0.0000$$

TABLE 6.6 VALUES OF HARMONIC COEFFICIENTS - FIVE-DAILY RIV-
ERFLOW SERIES

Watershed	Ist harmonic		IInd harmonic		IIIRD harmonic	
	a_1	b_1	a_2	b_2	a_3	b_3
Castor	0.2670	-0.4310	-0.1335	0.1242	0.1355	0.0639
Carp 1	0.3804	-0.3486	-0.1566	0.0809	0.0995	0.0524
Carp 2	0.3804	-0.3486	-0.1566	0.0809	0.0995	0.0524

Watershed	IVth harmonic		Vth harmonic		VIth harmonic	
	a_4	b_4	a_5	b_5	a_6	b_6
Castor	-0.0041	-0.0040	-0.0091	-0.0696	-0.0193	0.0091
Carp 1	-0.0174	-0.0211	0.0193	-0.0241	-0.0207	-0.0216
Carp 2	-0.0174	-0.0211	0.0193	-0.0241	-0.0207	-0.0216

The harmonic model used for the periodic series is.

$$g_t = a_0 + \sum_{i=1}^6 \left(a_i \cos \frac{2\pi}{12} t + b_i \sin \frac{2\pi}{12} t \right)$$

but since the series are standardized (i.e. mean zero, variance unity)

$$a_0 = 0.0000$$

6.2.2 Cross spectral and cross correlation analysis

To arrive at smoothed sample auto- and cross spectral estimates, an IMSL (International Mathematical and Statistical Library) subroutine was used, namely subroutine FTFREQ, which yields all the estimates of relevant functions. The number of lags used was determined earlier in 6.2.1, according to $l=N/10$, where N is the number of five-daily periods used for the analysis. Output from the computations includes the sample auto- and cross covariance, and auto- and cross correlation functions, the auto power spectra for both the input and output series, the frequency, cross power and quadrature spectra, phase and gain spectra along with the coherence spectrum. The power spectrum of the output series and coherence spectrum provided a means for the computation of the noise component (i.e. the power spectrum of the residuals). Indication as to the form of the ARIMA process for the noise component can then be extracted from the power spectrum of the residuals.

The nonstationary input (riverflow) and output (groundwater level) series were standardized using Eq. 5.24 before being subjected to cross correlation and cross spectral analyses. These analyses provide insight on the nature of

the delay parameter, and memory content of the system. Modeling requirements can also be determined from the phase and gain functions, such that if these functions are not different from each other for both the nonstationary input and output series significantly, then system identification may be accomplished by a multiple linear regression algorithm. Different models are in order if the difference between them found to be significant: one for the deterministic (periodic) and one for the stochastic input-output pairs.

6.2.2.1 Cross correlogram

Inspection of the cross correlograms reveals that the cross correlation functions between riverflows and groundwater levels for both the stationary and nonstationary series are almost symmetrical and damp out very slowly (see Figures 6.4, 6.5 and 6.6). For both watersheds the cross correlograms have well defined peaks at near zero lag, indicating that the delay parameter is zero. This in turn suggests a concurrent occurrence of the riverflow and groundwater processes.

Another point emerges by noting that the cross correlations are negative, and this suggests that an increase in input produces a decrease in the output. However, the con-

FIGURE 6.4 CROSS-CORRELOGRAM (CASTOR)

FIGURE 6.5 CROSS-CORRELOGRAM (CARP 1)

FIGURE 6.6 CROSS-CORRELOGRAM (CARP 2)

tradiction may be explained by the watershed groundwater storage. Both rivers have high flows at spring when groundwater levels are at their low levels, whereas during the rest of the year the groundwater levels exhibit high values compared to the low flows recorded at corresponding riverflow stations. The concurrent occurrences of low groundwater levels and low riverflows during autumn remain to be explained.

If the memory terms are disregarded in the model, it can be determined that more than 52% variability in the groundwater level fluctuation can be explained by the variability in the riverflow, in terms of the nonstationary series (Table 6.7). However, less than 10% variance in the groundwater level fluctuation is explained by the riverflow in terms of the stationary series. This suggests to utilize nonstationary input-output series in setting up a linear model for the process under study.

In conclusion, the peak cross correlation at the near zero lag suggests close correlation in riverflow and groundwater level for both nonstationary and stationary series; and they are in phase when each series is considered as a whole.

TABLE 6.7 VALUES OF CROSS CORRELATION COEFFICIENTS AT ZERO LAG AND THE VARIANCE EXPLAINED

Watershed	Non-stationary series		Stationary series	
	r_{12}	$r_{12}^2(\%)$	r_{12}	$r_{12}^2(\%)$
Castor	- 0.814	66.26	- 0.312	9.73
Carp 1	- 0.723	52.27	- 0.303	9.18
Carp 2	- 0.724	52.42	- 0.280	7.84

6.2.2.2 Phase spectra

Results from the phase diagram, in general, yield much more accurately the difference in phase between the annual cycles in both series than the cross correlogram, due to the fact that when working in the time domain we consider each process as a whole without any consideration of frequency components. From the phase diagrams in Figures 6.7, 6.8, 6.9, it is observed that oscillation occurs around the value of .475 (on a scale from 0 to 1), or in other words, the riverflow series appears to be, as a whole, 1.14 five days (i.e. 5.7 days) ahead of the groundwater level series. This corresponds to some instability in the linear model for high frequencies, which -as it will be shown- is also the conclusion in the analysis of coherence spectra. On long term basis (i.e. low frequencies - long cyclic periods), however, a linear relationship between riverflow and groundwater levels appears to be a fair approximation.

This is supported by the fact that no appreciable lag or lead (+1.14 five-days) is present for both stationary and nonstationary input-output series; therefore the delay parameter can be accepted to be equal to zero. Furthermore, if the presence of trend is shown by the phase function with delay parameter being zero, then the system contains memory.

FIGURE 6.7 PHASE SPECTRA (CASTOR)

— STATIONARY
- - NON-STATIONARY

FIGURE 6.8 PHASE SPECTRA (CARP 1)

σ^2
 — STATIONARY
 --- NON-STATIONARY

FIGURE 6.9 PHASE SPECTRA (CARP 2)

STATIONARY
NON-STATIONARY

TABLE 6.8 RESULTS OF TEST OF RANDOMNESS FOR THE PHASE FUNCTION

Watershed	Non-stationary series		Stationary series	
	Run statistic	Hypothesis	Run statistic	Hypothesis
Castor	- 1.676	rejected	- 2.320	rejected
Carp 1	- 1.742	rejected	- 1.765	rejected
Carp 2	- 1.784	rejected	- 0.850	accepted

The hypothesis of randomness was tested at 10% level of significance. The critical value of test statistic at this level is ± 1.65 .

TABLE 6.9 RESULTS OF MANN-WHITNEY U TEST OF IDENTICAL POPULATION FOR PHASE FUNCTION

Watershed	Phase function	
	Statistic "U"	Hypothesis
Castor	0.000	accepted
Carp 1	- 0.340	accepted
Carp 2	- 0.718	accepted

The hypothesis of identical population was tested at 5% level of significance. The critical value of test statistic at this level is $= \pm 1.96$.

If this memory is very small, trends can be hardly detected in the estimated phase function because of the inconsistency of the spectral estimates.

Thus, non-parametric tests were employed to investigate whether the phase functions for both nonstationary and stationary input-output series are identical or not. The test used was the Mann-Whitney U-test of identical population [11]. Also a run test of randomness [43] about zero radian was conducted to see whether or not the phase function is trend free. Results from above tests are given in Tables 6.8 and 6.9.

The nonparametric tests reveal that the phase functions for nonstationary and stationary input-output series are insignificantly different, with the exception of Carp 2 watershed, thus suggesting that $\phi(f) = \phi_3(f)$ in Eq.'s 5.13 through 5.15. Also the run test of randomness indicates the presence of some memory associated with the nonrandom fluctuations in the phase function.

In conclusion, the phase spectra indicate that the delay parameter is zero, and no appreciable memory is present in the system. Therefore the proposed linear systems model can be accepted. Moreover, the frequency response functions $H(f)$ and $H_3(f)$ can be considered insignificantly different, pending results from the nonparametric tests of gain function.

6.2.2.3 Gain spectra

The ratio between the amplitudes of the frequency components of input-output series at each frequency is called the gain function, and the particular gain functions, for each watershed, are shown in Figures 6.10, 6.11, and 6.12. The gain function should fluctuate about a constant value throughout the frequency range if the system is memoryless. In case of significant memory, the gain function will show some trend of variation with respect to frequency.

Keeping the above in mind, tests were conducted to see whether the gain functions are trend free and whether they are random about their median or not, by using run test of randomness [43]. Also Mann-Whitney U-test [11] of identical population was employed to determine whether or not the gain functions for both stationary and nonstationary input-output series are identical.

The nonparametric tests reveal that the gain functions for nonstationary and stationary input-output series are insignificantly different, thus suggesting that $G(f)=G_3(f)$ in Eqs. 5.13 through 5.15. Some trend can be detected in the fluctuations of the gain function with respect to frequencies. However, Carp 2 watershed defies both hypothesis at 10% level of significance.

FIGURE 6.10 GAIN SPECTRUM (CASTOR)

— NON-STATIONARY
- - - STATIONARY

FIGURE 6.11 GAIN SPECTRUM (CARP 1)

— NON-STATIONARY
--- STATIONARY

FIGURE 6.12 GAIN SPECTRUM (CARP 2)

TABLE 6.10 RESULTS OF TEST OF RANDOMNESS FOR THE GAIN FUNCTION

Watershed	Non-stationary series		Stationary series	
	Run statistic	Hypothesis	Run statistic	Hypothesis
Castor	- 2.926	rejected	- 2.528	rejected
Carp 1	- 2.130	rejected	- 2.149	rejected
Carp 2	- 2.187	rejected	- 2.339	accepted

The hypothesis of randomness was tested at 10% level of significance. The critical value of test statistic at this level is $= \pm 1.65$.

TABLE 6.11 RESULTS OF MANN-WHITNEY U TEST OF IDENTICAL POPU-
LATION FOR GAIN FUNCTION

Watershed	Gain function	
	Statistic "U"	Hypothesis
Castor	- 1.788	accepted
Carp 1	- 0.836	accepted
Carp 2	- 3.178	rejected

The hypothesis of identical population was tested at 5% level of significance. The critical value of test statistic at this level is ± 1.96 .

Cursory examination of gain spectra reveals that some decrease in gain can be observed at high frequencies relative to that of at low frequencies. Results of the tests are shown in Tables 6.10 and 6.11.

In conclusion it can be definitely stated that the frequency response functions $H(f)$ and $H_3(f)$ are satisfying the criterion of equality in light of the nonparametric tests conducted on the phase and gain functions. Consequently, in Eq.'s 5.13 through 5.15 $H(f)=H_2(f)=H_3(f)$, therefore a model in terms of nonstationary input-output series would be more realistic than separate models for the periodic and stochastic input-output series.

6.2.2.4 Coherency spectra

The coherence is insignificantly different from zero at frequencies higher than 0.11 cycles/five-days -based on a chi-square test- for all watersheds and for both stationary and nonstationary input-output series (see Figures 6.13, 6.14, and 6.15). The main features of the coherency spectra are the presence of a large peak around 0.01 cycles/five-days and a flatter region with an average value of approximately $\kappa^2(f)=0.13$, stretching over the entire frequency range. Since the coherence estimates are obtained by using

FIGURE 6.13 COHERENCE SPECTRUM (CASTOR)

— STATIONARY
 - - - NON-STATIONARY

FIGURE 6.14 COHERENCE SPECTRUM (CARP 1)

— STATIONARY
 --- NON-STATIONARY

FIGURE 6.15 COHERENCE SPECTRUM (CARP 2)

— STATIONARY
 --- NON-STATIONARY

TABLE 6.12 COHERENCE VALUES AT ANNUAL CYCLE FOR
NON-STATIONARY AND STATIONARY SERIES

Watershed	Non-stationary series	Stationary series	Test statistic at 5% level of sign.
Castor	0.889	0.888	0.451
Carp 1	0.738	0.831	0.451
Carp 2	0.822	0.745	0.451

IMSL subroutine FTFREQ, they were based on the smoothed sample auto- and cross spectral estimates of input-output series.

The shape of the coherence spectra indicate that there is very little coherence beyond 0.11 cycles/five-days. It is noted that approaching very low frequencies the coherence is decreasing. This would indicate that the groundwater storage is more sensitive to long term changes than the riverflow. Therefore on the long term basis a linear relationship appears to be a good approximation. The validity of a linear regression model at annual cycle is evident from the high values of coherence for nonstationary series except for Carp 1 watershed. On the average, however, the coherence is lower at low frequencies for stationary series than for the corresponding nonstationary series.

The values of coherence at annual cycle for both nonstationary and stationary series are shown in Table 6.12. As it has been shown, the larger part of variance for the groundwater and riverflow processes is contained in the lower frequencies and as it can be seen the coherence is higher at the same frequencies for the nonstationary series than for the stationary ones. It follows then, that a successful attempt can be made to explore a linear model in terms of the nonstationary input-output series.

6.2.2.5 Power spectra of the noise component

It was stated in Section 6.2.2. that the coherence and power spectra of the output series provide an excellent means to obtain the power spectrum of the noise component. As can be seen from Figures 6.16, 6.17, 6.18 nearly equal power is exhibited at all frequencies in the spectrum of the noise component. Therefore, the noise component can be regarded as a white noise series by virtue of the definition, which states that a process is termed white noise if it has approximately the same intensity at all frequencies [34].

The hypothesis of equality of the impulse response functions $V_2(B)$, $V_3(B)$, $V(B)$ -Eqs. 5.3 through 5.6- is supported by the results of the Mann-Whitney U-test (Table 6.13) whereby it was shown that the power spectrum of the noise component for the nonstationary (q_t, g_t) and stationary (a_t, b_t) input-output series are identical. It is therefore possible to develop a linear model using either the nonstationary or stationary input-output series, since the noise component is identical for both cases and most importantly they are white noise. It is generally accepted that if the noise component found to be white noise series, then the model fitting process is simplified.

FIGURE 6.16 POWER SPECTRUM OF THE NOISE COMPONENT (CASTOR)

NON-STATIONARY
STATIONARY

FIGURE 6.17 POWER SPECTRUM OF THE NOISE COMPONENT (CARP 1)

NON-STATIONARY
STATIONARY

FIGURE 6.18 POWER SPECTRUM OF THE NOISE COMPONENT (CARP 2)

— NON-STATIONARY
 --- STATIONARY

TABLE 6.13 MANN-WHITNEY U-TEST OF IDENTICAL POPULATION FOR NOISE COMPONENT

Watershed	Noise component	
	U statistic	Hypothesis
Castor	- 1.5379	accepted
Carp 1	- 1.8737	accepted
Carp 2	- 1.9127	accepted

Hypothesis of identical population was tested at 95% level of significance. The critical value associated with this level is ± 1.96 .

In light of the results obtained by examining the gain, phase and power spectra (of the noise term), it was decided that for model fitting procedures the nonstationary input-output series will be utilized.

6.3 MODEL FITTING FOR FIVE-DAILY GROUNDWATER AND RIVERFLOW PROCESSES

6.3.1 Frequency versus time domain analysis

The preceeding analyses of the five-daily groundwater level and riverflow processes led to the acceptance of using nonstationary input-output series in building a parsimonious model. However, both time domain (a multiple linear regression) and frequency domain (phase and gain spectra) techniques could be used for identifying the groundwater level and riverflow processes as a linear system.

In order to employ frequency domain techniques for a linear system identification it is noted that the low frequency range is dominant (see Figures 6.1 through 6.3), therefore to achieve proper frequency resolution (for five-daily series) one is required to work with a large number of lags, which in turn, requires considerable computing time for Fourier transformation of the auto- and cross-covariance

functions in estimating the auto- and cross-spectra. Suppose, if one is determined to employ frequency domain procedures there will still be a need for time domain analysis because of parameter estimation for the governing equation of the system. Besides, model diagnostic checks may indicate that further improvements are due in the model structure, therefore an alternative model may be considered and fitting procedures are to be executed again.

Time domain analysis do provide advantages not only in computing time, but in resulting satisfactory systems equation through a more simple multiple regression procedure. Accordingly, the model hypothesis was implemented by using time domain techniques in fitting a satisfactory systems model. Since the noise component was found to be white noise and the delay parameter has been equal to zero, it was decided that for model building purposes the procedure described in Section 5.4.1 should be utilized.

6.3.2 Model fitting in terms of time domain techniques

The model sought was in the form of

$$\begin{aligned}
 g_t - \delta_1 g_{t-1} - \delta_2 g_{t-2} - \dots - \delta_p g_{t-p} &= \\
 &= \omega_0 q_t - \omega_1 q_{t-1} - \omega_2 q_{t-2} - \dots - \omega_n q_{t-n} + a_t
 \end{aligned}
 \quad \dots 6.3$$

It is different from the model given by Eq. 5.27, since the delay parameter was found to be zero and the noise term was shown being a white noise series, warranted above changes.

To identify the orders, p and n , associated with the model described in Eq. 6.2, the IMSL subroutine RLSTP was utilized. This subroutine not only provides estimators for model orders, but the corresponding parameters are also estimated. RLSTP uses a forward stepwise algorithm in the regression model selection and forces the chosen independent variables into the model [18].

A zero order model ($p=0, n=0$) was tried first, residuals obtained and diagnostic checks performed. The autocorrelations were found to be significantly different than zero at past lags. The result was the same for the cross-correlations at present and past lags.

A model with $p=1, n=0$ was attempted next, but the residuals still exhibited significantly different than zero auto- and cross-correlations. First order model ($p=1, n=1$) model was then tested and again the auto- and cross-correlation checks gave doubtful results.

Finally, a model with $p=2$ and $n=0$ yielded residuals for which the auto- and cross-correlation checks revealed that insignificant differences from zero exist at present and past lags (Tables 6.14 and 6.15). In general, model orders

TABLE 6.14 CROSS-CORRELATION CHECK FOR RESIDUALS

Watershed	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Test statistic
Castor	-0.07	-0.07	-0.05	-0.03	0.00	0.02	0.05	0.06	0.10	0.12	0.09	0.08	0.07	0.05	0.08	± 0.15
Carp 1	0.01	0.03	0.04	0.07	0.07	0.05	0.07	0.10	0.08	0.06	0.06	0.10	0.06	0.04	0.03	± 0.15
Carp 2	0.04	0.06	0.07	0.10	0.10	0.09	0.10	0.12	0.11	0.09	0.08	0.10	0.06	0.05	0.04	± 0.15

Watershed	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Test statistic
Castor	0.08	0.09	0.14	0.11	0.08	0.05	0.05	0.05	0.01	0.03	0.01	0.01	0.06	0.07	0.04	± 0.15
Carp 1	0.03	0.02	0.01	-0.01	-0.03	-0.04	-0.04	-0.05	-0.07	-0.07	-0.08	-0.10	-0.10	-0.09	-0.10	± 0.15
Carp 2	0.02	0.02	0.01	-0.01	-0.03	-0.06	-0.07	-0.06	-0.08	-0.09	-0.10	-0.11	-0.11	-0.10	-0.12	± 0.15

Watershed	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	Test statistic
Castor	0.06	0.05	0.05	0.07	0.05	0.03	-0.00	0.02	0.01	0.01	0.02	0.04	0.03	-0.00	0.01	± 0.15
Carp 1	-0.09	-0.07	-0.03	-0.03	-0.03	-0.01	-0.01	0.01	0.02	0.03	0.05	0.12	0.12	0.08	0.06	± 0.15
Carp 2	-0.11	-0.10	-0.07	-0.08	-0.07	-0.06	-0.06	-0.04	-0.01	0.00	0.03	0.10	0.09	0.05	0.03	± 0.15

TABLE 6.15 AUTO-CORRELATION CHECK FOR RESIDUALS

Watershed	1	2	3	4	5	6	Lag			11	12	13	14	15	Test statistic
							7	8	9						
Castor	0.02	0.03	-0.03	0.01	0.01	0.05	0.06	-0.02	0.01	-0.02	0.01	0.06	0.09	0.03	-0.05
Carp 1	0.04	0.03	-0.05	0.17	0.05	0.05	0.00	-0.03	-0.01	-0.01	0.04	-0.01	0.05	-0.03	-0.01
Carp 2	0.04	-0.01	-0.02	0.13	0.04	0.05	-0.01	-0.04	-0.05	-0.02	0.01	0.07	-0.06	0.00	0.01
															± 0.15
															± 0.15
															± 0.15
Watershed	16	17	18	19	20	21	Lag			26	27	28	29	30	Test statistic
							22	23	24						
Castor	0.01	0.01	0.00	0.00	-0.07	0.01	-0.02	0.07	0.01	-0.01	-0.02	0.03	-0.03	0.03	0.02
Carp 1	0.00	-0.01	-0.02	-0.02	0.01	-0.03	-0.01	-0.05	0.00	0.04	0.02	0.04	0.00	0.03	0.03
Carp 2	0.02	-0.03	-0.02	-0.02	-0.00	-0.04	-0.02	-0.02	0.02	0.02	0.03	0.04	0.03	0.03	0.02
															± 0.15
															± 0.15
															± 0.15
Watershed	31	32	33	34	35	36	Lag			41	42	43	44	45	Test statistic
							37	38	39						
Castor	-0.06	-0.06	-0.06	-0.02	0.01	0.02	-0.01	-0.08	-0.06	-0.06	-0.02	0.03	-0.05	-0.06	-0.06
Carp 1	0.03	-0.00	0.01	0.06	-0.00	0.02	-0.06	0.06	-0.01	0.00	-0.03	0.01	0.00	0.01	-0.04
Carp 2	0.02	0.01	0.03	0.02	-0.01	0.01	-0.01	0.01	0.00	0.01	-0.01	0.00	0.00	0.01	-0.04
															± 0.15
															± 0.15
															± 0.15

TABLE 6.16 TESTING OF RESIDUALS FOR WHITENESS (PORTMANTEAU TEST)

Watershed	Portmanteau statistic	Chi-square statistic $\chi^2_{43,0.05}$	Hypothesis Ho
Castor	32.33	60.5	accepted
Carp 1	30.14	60.5	accepted
Carp 2	21.54	60.5	accepted

Ho : The residuals are independent and trend-free, with zero mean and variance σ_a^2 .

Ha : The residuals obey an AR (or ARIMA) process.

FIGURE 6.19 CUMMULATIVE PERIODOGRAM (CASTOR WATERSHED)

FIGURE 6.20 CUMMULATIVE PERIODOGRAM (CARP 1 WATERSHED)

FIGURE 6.21 CUMMULATIVE PERIODOGRAM (CARP 2 WATERSHED)

$p=2$, $n=0$ were accepted for all watersheds in light of the results of diagnostic checks. Portmanteau statistic (Table 6.16) and cumulative periodogram (Figures 6.19 through 6.21) demonstrated that the residuals, from models fitted, can be considered white and trend-free.

It was mentioned before; that the multiple regression analysis uses a linear least squares algorithm for parameter estimation. These estimates can be considered acceptable since the residuals (noise component) were found to be un-autocorrelated, therefore possess the characteristics of a white noise series.

The final model of the five-daily groundwater and riverflow process therefore takes the following form:

$$g_t - \delta_1 g_{t-1} - \delta_2 g_{t-2} = \theta_0 + \omega_0 q_t + a_t$$

....6.4

where a_t is a white noise series with mean zero, and variance $\sigma_{a_t}^2$. It can be seen, that some memory is present in the system, since the condition $\delta_1=0$ and $\omega_1=0$ is not completely fulfilled. The parameter estimates associated with the accepted model orders, along with the residuals' variance, are shown in Table 6.17.

Stability conditions require that, for a second order model,

$$\delta_2 \pm \delta_1 < 1$$

TABLE 6.17 PARAMETER ESTIMATES FOR ACCEPTED MODEL ORDERS

Watershed	$\hat{\Lambda}_{\theta_0}$	$\hat{\Lambda}_{\delta_1}$	$\hat{\Lambda}_{\delta_2}$	$\hat{\Lambda}_{\omega_0}$	Model orders p n	Variance of residuals
Castor	0.002	1.103	-0.249	-0.141	2 0	0.047
Carp 1	0.020	1.139	-0.271	-0.121	2 0	0.060
Carp 2	0.015	1.081	-0.216	-0.128	2 0	0.055

Residuals are white noise. Accordingly obey ARIMA (0,0,0) model where $p=0, d=0, q=0$.

For stability:

Watershed	$\delta_2 \pm \delta_1 < 1$	$-1 < \delta_2 < 1$
Castor	0.854	- 0.249
Carp 1	0.868	- 0.271
Carp 2	0.865	- 0.216

TABLE 6.18 ANALYSIS OF VARIANCE

Watershed	Variance of groundwater levels	Variance of residuals	Variance explained (percent)	Nature of residuals
Castor	2.366	0.047	95.37	white noise
Carp 1	0.165	0.060	92.22	white noise
Carp 2	0.276	0.055	94.03	white noise

and

$$-1 < \delta_2 < 1$$

be true. These requirements for model stability are met (see Table 6.17), therefore the dynamic groundwater level-river-flow model shown by Eq. 6.4 is stable for all watersheds considered.

6.3.3 Remarks on the performance of model hypothesis

A thorough examination of the variance provides an acceptable way in evaluating the model hypothesis for the groundwater level fluctuation process. The model arrived at was of a second order ($p=2, n=0$), and the results of diagnostic checks suggested that the residuals follow an ARIMA model, namely ARIMA (0,0,0). As can be seen from Table 6.18, the model with five parameters explained more than 92% variance in all watersheds for five-daily groundwater level series. The remaining less than 8% variance is therefore accounted for by the residual white noise component.

Thus, the model hypothesis is accepted in view of the high values of the explained variance. Furthermore, the application of the model hypothesis provides a quick and sensible way to simulate groundwater level - riverflow system,

since a simple multiple linear regression algorithm is used to identify relevant parameters in fitting the model.

6.4 DISCUSSION ON GROUNDWATER LEVEL FLUCTUATION PROCESS

6.4.1 Introduction

The unit step and frequency response functions of the groundwater level - riverflow system allows meaningful interpretations on the behaviour of the watershed groundwater level fluctuation process. Therefore, the dynamic models secured on a five-daily basis for all watersheds were used to calculate the unit step and frequency response functions. To be able to compare the systems' response functions, the input (riverflow) and output (groundwater level) series were standardized. The standardization results in a series which will have zero mean and unit variance.

6.4.2 Unit step response function

The unit step response function is defined as the variation of the output (groundwater levels) with respect to time ($t=0$ to $t<\infty$), when the input (riverflow) to the system is made unity throughout the timespan $t=0$ to $t<\infty$.

As it has been determined earlier, the second order model proposed explains more than 92% variance in the groundwater series while the noise component contributes only less than 8% variance towards the total. Since the noise term is small, one could neglect its contribution to attain the unit step response function of a given system. Therefore the model for five-daily groundwater level fluctuation process can be reduced to

$$g_t = \delta_1 g_{t-1} - \delta_2 g_{t-2} + \omega_0 q_t \quad \text{....6.5}$$

or by employing the backshift operator, Eq. 6.5 can be written as

$$(1 - \delta_1 B - \delta_2 B^2) g_t = \omega_0 q_t \quad \text{....6.6}$$

Present and past values of the input can be represented in the frame of Eq. 6.6 as

$$g_t = (v_0 + v_1 B + v_2 B^2 + \dots + v_k B^k) q_t \quad \text{....6.7}$$

where the impulse response weights are denoted by v_i ($i=0,1,\dots,k$). The two equations then can be equated to yield:

$$(1 - \delta_1 B - \delta_2 B^2)(v_0 + v_1 B + v_2 B^2 + \dots + v_k B^k) = \omega_0 \quad \dots 6.8$$

By equating the coefficients in Eq. 6.8, the impulse response weights can easily be obtained. Therefore, the equations below represent the computational version of the impulse response weights

$$\begin{aligned} v_0 &= \omega_0 \\ v_1 &= v_0 \delta_1 \\ v_2 &= \delta_1 v_1 + \delta_2 v_0 \\ &\vdots \\ v_k &= \delta_1 v_{k-1} + \delta_2 v_{k-2} \end{aligned}$$

Since the step input is $q_t=0$ for $t<0$ and $q_t=1$ for $t\geq 0$, the impulse response weights can be utilized to obtain the system's step response to unit step input in the following manner:

$$\begin{aligned} g_t &= v_0 q_t = v_0 && \text{at } t=0 \\ g_t &= v_0 q_t + v_1 q_{t-1} = v_0 + v_1 && \text{at } t=1 \\ g_t &= v_0 q_t + v_1 q_{t-1} + v_2 q_{t-2} = v_0 + v_1 + v_2 && \text{at } t=2 \\ &\vdots \end{aligned}$$

$$\begin{aligned}
 g_t &= v_0 q_t + v_1 q_{t-1} + \dots + v_k q_{t-k} = \\
 &= v_0 + v_1 + \dots + v_k \quad \text{at } t=k
 \end{aligned}$$

Using above expressions, the step response weights were calculated and plotted for all watersheds under consideration (see Figure 6.22).

It is worthwhile to mention that because of the standardization (i.e. to be able to compare among the systems' response functions) the parameter(s) associated with the riverflow series has to be adjusted as well, however values corresponding to the groundwater level series would remain unchanged. The values to be used then can be calculated by multiplying the parameter, w_0 , in Table 6.17 by σ_q/σ_g , which are estimates of the standard deviations of series q_t and g_t respectively.

Examination of the plot of unit step functions indicates that a sudden rise in the riverflow to unity is followed by a gradual increase in the groundwater level. In general the groundwater level then rises rather rapidly although the input rate remained constant. However, after a lag of 18 days the output reaches a constant plateau, which suggest that the regime conditions are established by then.

It is well known, that any natural disruption causes changes in the level of the groundwater table, but man made

FIGURE 6.22 STEP RESPONSE FUNCTION FOR FIVE-DAILY GROUNDWATER LEVEL FLUCTUATION SYSTEM

• CUSTOM MEASURED
• CIP 1 MEASURED
• CIP 2 MEASURED

influences are particularly pronounced [6]. For example cultivation affects infiltration, evapotranspiration, and overland flow; pumping or artificially recharging groundwater aquifers, irrigation further accentuates the fluctuation in the groundwater table. In severe cases the effect of riverflow may diminish altogether. It is expected that above changes may amplify the fluctuation in groundwater levels before finally reaching regime conditions.

The general shape of the unit step response function of the watershed groundwater level - riverflow system also reveals that no particular changes expected during the rising phase of flood. A stream is reacting steadily to high seasonal differences such as spring thaw and consequent runoff, the groundwater in an aquifer connected to this stream probably exhibit rise also, but not as pronounced. On the later stages of flood when it maintains near constant volume, one could expect high values in groundwater levels. While the flood volume diminishes sharply in the receding phase, the groundwater level may be still rising before finally leveling off to a constant value.

It is also evident, that if a unit step input is introduced to the watershed fluvial system, the unit step function reaches an equilibrium value in approximately 20 days for studied watersheds.

6.4.3 Frequency response function

The calculation of the frequency response function is simple by making use of Fourier analysis, which transforms the system's impulse response into frequency domain. The frequency response function contains the system's response to a signal at certain frequency, but scaled in amplitude by the gain function and shifted in phase by a phase angle. The frequency response function, as is well known, is a complex valued function and is shown below in mathematical terms

$$H(f) = G(f) e^{j\phi(f)}$$

....6.9

The phase function, denoted by $\phi(f)$, measures the phase shift at a given frequency, while the gain function ($G(f)$) gives the amplitude magnification at the same frequency. Therefore, the equation governing the system yields the frequency response function upon Fourier transformation. It should also be noted that because of the noise component's negligible contribution to the variance explained (less than 8%), it was disregarded throughout the following calculations. The model for five-daily groundwater level fluctuation process therefore is presented below in the absence of the noise term as:

$$s_t = \delta_1 s_{t-1} + \delta_2 s_{t-2} + \omega_0 q_t \quad \dots 6.10$$

The difference equation shown in Eq. 6.10 can be reproduced by employing groundwater level and riverflow values at chosen discrete time points. The amounts g_t , g_{t-1} , g_{t-2} , q_t can be regarded as continuous g_t and q_t series at times $t=t\tau$, $(t-1)\tau$, $(t-2)\tau$, where τ is equal to one five-day interval in our case. Accordingly Eq. 6.10 becomes:

$$g(t) = \delta_1 g(t-1) + \delta_2 g(t-2) + \omega_0 q(t) \quad \dots 6.11$$

which after transformation yields

$$g(f) = \frac{\omega_0}{1 - \delta_1 e^{-j2\pi f} - \delta_2 e^{-j4\pi f}} q(f)$$

or

$$g(f) = H(f)q(f) \quad \dots 6.12$$

The complex exponentials can be expressed in terms of sine and cosine functions, therefore

$$H(f) = \frac{\omega_0}{1 - \delta_1 (\cos 2\pi f - j \sin 2\pi f) - \delta_2 (\cos 4\pi f - j \sin 4\pi f)} \quad \dots 6.13$$

Defining the real and imaginary components as $R(f)$ and $I(f)$ in Eq. 6.9, the following can be obtained:

$$H(f) = G(f) e^{j\phi(f)} = R(f) + j I(f) \quad \text{....6.14}$$

After equating the right hand side of Eqs. 6.13 and 6.14 and performing algebraic simplifications, the real and imaginary components can be written as

$$R(f) = \frac{\omega_0 - \delta_1 \omega_0 \cos 2\pi f}{1 - \delta_1^2 + \delta_2^2 - 2(\delta_1 - \delta_1 \delta_2) \cos 2\pi f - 2\delta_2 \cos 4\pi f} \quad \text{....6.15}$$

and

$$I(f) = \frac{\delta_2 \omega_0 \sin 4\pi f - \delta_1 \omega_0 \sin 2\pi f}{1 - \delta_1^2 + \delta_2^2 - 2(\delta_1 - \delta_1 \delta_2) \cos 2\pi f - 2\delta_2 \cos 4\pi f} \quad \text{....6.16}$$

Accordingly from Eq. 6.14

$$G(f) = [R^2(f) + I^2(f)]^{\frac{1}{2}} \quad \text{....6.17}$$

and

$$\phi(f) = \arctan [I(f)/R(f)] \quad \text{....6.18}$$

provide a means to compute the gain and phase functions. Graphical representation of these functions are shown in Figures 6.23 and 6.24.

Generally speaking, the gain ranges between 1.3 and 0.5 at all frequencies (i.e. 0.0 - 0.5 cycles/five-days). After an initial rise, the gain function (Figure 6.23) damps out quickly indicating that the low frequency components of the groundwater level fluctuation process have greater amplitudes than of the high frequency components. The variability in the groundwater levels for rare floods thus is higher than for the more frequently occurring riverflow events as it is suggested by the high values of gain at low frequencies (which correspond to monthly, yearly, and higher time intervals). In other words the groundwater level fluctuation process seems more probabilistic in the low flow range than in the range of high flows.

Examination of the phase diagram (Figure 6.24) reveals that in general the groundwater level process slightly (on the average by 0.2 days for all watershed) lags behind the riverflow process at all frequencies and that the phase shift ranges between 0.00-1.60 days. This observation suggests a tendency for peak riverflows to precede the maximum groundwater levels.

FIGURE 6.23 . GAIN FUNCTION FOR FIVE-DAILY GROUNDWATER LEVEL

FLUCTUATION AND RIVERFLOW SYSTEM

- CISTON WATERSHED
- CAPP 1 WATERSHED
- △ CAPP 2 WATERSHED

FIGURE 6.24 PHASE FUNCTION FOR FIVE-DAILY GROUNDWATER LEVEL
FLUCTUATION AND RIVERFLOW SYSTEM

TABLE 6.19 VARIANCES EXPLAINED BY MODEL ORDERS

Watershed	Zero order			First order			Second order		
	r	s	$\sigma^2(\%)$	r	s	$\sigma^2(\%)$	r	s	$\sigma^2(\%)$
Castor	0	0	79.65	1	0	95.18	2	0	95.37
Carp 1	0	0	76.62	1	0	92.10	2	0	92.22
Carp 2	0	0	78.69	1	0	93.94	2	0	94.03

Note: The results of model diagnostic checks indicate, that the second order model adequately describe the groundwater level fluctuation riverflow process.

The gradual change in gain and phase in the frequency range of 0.0 to 0.5 cycles/five-days implies the apparent presence of memory terms. The extent of the contribution of memory components towards explaining the variance in the groundwater process is an indication of whether this memory is weak or strong. In case of weak memory or memoryless model in terms of groundwater level and riverflow would not explain significant amount of variance when comparing zero, first, and second order models. In fact the contribution is small as can be seen from Table 6.19.

Chapter VII

CONCLUSION

The single input, single output model developed in this thesis can be used to describe the effects of riverflow on groundwater. These models were developed for two eastern Ontario watersheds.

The analyses performed would imply that the groundwater level fluctuation process can at any point in time and/or frequency be adequately described by proposed models. Groundwater hydrology would certainly be benefited by the use of input-output techniques for simulation and/or forecasting purposes. The conclusions therefore can be collected in the following points:

1. Logarithmic transformed riverflow and groundwater level series on a five-daily basis can be regarded as input and output for the watershed fluvial system and a discrete linear dynamic model for groundwater level fluctuation process can thus be developed;
2. Although these series are nonstationary in nature, frequency and time domain analyses have proved that a parsimonious systems model can be built directly in terms of these series without stationarizing;

3. Both the groundwater level and riverflow series have shown periodicities in the mean and standard deviations, which were promptly described by Fourier series with its principal and subharmonics, and accordingly removed as deterministic components;
4. A second order model was found to be yielding satisfactory description of the watershed's fluvial system with riverflow and groundwater level as input and output respectively;
5. The noise component present in the watershed fluvial system on a five-daily basis was found to possess the characteristics of a white noise series, while containing less than 8% variance in the five-daily groundwater level series;
6. The variance of the five-daily groundwater level series was explained by the linear dynamic input-output model in the excess of 92%;
7. The parsimonious structure of the linear dynamic models enables synthesizing groundwater level data by considering riverflow series only;
8. The unit step and frequency response functions have shown to be useful in providing a better understanding and insight to the groundwater level fluctuation and riverflow processes, while the information con-

tent regarding the model was also improved by these functions. The frequency response function also indicated the presence of weak memory components in the groundwater level fluctuation - riverflow process, which proved to be insignificant.

In conclusion it is worthwhile to mention, that the model validation procedure, namely testing the characteristics of residuals is not a common occurrence in the hydrologic literature. It has to be stressed here, that unless the residuals are found to be uncorrelated the models have to be rejected. The model proposed in this study has passed these tests and found to satisfactorily describe the groundwater level fluctuation - riverflow process.

Chapter VIII

SUGGESTIONS FOR FUTURE STUDIES

Groundwater levels in a watershed environment and processes related to them can be modelled by stochastic techniques. The stochastic models obtained can be used in turn for simulation and/or forecasting. In the present study only a single input and single output system was identified. It is possible, however, to extend this work to multivariate processes of groundwater level, riverflow, precipitation and other related processes, since the more variables introduced, the more accurate the results are likely to be.

Further investigation should be directed towards applying other than logarithmic transformation methods such as square, cube, squareroot, etc. [10], in the hope of improving the structure of the residuals. Exploration of other methods in supplement to correlation analysis of residuals would also be of great use while working with hydrologic series.

Also, the present model disregards the problem posed by intermittent streams. Future studies could address this phenomenon and a satisfactory solution will be found, thereby allowing analyses to be made with increased accuracy.

Across Canada there are numerous watersheds gauged for groundwater level observation, a legacy of the International Hydrologic Decade (1965-1974). The data provided by these experimental stations could be a basis for fitting improved dynamic input-output models. The parameters obtained may be correlated with other watersheds similar characteristics, thus forming a nucleus for an information base of fluvial characteristics in several regions in Canada.

BIBLIOGRAPHY

- 1 Adamowski, K., "Spectral density of a river flow time series", J. of Hydrology, Vol. 14, 1971.
- 2 Adamowski, K., "Lecture notes on stochastic hydrology", (unpublished), University of Ottawa, Ottawa, On., 1979.
- 3 Adamowski, K., and Oosterveld, M., "Time series analysis of a watershed response variable", Water Resources Research, Vol. 11, No. 5, October 1975.
- 4 Adamowski, K., and Smith, A.F., "Stochastic generation of rainfall", J. of the Hydraulic Division, ASCE, Vol. 98, No. HY11, Proc. Paper 9353, November 1972.
- 5 Amorocho, J., and Brandstetter, A., "Determination of non-linear functional response functions in rainfall runoff processes", Water Resources Research, Vol. 7, No. 5, 1971.
- 6 Balogh, J., and Matrai, I., "Man's influence on hydrological phenomena", 2nd International Postgraduate Course on Hydrological Methods for Development of Water Resources Management, Budapest, Hungary, 1968. (in Hungarian)
- 7 Bear, J., DYNAMICS OF FLOW IN POROUS MEDIA, American Elsevier, New York, N.Y., 1972.
- 8 Bendat, J.C., and Piersol, A.G., RANDOM DATA ANALYSIS AND MEASUREMENT PROCEDURES, 1st ed., Wiley & Sons, Inc., New York, N.Y., 1971.
- 9 Bourodimos, E.L., and Oguntase, A.M., "Cross-spectral analysis of rainfall and runoff for Raritan and Mullica river basins in New Jersey", J. of Hydrology, Vol. 21, 1974.
- 10 Box, G.E.P., and Jenkins, G.M., TIME SERIES ANALYSIS, FORECASTING AND CONTROL, 2nd ed., Holden-Day, San Francisco, Calif., 1976.

- 11 Breiman, L., STATISTICS: WITH A VIEW TOWARD APPLICATIONS, 1st ed., Houghton Mifflin Co., Boston, Mass., 1973.
- 12 Charron, J.E., "A study of groundwater flow in Russell County, Ontario", Sci. Ser. No. 40, Water Resources Branch, Environment Canada, 1974.
- 13 Chin, V.I., Wang, K.T., and Vallery, D.J., "Water resources of the South Nation River Basin - Summary", Water Resources Report No. 13, Ontario Ministry of the Environment, 1980.
- 14 Chiu, C., "Stochastic open channel flow", J. of the Eng. Mech. Division, ASCE, Vol. 94, No. EM3, 1968.
- 15 Chow, V.T., and Kareliotis, S.J., "Analysis of stochastic hydrologic systems", Water Resources Research, Vol. 6, No. 6, 1970.
- 16 Deininger, R.A., "Linear programing for hydrologic analyses", Water Resources Research, Vol. 5, No. 5, 1969.
- 17 Dooge, J.C., "Linear theory of hydrologic systems", Tech. Bulletin No. 1468, Agr. Res. Serv., U.S. Dept. of Agr., Washington, D.C., 1973.
- 18 Efraymson, M.A., "Multiple regression analysis", in MATHEMATICAL METHODS FOR DIGITAL COMPUTERS, Wiley & Sons, New York, N.Y., 1960.
- 19 Emsellem, Y., and deMarsily, G., "An automatic solution for the inverse problem", Water Resources Research, Vol. 7, No. 4, 1971.
- 20 Eriksson, E., "Cross spectrum analysis of groundwater levels in an esker", Nordic Hydrology, Vol. 1, No. 3, 1970.
- 21 Evans, D.H., Harley, B.M., and Brös, R., "Application of linear routing systems to regional groundwater problems", Rep. No. 155, Ralph M. Parsons Lab. for Water Resour. and Hydrodyn., Dept. of Civil Eng., Massachusetts Inst. of Technology, Cambridge, Mass., 1972.
- 22 Fiering, M.B., STREAMFLOW SYNTHESIS, Harvard University Press, Cambridge, Mass., 1967.
- 23 Freeze, R.A., "Theoretical analysis of regional groundwater flow", Sci. Ser. No. 3, Water Resources Branch, Environment Canada, 1969.

- 24 Freeze, R.A., "A stochastic conceptual analysis of one-dimensional groundwater flow in non-uniform homogeneous media", Water Resources Research, Vol. 11, No. 5, 1975.
- 25 Freeze, R.A., and Smith, L., "Stochastic analysis of steady state groundwater flow in a bounded domain, 1, One-dimensional simulations", Water Resources Research, Vol. 15, No. 3, 1979.
- 26 Freund, J.E., MATHEMATICAL STATISTICS, 2nd ed. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1970.
- 27 Graham Engineering Consultants Ltd., "Carp River channelization study", for the Mississippi Valley Conservation Authority, 1974.
- 28 Haan, C.T., STATISTICAL METHODS IN HYDROLOGY, The Iowa State University Press, Ames, Iowa, 1977.
- 29 Hannan, E.J., TIME SERIES ANALYSIS, Wiley & Sons Inc., New York, N.Y., 1962.
- 30 Hino, M., "Stochastic approach to linear and non-linear runoff analysis", Technical Report No. 12, Dept. of Civil Eng., Tokyo Inst. of Technol., Tokyo, July 1972.
- 31 Hipel, K.W., McLeod, A.I., and Lennox, W.C., "Advances in Box and Jenkins modeling, 1. Model construction", Water Resources Research, Vol. 13, No. 3, June 1977.
- 32 Hubbert, M.K., "The theory of groundwater motion", J. of Geology, Vol. 48, No. 8, 1940.
- 33 Huthmann, G., "Modeling of water quality systems by multiple frequency response analysis", in SURFACE AND SUBSURFACE HYDROLOGY, ed. by H.J. Morel-Seytoux, Water Resources Publications, Fort Collins, Colo., 1979.
- 34 Jenkins, G.W., and Watts, D.G., SPECTRAL ANALYSIS AND ITS APPLICATIONS, 2nd ed., Holden-Day, San Francisco, Calif., 1969.
- 35 Julian, P.R., "Variance spectrum analysis", Water Resources Research, Vol. 3, No. 3, Third Quarter, 1967.
- 36 Kostuch Associates Ltd., "South Nation River channel improvement above Chesterville", for the South Nation River Conservation Authority, 1973.

- 37 Law, A.G., "Stochastic analysis of groundwater level time-series in the western United States", Hydrology Paper No. 68, Colorado State University, Fort Collins, Colo., May 1974.
- 38 Matalas, N.C., "Time series analysis", Water Resources Research, Vol. 3, No. 3, 1967.
- 39 Meyboom, P., "Patterns of groundwater flow, Prairie Profile", Proc. Hydrol. Symp. No. 3, 1963.
- 40 Natale, L., and Todini, E., "A stable estimator for linear models, 2. Real world applications", Water Resources Research, Vol. 12, No. 4, 1976.
- 41 Nemeth, E., Dr., HYDROMECHANIKA, 1st ed., Textbook Publishing Co., Budapest, 1963. (in Hungarian)
- 42 Neumann, S.P., and deMarsily, G., "Identification of linear systems response by parametric programming", Water Resources Research, Vol. 12, No. 2, 1976.
- 43 Noether, G.E., INTRODUCTION TO STATISTICS - A FRESH APPROACH, 1st ed., Houghton Mifflin Co., Boston, Mass., 1971.
- 44 O'Connell, E.P., "Stochastic modelling of long term persistence in streamflow sequences", Ph.D. thesis, University of London, London, 1974.
- 45 Rao, A.R., Rao, R.G.S., and Kashyap, R.L., "Stochastic models for groundwater levels", Technical Report No. 67, Water Resources Research Centre, Purdue University, West Lafayette, Ind., August 1975.
- 46 Remson, I., Appel, C.A., and Webster, R.A., "Groundwater models solved by digital computer", Proc. Amer. Soc. Civil Eng., J. of Hydr. Div., Vol. 91, No. HY3, 1965.
- 47 Rodriguez-Iturbe, I., "The application of cross-spectral analysis to hydrological time-series", Hydrology Paper No. 24, Colorado State University, Fort Collins, Colo., 1967.
- 48 Rodriguez-Iturbe, I., and Nordin, C.F., "Time series analyses of water and sediment discharges", Bull. Int. Ass. of Sci. Hydrol., Vol. 13, No. 2, 1968.

- 49 Roesner, L.A., and Yevjevich, V., "Mathematical models for time-series of monthly precipitation and monthly runoff", Hydrology Paper No. 15, Colorado State University, Fort Collins, Colo., 1966.
- 50 Salas-LaCruz, J.D., and Yevjevich, V., "Stochastic structure of water use time-series", Hydrology Paper No. 52, Colorado State University, Fort Collins, Colo., 1972.
- 51 Scheidegger, A.E., "Stochastic models in hydrology", Water Resources Research, Vol. 6, No. 3, 1970.
- 52 Shanane, A.N., Thomas, D., and Bock, P., "Spectral analysis of hydrometeorological time series", Water Resources Research, Vol. 13, No. 1, February 1977.
- 53 Skibitzke, H.E., "Electronic computers as an aid to the analysis of hydrologic problems", Int. Assoc. of Sci. Hydrol., Publ. No. 52, 1960.
- 54 Smith, L., and Freeze, R.A., "Stochastic analysis of steady state groundwater flow in a bounded domain, 2, Two-dimensional simulations" Water Resources Research, Vol. 15, No. 6, 1979.
- 55 Thomann, R.V., "Time series analysis of water quality data", J. of the Hydraulic Division, ASCE, Vol. 93, No. SA1, Proc. Paper 1, 1967.
- 56 Tison, G., Jr., "Fluctuations of groundwater levels", in ADVANCES IN GEOPHYSICS, Vol. 11, ed. by H.E. Landsberg and J. van Miegheem, Academic Press, New York, N.Y., 1965.
- 57 Todd, D.K., "Groundwater flow in relation to a flooding stream", Proc. Amer. Soc. Civil Eng., Vol. 81, 1955.
- 58 Toth, J., "A theoretical analysis of groundwater flow in small drainage basins", Proc. Can. Hydrol. Symp. No. 3, Ottawa, 1962.
- 59 Toth, J., "A theory of groundwater motion in small drainage basins in central Alberta, Canada", J. of Geophys. Res., Vol. 67, No. 11, 1962.
- 60 Walton, C.W., GROUNDWATER RESOURCE EVALUATION, McGraw-Hill Book Co., New York, N.Y., 1970
- 61 Yevjevich, V., STOCHASTIC PROCESSES IN HYDROLOGY, 1st ed., Water Resources Publications, Fort Collins, Colo., 1972.

- 62 Yevjevich, V., " Structural analysis of hydrologic time series", Hydrology Paper No. 56, Colorado State University, Fort Collins, Colo., 1972.