Systematic Development of Methodologies in Planning Urban Water Resources for Medium Size Communities

STOCHASTIC MODELS FOR GROUND WATER LEVELS

by
A. Ramachandra Rao
R. G. Srinivasa Rao
R. L. Kashyap

August 1975

PURDUE UNIVERSITY
WATER RESOURCES RESEARCH CENTER
WEST LAFAYETTE, INDIANA
STOCHASTIC MODELS FOR GROUND WATER LEVELS

by

A. Ramachandra Rao
R. G. Srinivasa Rao
R. L. Kashyap

The work upon which this report is based was supported in part by funds provided by the United States Department of the Interior, Office of Water Research and Technology, as authorized by the Water Resources Research Act of 1964 (PL88-379) as amended.

Period of Investigation: September 1971 - August 1975
Partial Completion Report for OWRT-C-3277
Grant No. 14-31-0001-3713
PREFACE

This report covers part of the work performed under the title II project C-3277 entitled "Systematic Development of Methodologies in Planning Urban Water Resources for Medium Size Communities."

The financial support from the Office of Water Research and Technology has made the study possible. We appreciate the support.

ABSTRACT

The ground water level fluctuations in an aquifer are affected by external causes such as precipitation, outflows from the aquifer, and pumping from the aquifer. In the present study stochastic models have been proposed for the fluctuations in ground water levels.

Both multiinput-single output and multivariate models have been considered. In the univariate models, variables such as precipitation, stream flows and others which have been known to affect the ground water levels have been considered as external inputs. The fluctuations in ground water levels and the causal variables are directly built into the multivariate model. The parameter estimation and the residual tests are discussed in detail. The multivariate model was also tested by simulation.

The results indicate that the ground water level series can be successfully modeled by stochastic models. These models can be used to evaluate the time lag between precipitation and changes in ground water levels, the effect of the streamflows on water levels in adjacent wells and in evaluating the ground water resources in a region.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>I.  Introduction and data used in the study</td>
<td>1.1</td>
</tr>
<tr>
<td>1.1  Introduction and statement of the problem</td>
<td>1.1</td>
</tr>
<tr>
<td>1.2  Data used in the study</td>
<td>1.3</td>
</tr>
<tr>
<td>1.3  Statistical characteristics of the data used in the study</td>
<td>1.13</td>
</tr>
<tr>
<td>1.3.1 Continuous means and standard deviations</td>
<td>1.13</td>
</tr>
<tr>
<td>1.3.2 Histograms of the data</td>
<td>1.14</td>
</tr>
<tr>
<td>1.3.3 Correlograms and power spectra</td>
<td>1.16</td>
</tr>
<tr>
<td>1.3.4 Cross correlation properties of the data</td>
<td>1.22</td>
</tr>
<tr>
<td>II. Single output models for the well water level fluctuations</td>
<td>2.1</td>
</tr>
<tr>
<td>2.1  Model specification</td>
<td>2.1</td>
</tr>
<tr>
<td>2.2  Parameter estimation in the univariate model</td>
<td>2.2</td>
</tr>
<tr>
<td>2.3  The accuracy of the parameter estimates</td>
<td>2.6</td>
</tr>
<tr>
<td>2.4  Probability distribution of the random inputs V(·)</td>
<td>2.8</td>
</tr>
<tr>
<td>2.5  Parameter estimates of single output models - results</td>
<td>2.8</td>
</tr>
<tr>
<td>2.6  Characteristics of residuals of single output models</td>
<td>2.12</td>
</tr>
<tr>
<td>2.6.1 Simple statistics of residuals and prediction errors</td>
<td>2.12</td>
</tr>
<tr>
<td>2.6.2 Testing the residuals for whiteness</td>
<td>2.16</td>
</tr>
<tr>
<td>2.6.3 The results of the whiteness tests</td>
<td>2.21</td>
</tr>
<tr>
<td>2.6.4 Test for detecting periodicities in the residuals</td>
<td>2.23</td>
</tr>
</tbody>
</table>
III. The multivariate model

3.1 The model specification 3.1
3.2 The parameter estimation scheme 3.2
  3.2.1 Significance of additional terms $A_j$ 3.5
  3.2.2 Prediction error variance 3.6
3.3 Multivariate models for Indiana and Wisconsin data 3.6
3.4 Characteristics of residuals from the multivariate models 3.10
3.5 Whiteness tests on residuals - autocorrelation tests 3.13
3.6 Cross correlation tests on residual sequences 3.15

IV. Simulation of the multivariate model 4.1

4.1 Introduction 4.1
4.2 Comparison of statistical properties of observed and generated data 4.3
  4.2.1 Histograms 4.3
  4.2.2 The mean and variance 4.3
  4.2.3 Skewness coefficients 4.6
  4.2.4 Lag one correlation coefficient 4.6
  4.2.5 Lag zero and lag one cross correlation matrices 4.7
  4.2.6 Correlogram 4.7
  4.2.7 Power spectral density 4.8
  4.2.8 Summary 4.13
4.3 The multivariate AR(1) model 4.13
  4.3.1 Results 4.14

V. Discussion and conclusions 5.1

VI. List of references 6.1

VII. Computer programs 7.1
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Location of stations</td>
<td>1.5</td>
</tr>
<tr>
<td>1.2</td>
<td>Observed data traces and continuous means and standard deviations</td>
<td>1.9</td>
</tr>
<tr>
<td>1.3</td>
<td>Histograms and monthly means and standard deviations of observed data</td>
<td>1.15</td>
</tr>
<tr>
<td>1.4</td>
<td>Correlograms and power spectra of observed data</td>
<td>1.17</td>
</tr>
<tr>
<td>1.5</td>
<td>Correlograms and power spectra of transformed Wisconsin data</td>
<td>1.20</td>
</tr>
<tr>
<td>1.6</td>
<td>Correlograms and power spectra of transformed Indiana data</td>
<td>1.21</td>
</tr>
<tr>
<td>1.7</td>
<td>Cross covariance and coherence square of observed data</td>
<td>1.25</td>
</tr>
<tr>
<td>1.8</td>
<td>Cross covariance and coherence square of transformed data</td>
<td>1.26</td>
</tr>
<tr>
<td>2.1</td>
<td>One-month-ahead predicted (triangles) and observed (circles) ground water levels, prediction errors, error mean square, and residual mean square values, and histograms of residuals</td>
<td>2.15</td>
</tr>
<tr>
<td>2.2</td>
<td>Correlation characteristics of residuals</td>
<td>2.27</td>
</tr>
<tr>
<td>3.1</td>
<td>Histograms of residuals from multivariate models</td>
<td>3.14</td>
</tr>
<tr>
<td>3.2</td>
<td>Correlograms and power spectral densities of residuals from multivariate models</td>
<td>3.17</td>
</tr>
<tr>
<td>3.3</td>
<td>Cumulative periodograms of residuals from multivariate models</td>
<td>3.19</td>
</tr>
<tr>
<td>3.4</td>
<td>Cross correlation coefficients of residuals from multivariate models</td>
<td>3.21</td>
</tr>
<tr>
<td>3.5</td>
<td>Coherence diagram of residuals from multivariate models</td>
<td>3.23</td>
</tr>
<tr>
<td>3.6</td>
<td>Integrated phase spectra of residuals from multivariate models</td>
<td>3.25</td>
</tr>
<tr>
<td>3.7</td>
<td>Integrated co-spectra of residuals-multivariate models</td>
<td>3.26</td>
</tr>
<tr>
<td>4.1</td>
<td>Histograms of observed and back transformed generated data</td>
<td>4.4</td>
</tr>
<tr>
<td>4.2</td>
<td>Correlograms and power spectral densities of observed and back transformed generated data</td>
<td>4.9</td>
</tr>
<tr>
<td>4.3</td>
<td>Correlograms of transformed observed and generated data</td>
<td>4.11</td>
</tr>
<tr>
<td>4.4</td>
<td>Power spectral densities of transformed observed and generated data</td>
<td>4.12</td>
</tr>
</tbody>
</table>
LIST OF TABLES

1.1 Details of location of stations and data .............................................. 1.6
1.2 Statistics of data used in the study .......................................................... 1.7
1.3 Periodic components $\Theta_i$ (equation 1.1) ............................................. 1.12
2.1 Parameter estimates and their standard errors in single output models ...... 2.10
2.2 Characteristics of residuals and one-month-ahead prediction errors from single output models .......................................................... 2.13
2.3 $\chi^2$-test for the normality of residuals .................................................. 2.22
2.4 Results of the $F$ and $\chi^2$ tests on residuals ......................................... 2.22
2.5 Results of the Portmanteau test ............................................................... 2.24
3.1 Matrix coefficients $A_p$ ............................................................................ 3.8
3.2 $F_0$ and $G$ test statistics and final models .............................................. 3.9
3.3 Characteristics of residuals from multivariate models and results from $\chi^2$ test for normality ............................................................. 3.11
3.4 Results of whiteness tests on residuals from multivariate models ............ 3.16
4.1 Statistical characteristics of the observed and back transformed generated data .......................................................... 4.5
4.2 Lag-zero and lag-one cross correlation matrices of transformed observed and generated data ............................................................. 4.5
4.3 Estimates of $A$ and $B$ matrices (S1-Series) of the Matalas model .......... 4.15
4.4 Comparison of statistical characteristics of back transformed generated (by Matalas model) and observed data .............................................. 4.17
4.5 G-test on residuals from Matalas model .................................................... 4.18
I. INTRODUCTION AND DATA USED IN THE STUDY

1.1 Introduction and Statement of the Problem

The ground water resources of a region are traditionally modeled by means of analog, digital, and hybrid computer models (Bear, 1972). These models are deterministic. The inputs to the ground water system in a region such as the precipitation and inflows from adjoining aquifers are random processes. The outputs from the system such as the outflows to streams or to other aquifers are also random processes. Furthermore, the amount of water extracted from the aquifer by pumping, because of its dependence on demand, is variable and in several cases nonstationary. As the ground water system is affected by random factors such as those mentioned above and many others, the ground water levels in an aquifer are stochastic, and can be modeled by stochastic models, as demonstrated in this report.

In deterministic analysis of water table fluctuations, Tison (1965) has modeled large unconfined aquifers by using the continuity equation. The outflow from the aquifer is treated as linearly varying with the elevation of the water table and the inflow as a sinusoidal function of time. The time lag between inflows and outflows was estimated by using the model and the model was shown to be adequate. Finally, Tison considered the problems of sudden modifications in ground water levels, and fluctuations in the ground water levels in alluvial aquifers affected by
streams with changing levels, also by using deterministic models. A brief review of the literature dealing with the fluctuations of water table is also given by Tison.

A stochastic model of ground water levels may be used to forecast the future water levels in the aquifer, which in turn is useful in the management of the aquifer. It is also useful in estimating the response of the aquifer to precipitation and finally in the drought analysis. Consequently, the analysis of ground water levels and the other factors which affect them is receiving some attention recently.

A statistical analysis of the long term (seventy years) observations of ground water levels in a well in Czechoslovakia has been reported by Kriz (1972). The frequency distributions of ground water levels, the level-duration curves, the variation in the water levels in individual months, and a hydrological classification of the period of record have been discussed by Kriz. No attempt has been made by Kriz to propose stochastic models of the data.

Time series models have been fitted to thirty years of ground water level data sampled at intervals of ten days by Eriksson (1970). The data used were from the Ultuna area, south of Uppsala, in Sweden. The wells are located near Fyris River. The deterministic components are removed and the residuals were fitted by AR models. Models were fitted to the series of water levels in the river Fyris, the total ten-day precipitation, and the ten-day means of air temperature at Ultana. A model based on the continuity equation, which is similar to Tison's model has also been proposed and discussed by Eriksson.

The objectives of the present study are to develop stochastic models for the interaction of rainfall and ground water levels. Univariate models
relating the ground water levels and rainfall at several stations are developed and validated. A multivariate model of the ground water levels and rainfall is also formulated and its validity tested by simulation.

The report is organized as follows. The characteristics of the data used in the study are discussed in this chapter. The univariate models are discussed in Chapter II. The multivariate model is discussed in Chapter III. Results of simulation of the multivariate model are discussed in Chapter IV. The results are discussed and a set of conclusions are presented in Chapter V.

1.2 Data used in the Study

The data used in the study were measured at two locations, one of which is in Indiana and the other in Wisconsin. The rainfall, the stages of the Wabash River, and the water levels in two wells at Lafayette, Indiana, constituted the data from Indiana stations. One of the wells (L1) at West Lafayette is located close to the Wabash River and hence reflects the effects of stage variations in the Wabash River, whereas the other well (L2) is located in the Purdue University Campus and is not as much affected by the changes in the Wabash River flows.

The water levels in a well (W1) located in Wood County in north central Wisconsin, and the rainfall measured at Stevens Point, Marshfield, Wisconsin Rapids and Neilsville, which are around W1 and located in Wisconsin, were also used in the study. The well W1 is not pumped and is used only as an observation well.

The geology of the area, and the dependence of Marshfield and Neilsville on ground water for their domestic and industrial supplies has been discussed by Drescher (1956), and Audini et al. (1959). The trends in the ground water levels in Wisconsin (through 1966) have been
discussed by DeVaul (1967). The Wisconsin River is about 16 miles away from the well W1, and the water levels in the aquifer in which W1 is located do not seem to be significantly affected by the river.

The locations of the stations, the data from which were used in the study, are shown in fig. 1.1. Among the Wisconsin stations, Wisconsin Rapids is closest to W1, whereas Neillsville is farthest from W1. The Lafayette stations are not as far apart as the Wisconsin stations. The rainfall data from Lafayette are included mainly to investigate the effects of rainfall on the changes in level in L2 at Lafayette. The flow in Wabash River at Lafayette and the changes in water level in L1 are influenced not only by the rainfall at Lafayette but also by rainfall upstream. The details of locations of the stations and the symbols used to designate them are shown in Table 1.1.

The monthly data were used in the study. The monthly rainfall data were taken from the publications of the National Weather Service. The daily stage values in the Wabash River at Lafayette were averaged to obtain the monthly values. The well water level data, which are reported at irregular intervals, were also averaged to obtain the monthly values. The water level data were taken from the water supply papers of the USGS.

The elementary statistics of the data used in the study are shown in Table 1.2. The variance of water levels in L1 is much larger than the variance of levels in L2, which indicates the general effect of the Wabash River stage on increasing the fluctuations of water levels in L1. The variance of rainfall at Lafayette is relatively large, indicating the rather large variability of monthly rainfall at Lafayette. The Wabash River stages and the Lafayette rainfall are highly skewed,
FIGURE 1.1 LOCATION OF STATIONS
Table 1.1
Details of Location of Stations and Data

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Measured at</th>
<th>Duration of data</th>
<th>Symbols used in the text</th>
<th>Source of data and remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>Purdue Ag. Farm West Lafayette, Ind. Lat: 40° 28' N Long: 87° 00' W</td>
<td>1945 - 1970</td>
<td>&quot;Lafayette&quot; (LAF)</td>
<td>National Weather Service</td>
</tr>
</tbody>
</table>
| Water levels in well TC4 | NE_{1/4}, SW_{1/4}, NE_{1/4}  
Sec 20, T23N, R4W Canal St., Lafayette, Ind. | 1945 - 1970      | "L_1"                    | U.S.G.S. Water Supply Papers. Elevation recorded with respect to MSL |
| Water levels in well TC7 | SE_{1/4}, SE_{1/4}, Sec 13, T23N, R5W  
Purdue Univ. Campus West Lafayette, Ind. | 1945 - 1970      | "L_2"                    | U.S.G.S. Water Supply Papers. "Ground water levels in North Eastern States." Elevations recorded with respect to MSL |
| Precipitation        | Wisconsin Rapids, Wisconsin                     | 1945 - 1964      | WR                       | National Weather Service                                        |
| Precipitation        | Stevenspoint, Wisconsin                         | 1945 - 1964      | SP                       | National Weather Service                                        |
| Precipitation        | Marshfield, Wisconsin                            | 1945 - 1964      | MF                       | National Weather Service                                        |
### Table 1.2

**Statistics of Data Used in the Study**

**Wisconsin Data**

whereas the water levels in L1 and L2 are not as highly skewed as indicated by the skewness coefficients given in Table 1.2.

The correlation coefficients of water levels in L1, L2, Wabash stages and Lafayette rainfall given in Table 1.2 indicate that the levels in L1 are highly correlated with stages in Wabash, and are not as highly correlated with Lafayette rainfall. The water levels in L2, on the other hand, are very weakly correlated with all the other variables such as Wabash stages, rainfall at Lafayette, etc. The lag-one correlation coefficients given in Table 1.2 indicate that the water levels in L1 are highly correlated with the lagged value of L1, and with the lagged value of Wabash river stages. The lag-one correlation coefficient between the water levels in L1 and Wabash river stages is higher than the lag-zero correlation coefficient which indicates the delayed effect of the Wabash stages on the water levels in L1. Once again, the water levels in L2 are weakly correlated with the other variables.

The monthly precipitation from Wisconsin stations vary considerably more than the water levels in W1. Both the precipitation and water level data from Wisconsin are highly skewed, and the rainfall and water level data are weakly correlated. The water levels in W1 exhibit highest correlation with the rainfall at Marshfield whereas it is lowest with the rainfall data from Wisconsin Rapids. This may be due to the fact that the well W1 and Marshfield are both located in the regions of thin glacial deposits of the Iowan age whereas Wisconsin Rapids is in the Central Sand Plain (Drescher, 1956).

The monthly data used in the present study are plotted in fig. 1.2.
FIGURE 1.2  OBSERVED DATA TRACES AND CONTINUOUS MEANS AND STANDARD DEVIATIONS
The rainfall at Lafayette (Labelled LAFAYETTE in fig. 1.2) indicates generally increasing trends initially and decreases later to reach the lowest level at about the 100th sample, after which it appears to oscillate more or less at a stable level. A similar variation, although not as pronounced as that of rainfall may also be seen in Wabash river stages (labelled WABASH). The water levels in L1 and L2 also indicate these changes. The fluctuations in L1 are larger than those in L2, and have higher frequencies than those in L2, which clearly reflects the effect of the flows in Wabash River on the water levels in L1. An annual cycle is clearly apparent in the rainfall, Wabash river stages and the water level fluctuations in L1. The annual cycle is not as apparent in the water level fluctuations in L2.

The rainfall at Stevenspoint, Wisconsin Rapids and Marshfield appears to fluctuate around a constant mean and do not have as much fluctuation as the rainfall at Lafayette. The fluctuations in the levels in W1 are also not as large as those in L1 and L2.

Apart from the observed data, three types of transformed data were also used in the study. In the first of these the deterministic trends were removed from the data by Fourier analysis and the residuals from the Fourier analysis were used. If $x(t)$ represents the observed data sequence, then the transformed data sequence $x_1(t)$ is defined by eq. 1.1

$$x_1(t) = x(t) - \theta_1 \sin \left(\frac{2\pi t}{12}\right) - \theta_2 \cos \left(\frac{2\pi t}{12}\right)$$

$$- \theta_3 \sin \left(\frac{2\pi t}{6}\right) - \theta_4 \cos \left(\frac{2\pi t}{6}\right) \quad (1.1)$$
The coefficients \( \theta_i \) are listed in Table 1.3, for both the Indiana and Wisconsin data. In both the Indiana and Wisconsin data, the annual cycle is much stranger than the semi-annual cycle. The \( x_i(t) \) series from eq. 1.1 is designated S1 in the present report.

In the second type of transformation, \( \bar{x} \), the mean of the \( x(t) \) series is subtracted from \( x(t) \) and the resulting values are divided by \( \sigma \), the standard deviation of the \( x(t) \) series as in eq. 1.2.

\[
x_2(t) = \frac{x(t) - \bar{x}}{\sigma} \quad \text{(1.2)}
\]

The data transformed according to eq. 1.2 are designated S2 in the present report.

In the third type of transformation, \( \bar{x}_s \), the mean of the square root transformed \( x(t) \) series is subtracted from \( \sqrt{x(t)} \) and the resulting values are divided by \( \sigma_s \), the standard deviation of square root transformed data as in eq. 1.3.

\[
x_3(t) = \frac{\sqrt{x(t)} - \bar{x}_s}{\sigma_s} \quad \text{(1.3)}
\]

The data transformed according to eq. 1.3 is designated S3 in the present report.

The S1 sequence is referred to as the detrended data and the S2 and S3 sequences as the normalized or standardized data in the following discussion.
## Indiana Data

<table>
<thead>
<tr>
<th>DATA</th>
<th>12 MONTHS</th>
<th>6 MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SINE</td>
<td>C0SINE</td>
</tr>
<tr>
<td></td>
<td>$\theta_1$</td>
<td>$\theta_2$</td>
</tr>
<tr>
<td>L1</td>
<td>3.1490</td>
<td>-0.1220</td>
</tr>
<tr>
<td>L2</td>
<td>0.3328</td>
<td>-0.4005</td>
</tr>
<tr>
<td>WAB</td>
<td>2.6863</td>
<td>-0.2258</td>
</tr>
<tr>
<td>LAF</td>
<td>0.0194</td>
<td>-0.8836</td>
</tr>
</tbody>
</table>

## Wisconsin Data

<table>
<thead>
<tr>
<th></th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>0.2636</td>
<td>1.4210</td>
<td>0.4122</td>
<td>-0.36670</td>
</tr>
<tr>
<td>WR</td>
<td>-0.6187</td>
<td>-1.4222</td>
<td>-0.2979</td>
<td>-0.00380</td>
</tr>
<tr>
<td>SP</td>
<td>-0.6121</td>
<td>-1.2854</td>
<td>-0.1109</td>
<td>0.04779</td>
</tr>
<tr>
<td>MF</td>
<td>-0.7244</td>
<td>-1.5010</td>
<td>-0.1230</td>
<td>0.04186</td>
</tr>
</tbody>
</table>

### Table 1.3 Periodic Components $\theta_i$ (Equation 1.1)
1.3 Statistical Characteristics of the Data Used in the Present Study

1.3.1 Continuous Mean and Standard Deviations

The continuous mean values of the variates were computed by 1.4, where \( x(i) \) may represent the monthly rainfall, water levels, etc.

\[
\bar{x}(j) = \frac{1}{j} \sum_{i=1}^{j} x(i).
\] (1.4)

The continuous standard deviations given by 1.5 were also computed.

\[
S(j) = \sqrt{\frac{\sum_{i=1}^{j} (x(i) - \bar{x}(j))^2}{j-1}}.
\] (1.5)

The continuous means and standard deviations of the data used in the present study are shown in fig. 1.2.

The continuous means and standard deviations of both the rainfall at Lafayette and the Wabash River stages fluctuate markedly up to about 150 samples after which they appear to remain approximately constant. The means and standard deviations of water levels in L1 and L2 markedly vary although the water levels in L1 appear to reach a steady state towards the end of the available observations. On the other hand, the continuous means of water levels in L2 changes over the entire sample size.

The continuous means and standard deviations of Wisconsin data reach a steady value with very few samples and remain close to that value. The continuous means and standard deviations in WI show larger variability than the continuous means of the precipitation data from Stevens Point and other stations. These fluctuations in the continuous
means and standard deviations in the water levels in W1, however, are small compared to those from L1 and L2.

The variation in means and standard deviations in water levels in L1, L2 and W1 and the stability of the precipitation means and standard deviations indicates the overall effects of pumping in these aquifers on the well water levels. The effect of pumping from the aquifers appears to be smallest on the levels in W1, and greatest on the levels in L2. The levels in L1 follow the trend in the Wabash stages.

1.3.2 Histograms of the Data

The histograms of the observed data, which are shown in Fig. 1.3, show that the data are strongly skewed. The rainfall data from Lafayette as well as from the Wisconsin stations show a negative skewness. The histogram of Wabash river stages is similarly skewed as that of rainfall at Lafayette. The water level data, on the other hand, are not as skewed as the rainfall data. However, a predominance of small fluctuations in the water levels in L1 and W1 is apparent. The fluctuations in L2 are comparatively larger and their distribution is flatter than those in L1 and W1.

The monthly means and standard deviations for different months (January, February, etc.) of precipitation data shown in Fig. 1.3 indicate a roughly constant mean over the different months. The monthly standard deviations of precipitation data, however, show a pronounced seasonal effect. The monthly means of the water levels do not vary very much. The monthly standard deviations of water levels indicate a more pronounced variation over the year which is not as pronounced as that of the rainfall. It is interesting to note that the monthly means and standard deviations
of rainfall and water levels shown in Fig. 1.3 vary in opposite senses. The largest means and standard deviations of precipitation correspond to the smallest means and standard deviations of water levels and vice versa, again indicating the existence of a lag effect between the precipitation and ground water levels.

1.3.3 Correlograms and Power Spectra

The correlograms and power spectra of the data used in the present study are shown in Fig. 1.4. The autocovariance $\gamma(k)$ of the variables $x(i)$, such as the rainfall, water levels, etc., at lag $k$ were computed by using eq. 1.6.

$$\gamma(k) = \frac{1}{N-K} \sum_{i=1}^{N-K} [x(i) - \bar{x}(N)] [x(i+k) - \bar{x}(N)] \quad (1.6)$$

$$k = 0, 1, \ldots, M; \quad \bar{x}(N) = \frac{1}{N} \sum_{i=1}^{N} x(i)$$

$M = \text{maximum lag}$

In eq. 1.6, $N$ is the total number of data points. The correlation coefficients $r(k)$ were obtained by dividing the autocovariances by the variance $\gamma(0)$. The correlation coefficients are shown as $R(k)$ in fig. 1.4.

The power spectral density functions of the data used in the study are also shown in fig. 1.4. The raw estimates of the power spectrum were obtained by using eq. 1.7.

$$PS(\omega_n) = \frac{2}{\pi} \sum_{k=0}^{M} E_k \gamma(k) \cos \frac{hk\pi}{M} \quad (1.7)$$
\[ \omega_h = \frac{h\pi}{M} = \text{frequency in radians/month} \]

\[
E_k = \begin{cases} 
1 & 0 < k < M \\
\frac{1}{2} & k = 0, M 
\end{cases}
\]

\[ h = 0, 1, 2, \ldots, M \]

\[ M = \text{maximum lag} \]

\[ \gamma(k) = \text{autocovariance at lag } k \]

\[ \text{PS}(\omega_h) = \text{raw estimate of the power spectral density at } \omega_h. \]

These raw estimators of the power spectral density were smoothed by using the Hamming window (Jenkins and Watts, 1968) to obtain the final estimates of the power spectrum. The smoothed estimators of the ordinates of power spectrum are given by eq. 1.8.

\[
\begin{align*}
  k = 0: \quad & S(\omega_0) = 0.54 \text{PS}(\omega_0) + 0.46 \text{PS}(\omega_1) \\
  0 < k < M: \quad & S(\omega_k) = 0.23 \text{PS}(\omega_{k-1}) + 0.54 \text{PS}(\omega_k) + 0.23 \text{PS}(\omega_{k-1}) \\
  k = M: \quad & S(\omega_M) = 0.54 \text{PS}(\omega_M) + 0.46 \text{PS}(\omega_{M-1}).
\end{align*}
\]  

The accuracy of computations was checked by computing the area under the power spectrum which must be equal to the variance of the process. Further computational details are found in Dixon (1970).

The correlograms and power spectral density functions of the observed data used in the study are shown in fig. 1.4. The 95% confidence limits of (Box and Jenkins, 1970) of the correlation coefficients are also indicated on the correlograms.

The correlograms of the observed precipitation data indicate the presence of an annual cycle which is stronger for the Wisconsin data
than for the Indiana data. The prominence of the annual cycle is also apparent in the power spectral densities of the precipitation data. There is some power at low frequencies in the precipitation data from Lafayette, Stevens Point and Wisconsin Rapids which is surprisingly not apparent in the power spectral density of the precipitation data from Marshfield.

The correlogram and power spectral density of the observed Wabash river stage data indicates the existence of strong annual periodicity and some concentration of power at low frequencies.

The low frequency nature of the observed water level data, which were hinted by the original data traces (fig. 1.2) are clearly brought about by the plots of correlograms and power spectral densities of the observed water levels. The annual cycle is superimposed on the low frequency oscillations in the data from L1 and W1, whereas it is absent from the data from L2. The comparatively strong annual cycle in L1 may be attributed to its proximity to the Wabash River. The correlogram of the observed water levels in L2 indicates the highly correlated nature of the water levels in L2.

From the above discussion it is clear that the models fitted to the observed data must account for the annual cycle and the lower frequency effects, especially in the well water level data.

The correlograms and power spectral densities of the transformed (S1, S2, S3) data are shown in figs. 1.5 and 1.6 respectively for the data from Wisconsin and Indiana. The correlograms of the transformed precipitation data show negligibly small correlations at all lags, in comparison with the correlations in the observed data. Consequently,
FIGURE 1.5 CORRELOGRAMS AND POWER SPECTRA OF TRANSFORMED WISCONSIN DATA
it can be concluded that the annual cycle accounts for most of the correlation in the precipitation data. Both the detrending (S1) and normalizing (S2, S3) operations appear to be equally effective in eliminating the periodicity present in the precipitation data.

Elimination of the trends from Wabash river stages by the transformation S1 exaggerates the low frequency components present in the Wabash stage fluctuations. As the low frequency components are predominant in the observed water level data from W1 and L2 (fig. 1.2), the detrending (S1) and normalizing (S2, S3) operations do not alter the correlation structure significantly. The correlation structure of the water levels in L1 is significantly altered by the transformations.

In conclusion, the data transformations appear to eliminate most of the apparent periodicities present in the precipitation, Wabash river stage and of water level data from L1. The water levels in L2, and to a lesser extent in W1 do not have a predominant annual periodicity and hence the data transformations have lesser effect on changing their original correlation structure. The transformed data, which appear to be approximately stationary are used in the multivariate models which are discussed in Chapter III.

1.3.4 Cross Correlation Properties of the Data

The cross correlation among the several time series was examined by computing the cross covariances, cross spectra, etc. The computational procedure is described below.

The cross covariance at lag k between two series $x(i)$ and $y(i)$ is given by eq. 1.9.
\[ \gamma_{xy}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} x(i) y(i+k) \]  \hspace{1cm} (1.9) \\
\[ \gamma_{xy}(-k) = \frac{1}{N-k} \sum_{i=1}^{N-k} x(i+k) y(i) \] \\
\[ k = 0, 1, 2, \ldots, M. \]

The cross spectrum \( CS_{xy}(\omega_h) \) at frequency \( \omega_h \) is given by 1.10, where \( c_{xy}(\omega_h) \) is the co-spectrum \( Q_{xy}(\omega_h) \) is the quadrature spectrum defined in eqs. 1.11 and 1.12.

\[ CS_{xy}(\omega_h) = C_{xy}(\omega_h) + i Q_{xy}(\omega_h) \]  \hspace{1cm} (1.10) \\
\[ C_{xy}(\omega_h) = \frac{1}{\pi} \sum_{k=0}^{M} E_k [\gamma_{xy}(k) - \gamma_{xy}(-k)] \cos \frac{hk\pi}{M} \]  \hspace{1cm} (1.11) \\
\[ Q_{xy}(\omega_h) = \frac{1}{\pi} \sum_{k=0}^{M} E_k [\gamma_{xy}(k) - \gamma_{xy}(-k)] \sin \frac{hk\pi}{M}. \]  \hspace{1cm} (1.12)

The co- and quadrature spectra are smoothed by "Hamming window" as in eq. 1.8. The smoothed estimators of co- and quadrature spectra are used to estimate the amplitude (eq. 1.13) and phase (eq. 1.14) spectra. Let the amplitude be denoted by \( A_{xy}(\omega_h) \) and phase by \( P_{xy}(\omega_h) \).

\[ A_{xy}(\omega_h) = \sqrt{C_{xy}^2(\omega_h) + Q_{xy}^2(\omega_h)} \]  \hspace{1cm} (1.13) \\
\[ P_{xy}(\omega_h) = \text{Arg}[C_{xy}(\omega_h) + Q_{xy}(\omega_h)]. \]  \hspace{1cm} (1.14)

The coherence square between \( x(i) \) and \( y(i) \) is represented by \( CO_{xy}(\omega_h) \) and is given by eq. 1.15

\[ CO_{xy}(\omega_h) = \frac{[A_{xy}(\omega_h)]^2}{S_x(\omega_h) \cdot S_y(\omega_h)}. \]  \hspace{1cm} (1.15)
In eq. 1.15 $S_x(\omega_h)$ and $S_y(\omega_h)$ are the ordinates of power spectral density functions of $x$ and $y$ sequences.

The computational details of the cross covariance properties mentioned above are found in Jenkins and Watts (1968) and in Dixon (1970). A discussion of the hydrologic applications of cross spectral properties is found in Kisiel (1969).

As the cross correlation properties are more important in the context of multivariate models, the cross correlation properties of the transformed data are discussed below. These data, as mentioned earlier, are used to construct the multivariate models discussed in Chapter III.

The cross covariance (eq. 1.9), which indicates the correlation between two lagged time series in the time domain and the coherence square (eq. 1.15), which reflects the same property in the frequency domain, are shown in figs. 1.7 and 1.8.

The cross covariance between the detrended (S1) and transformed (S2, S3) levels in L1 and precipitation at Lafayette (indicated as "L1 - Lafayette" in fig. 1.7) is negligibly small at all lags. The squared coherence between levels L1 and precipitation at Lafayette indicates that approximately 50% of variation between levels in L1 and Wabash stages is higher than that between levels in L1 and precipitation at Lafayette indicating the greater influence of the Wabash stages on variations in L1 than the precipitation. The predominance of the low frequency effects in the coherence between L1 and Wabash is also noteworthy.

Perhaps the most interesting relationship is between the levels in L1 and L2. The cross covariances between the levels in L1 and L2 are
CROSS COVARIANCE AND COHERENCE SQUARE OF OBSERVED DATA
FIGURE 1.8  CROSS COVARIANCE AND COHERENCE SQUARE OF TRANSFORMED DATA
not much stronger than those between levels in L1 and precipitation. Consequently, although the wells L1 and L2 are in the same aquifer, the fluctuations in them are not highly correlated. This can be explained by the fact that the fluctuations in levels in L1 are affected by the Wabash stage variations whereas those in L2 are not.

The cross covariances between the precipitation at Stevens Point, Wisconsin Rapids and Marshfield and the water levels in W1 (fig. 1.7) indicate the response of the aquifer in which the well W1 is located to the variations in precipitation. The precipitation at Stevens Point, Wisconsin Rapids and Marshfield affect the levels in the well W1 with lags of approximately 5 to 6 months. This aspect was also brought out by the variation in monthly means and standard deviations (fig. 1.3). The predominance of low frequency components in the data from W1 are also apparent in the plots of coherence square between W1 and the precipitation at other stations.

The cross covariance properties of the detrended (S1) and normalized (S2, S3) (fig. 1.8) data indicate that the cross correlation relationships between the water levels in the wells and the other series are not entirely eliminated by transformation of the data. Consequently, a multivariate model may be fitted to the transformed data to extract the information available in these series.
II. SINGLE OUTPUT MODELS FOR THE WELL WATER LEVEL FLUCTUATIONS

In this chapter we will consider univariate models for the water level fluctuations in the wells. The precipitation and the stage variations in the river and other available information which affect the water levels in wells will be considered as inputs to the model. The parameters of the proposed model are estimated by using a recursive algorithm. The residuals from the model are tested for whiteness and presence of periodicity. The ability of the model to predict the well water levels one month ahead is also considered.

2.1 Model Specification

Let \( Y_1(k) \), \( Y_2(k) \), \( \ldots \) represent the ground water levels in wells 1, 2, etc. at times \( k = 1, 2, \ldots, N \). The precipitation at nearby stations are represented by \( P_1(k) \), \( P_2(k) \), \( \ldots \), etc. The river stages are represented by \( S(k) \). The general univariate model considered in the present study may then be represented as in eq. 2.1, where the fluctuation in ground water levels in well 1 is being modeled.

\[
Y_1(k) = \sum_{j=1}^{n_1} \alpha_j Y_1(k-j) + \sum_{j=1}^{n_2} \alpha_{j+n_1} Y_2(k-j) + \sum_{j=1}^{n_3} \alpha_{j+n_1} P_1(k-j)
\]

\[
+ \sum_{j=1}^{n_4} \alpha_{j+n_2} P_2(k-j) + \sum_{j=1}^{n_5} \alpha_{j+n_3} P_3(k-j) + \sum_{j=1}^{n_6} \alpha_{j+n_4} S(k-j)
\]

\[
+ \sum_{j=1}^{n_7} \alpha_{j+n_5} V(k-j) + U(k-1) + V(k) \quad (2.1)
\]
2.2

\[ U(k) = \alpha_0 + \sum_{j=1}^{n_8} \left( \alpha_j + n'_i + 2j \right) \cos \omega_j k + \alpha_j + n'_i + 2j \sin \omega_j k \]  \hspace{1cm} (2.1a)

\[ \omega_j = \text{annual cycle or its harmonics} = \frac{2\pi j}{12} \]

\[ n'_i = \sum_{j=1}^{i+1} n_j, \quad i = 1, 2, ..., 7. \]

\[ V(k) = \text{random component.} \]

Fluctuations in ground water levels in other wells 2, 3, etc. may also be represented in the same form as eq. 2.1. In eq. 2.1, \( U(k) \) represents the deterministic component of the ground water level series. As discussed in Sec. 1.3.3, the annual and semi-annual periodicities may be quite predominant in the well water level series and these deterministic components may be appropriately represented by \( U(k) \). The random input term \( V(k) \) represents the part of the process which is non-deterministic. The model given in 2.1 is a seasonal autoregressive-moving average model with external inputs.

2.2 Parameter Estimation in the Univariate Model

The following assumptions are made in order to estimate the parameters in eq. 2.1.

(1) Conditions on the random sequence \( V(i) \)

A1: The random inputs \( \{V(i)\} \) have a mean value equal to zero, are independent of each other and obey the conditions given in eqs. 2.2.

\[ E[V^2(k)] = \rho^0 \]

\[ E[V(k) V'_i(k-j)] = 0 \quad \forall \quad j \]  \hspace{1cm} (2.2)
2.3

(2) **Conditions on the characteristic polynomials** $\theta_1(\lambda)$ and $\theta_2(\lambda)$

Let $\theta_1(\lambda)$ and $\theta_2(\lambda)$ be the characteristic polynomials defined by eq. 2.3 in which $\lambda$ is the unit backward operator defined by the relation $\lambda Y(i) = y(i-1)$.

\[
\begin{align*}
\theta_1(\lambda) &= 1 - \sum_{j=1}^{n_1} \alpha_j \lambda^j \\
\theta_2(\lambda) &= 1 + \sum_{j=1}^{n_7} \alpha_{n_6+j} \lambda^j
\end{align*}
\]  
\[\text{(2.3)}\]

**A2:** The polynomial $\theta_1(\lambda)$ has all its zeros outside the Unit Circle and $\alpha_{n_1} \neq 0$.

**A3:** The polynomial $\theta_2(\lambda)$ has all its zeros outside the Unit Circle and it has no common zero with $\theta_1(\lambda)$.

For the purpose of estimating the parameters in the model, eq. 2.1 is written as eq. 2.4

\[
Y_1(k) = (\varphi)^T Z(k-1) + V(k)
\]  
\[\text{(2.4)}\]

\[
\alpha^T = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n)
\]  
\[\text{(2.5)}\]

where

\[
n = \text{the number of parameters to be estimated} = n_1 + \ldots + n_7 + 2n_8 + 1
\]

\[
Z^T(k-1) = [Y_1(k), \ldots, Y_1(k-n_1), Y_2(k-1), \ldots, Y_2(k-n_2), P_1(k-1), \ldots, P_1(k-n_3), \ldots, P_3(k-n_6), S(k-1), \ldots, S(k-n_6), V(k-1), \ldots, V(k-n_7), 1, \cos \omega_1(k-1), \sin \omega_1(k-1), \ldots, \cos \omega_{n_8}(k-1), \sin \omega_{n_8}(k-1)]^T
\]  
\[\text{(2.6)}\]

Let $\hat{\alpha}(k)$ be the estimate of the $\alpha$ vector computed by using the observations and minimizing the quadratic performance index.
2.4

It may often be necessary to compute the estimate $\tilde{a}(k)$ for all values of $k$. Such a necessity might arise, for example, in the management of an aquifer, when the water levels and the yield are forecast on a monthly basis. In such situations it is preferable to compute $\tilde{a}(k+1)$ by using the available estimate $\tilde{a}(k)$. The recursive estimation scheme given below is useful in such instances. The scheme is recursive in the sense that the estimate $\tilde{a}(k+1)$ is computed in terms of $\tilde{a}(k)$, the latest observations of $Y_1(k)$ and the other inputs. Such continuous monitoring of the estimates is also helpful in determining the possible nonstationary aspects which may be present in the process. Furthermore, the recursive estimation procedure is extremely helpful in scanning a number of models for the process in order to select the best model.

The recursive algorithm AL used for the estimation of parameters in the system eq. 2.1 is given below. The details of the development of the algorithm may be found elsewhere (Kashyap and Rao, 1973). The recursive computation of $\tilde{a}(i)$ involves the auxiliary matrices $\bar{S}(i)$ of dimension $(n \times n)$ which can also be recursively computed.

\[
\begin{align*}
\tilde{a}(k+1) &= \tilde{a}(k) + \bar{S}(k+1) Z(k) [Y_1(k+1) - a^T(k) Z(k)] \\
\bar{S}(k+1) &= \bar{S}(k) - [\bar{S}(k) Z(k) Z^T(k) \bar{S}(k) / [1 + Z^T(k) Z(k)]] \\
Z^T(k) &= [Y_1(k-1), ..., \sin \omega_{n_8}(k)] \\
\bar{V}(k) &= Y_1(k) - a^T(k) Z^T(k-1) \\
k &= r + 1, r + 2, ... \\
\bar{V}(k) &= \text{an estimate of } V(k).
\end{align*}
\]
2.5

The algorithm AL must be initialized. The following initialization procedure was adopted in the present study. The \( Z(k), k = n_1+1, \ldots, n_1+2, \ldots, r \) vectors were constructed by using the first \( r \) values of \( Y_1(k), Y_2(k), \ldots \), along with \( r \) Gaussian random numbers as estimates of \( V(i), i = 1, \ldots, r \). The \( S(r) \) matrix and the vector \( \alpha(r) \) were generated by using eqs. 2.7 and 2.8. In the generation of \( V(k) \) values for the initialization, the mean value of \( V(k) \) was assumed to be zero and its variance to be the estimated mean square value of \( Y_1(i) \).

\[
S(r) = \left[ \sum_{k=n_1+1}^{r} Z(k-1) Z^T(k-1) \right]^{-1} \tag{2.7}
\]

\[
\alpha(r) = S(r) \left[ \sum_{k=n_1+1}^{r} Y_1(k) Z^T(k-1) \right] \tag{2.8}
\]

By using the initial estimates \( S(r) \) and \( \alpha(r) \) given by eqs. 2.7 and 2.8 and the algorithm AL, \( V(r+1), S(r+1), \) etc. can be recursively computed.

At this point, a distinction between the residuals and the prediction errors should be made. Let us assume that \( N \) samples are available of \( Y_1(k), Y_2(k), P_1(k), \ldots \). Then the parameter vector \( \alpha(\cdot) \) can be estimated by using the recursive algorithm and the \( N \) available samples. Let this estimate be denoted by \( \alpha(N) \). The residual sequence computed by using the estimate \( \alpha(N) \) is given by eq. 2.9 and is designated \( V_F(k) \).

\[
V_F(k) = Y_1(k) - \alpha^T(N) Z_F(k-1) \tag{2.9}
\]

\[
Z_F(k-1) = [Y_1(k-1), \ldots, V_F(k-1), \ldots, V_F(k-n), \ldots, \sin \omega_n(k-1)]^T
\]
\[ \rho = \text{an estimate of } \rho^0 = \frac{1}{N-n_1} \sum_{k=n_1}^{N} V_F^2(k) \]

The prediction error \( e(k) \) is computed by using information available up to \( k-1 \) and hence it is the strict one step-ahead prediction error. \( e(k) \) is defined in eq. 2.10.

\[ e(k) = Y_1(k) - \tilde{a}^T(k-1) \tilde{Z}^T(k-1). \tag{2.10} \]

\( \bar{V}(k) \) defined in eq. 2.11 is the residual sequence used in the estimation of parameters and differs from \( e(k) \) in the sense that the estimate \( \tilde{a}^T(k) \) used in the computation of \( \bar{V}(k) \) uses the latest measurement \( Y_1(k) \) also.

\[ \bar{V}(k) = Y(k) - \tilde{a}^T(k) \tilde{Z}^T(k-1). \tag{2.11} \]

For a given sample size, the mean square value of \( \bar{V}_F(k) \) is usually smaller than the mean square value of \( \bar{V}(k) \) as the estimate of \( \bar{V}_F(k) \) is based on larger amounts of information.

2.3 The Accuracy of the Parameter Estimates

Under the conditions mentioned in section 2.2, the estimate \( \tilde{a}(k) \) is known to converge to the true value of \( a \) with probability one as the number of samples \( Y_1(k), Y_2(k), \text{ etc.} \) tend to infinity. In addition, if the moving average terms \( V(\cdot) \) are absent from the system eq. 2.1, then the estimates \( \tilde{a}(k) \) may be shown to have minimum variance for large \( N \).

For the case when no moving average terms \( (V(k-1), V(k-2), \ldots) \) are present in the system eq. 2.1, the mean and mean square error matrix of the parameter estimate \( \tilde{a}(N) \) are given by eqs. 2.12 and 2.13.
\[ \| E[\hat{a}(N)|\alpha, \rho^0] - \alpha \| = O(1/N) \quad (2.12) \]

\[ E[(\hat{a}(N) - \alpha)(\hat{a}(N) - \alpha)^T|\alpha, \rho^0] = \frac{R^{-1}_0(N)}{N} \rho^0 + O(N) \quad (2.13) \]

\[ R^{-1}_0(N) = E \left[ \frac{1}{N} \sum_{j=n_1}^N Z(j) Z^T(j) \right]^{-1}. \]

The mean square error matrix (eq. 2.13) of the parameter estimates \( a(N) \) is derived under the assumption that the sequence \( V(\cdot) \) is normally distributed with zero mean and variance \( \rho^0 \). However, in eq. 2.13, the matrix \( R_0(N) \) is unknown, but an estimate of \( R_0(N) \) is given by \( R_1(N) \) which can be evaluated by using \( S(N) \) available from the algorithm AL. Similarly, the unknown \( \rho^0 \) is replaced by its estimate \( \hat{\rho} \). Consequently, eq. 2.13 can be written as eq. 2.15

\[ R^{-1}_1(N) = N S(N) \quad (2.14) \]

\[ E[(\hat{a}(N) - \alpha)(\hat{a}(N) - \alpha)^T|\alpha, \rho^0] \equiv \text{an estimate of the mean square error matrix of} \]

\[ \hat{a}(N) = R^{-1}_1(N)(\hat{\rho}/N) \quad (2.15) \]

The formula given in 2.15 can also be used to obtain some idea of the mean square error of the parameter estimates even when the system eq. 2.1 has moving average terms in it. When moving average terms are present in eq. 2.1, the estimate of the mean square errors of the parameters computed by eq. 2.15 may be considerably different from the exact mean square error of the parameter estimates. In order to obtain an exact estimate of the mean square errors of the parameter estimates in systems with moving average terms, the corresponding parameter estimates
must be computed by the maximum likelihood method, which is computationally difficult and is sensitive to the initial conditions.

2.4 Probability Distribution of the Random Inputs \( V(\cdot) \)

A knowledge of the probability distribution of the random input \( V(i) \) is necessary for the simulation of the \( Y_1(i) \) process. The empirical probability distribution of \( V_F(i) \) may be regarded as a good estimate of the random input \( V(i) \). The histogram of residuals \( V_F(i) \), \( i = 1, 2, \ldots, N \), is computed, which may then be approximated by an appropriate distribution such as the normal distribution \( N(\hat{\mu}, \hat{\sigma}) \) when \( \hat{\mu} \) and \( \hat{\sigma} \) are defined below. The goodness-of-fit between the histogram of \( V_F(i) \) and the fitted normal or other distribution may be checked by using the \( \chi^2 \) or Kolmogorov-Smirnov tests.

\[
\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} V_F(i) \quad (2.16)
\]

\[
\hat{\sigma} = \frac{1}{N} \sum_{i=1}^{N} (V_F(i) - \hat{\mu})^2. \quad (2.17)
\]

We would like to emphasize, however, that our main interest in the distribution of residuals \( V_F(i) \) is only to the extent of their effect on the distribution of the simulated \( Y_1(\cdot) \) series, computed by using the model and the fitted distribution. The histograms of simulated or generated and the observed \( Y_1(\cdot) \) series may be compared, again, by using the goodness-of-fit tests.

2.5 Parameter Estimates of Single Output Models - Results

Several scalar models were selected by fixing different values of \( n_1, n_2, \ldots, \) etc. in eq. 2.1 and the parameters of these models were
estimated by using water level, rainfall and other data and the
algorithm AL. In the process of fitting these models the one step-
ahead prediction errors were also computed. The models were fitted
to the original data and to the transformed data series S1, S2 and
S3. After trying out various combinations of the inputs (rainfall,
river stages, etc.) sinusoidal and moving average terms, two of the
models which gave the smallest one step-ahead prediction error and
residual variance were selected. The parameters of these models and
their standard errors are listed in Table 2.1. Some of the character-
istics of these models are discussed below with reference to parameter
estimates. Only those parameter estimates whose values are greater
than their mean square errors are given in Table 2.1.

(a) **MODELS FOR FLUCTUATIONS IN WATER LEVELS IN WELL L1**

The models for water level fluctuations in L1, constructed by
using the observed data show a weak annual periodicity and a weaker
6-month periodicity. In the first model for the original data from
L1 both the autoregressive and moving average parameters are predominant.
In the second model, in which the fluctuation in the well L2 is also
included, the moving average terms are not significantly different from
zero. It is interesting to note that the parameters related to
precipitation in Lafayette are not very significant, which indicate
a rather weak dependence of the fluctuation in L1 on precipitation.

The models for transformed data are all similar. In these models,
terms related to the effects of changes in the stage and precipitation
are significantly different from zero and so are the autoregressive and
moving average terms. The terms representing the effects of changes in
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.25</td>
<td>0.05</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.36</td>
<td>0.07</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.47</td>
<td>0.08</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.58</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 2.1: Parameter Estimates and Their Standard Errors in Single Output Models
stage have a negative sign which indicates an inverse relationship between the levels in L1 and the Wabash river stages in these models of transformed data.

(b) MODELS FOR WATER LEVEL FLUCTUATIONS IN WELL L2

The parameter estimates in the models of observed data fitted for water level fluctuations in well L2 indicate very weak effects of annual and six month periodic fluctuations. The highly correlated structure of fluctuations in L2 are indicated by the closeness of the autoregressive coefficients to unity (0.979, 0.969, etc.). The water level fluctuations in L2 are not very much affected by the fluctuations in the Wabash River stages, but are affected more by the precipitation in Lafayette. The moving average terms in these models fitted to the original data are not as large as they are for the corresponding models for L1. The terms representing the water levels in L1 are not significantly different from zero, which indicates that the fluctuations in L1 do not affect those in L2. On the other hand, the models fitted to the water level fluctuations in L1 indicate a dependence of fluctuations in L1 on those in L2.

The models for the transformed data indicate very weak effects of precipitation and Wabash River stages on the water level fluctuations in L2, but very strong dependence of fluctuations in L2 on its own lagged values. The moving average terms in the models for the detrended data (S1) are highly significant but they are not as highly significant in the models of S2 and S3.
2.12

(a) Models for Water Level Fluctuations in Well WI

The models fitted to the observed data from WI confirms the presence of weak periodicities in WI, and in the rainfall. The parameter estimates of the AR terms and the moving average terms are also predominant. The rainfall at Wisconsin Rapids and Marshfield have greater effects on the water levels in WI than that at Stevens Point and Neillsville. Stevens Point and Neillsville are farther away from WI than Wisconsin Rapids and Marshfield and hence this effect is easily explainable. The rainfall at Marshfield has a higher and negative effect on the fluctuations in WI whereas the precipitation at Wisconsin Rapids has a weaker and positive effect. The negative effect of the rainfall at Marshfield on the water levels in WI may be explained by the lag which exists between the occurrence of rainfall and the time taken for it to manifest itself as water level changes in WI.

The parameters fitted to the detrended (S1) and normalized (S2, S3) data also indicate the high correlation in the water levels, the small dependence of fluctuations of water levels in WI on rainfall and the presence of strong moving average terms.

2.6 Characteristics of Residuals of Single Output Models

2.6.1 Simple Statistics of Residuals and Prediction Errors

The properties of the residuals $V_e(k)$ computed by eq. 2.9 and the one step-ahead prediction errors $e(k)$ computed by eq. 2.10 are listed in Table 2.2 for the models fitted to the Indiana and Wisconsin data. For the models fitted to the water levels in WI, the ratio of the residual variance to the signal variance varies from 0.203 (OD-1) to
<table>
<thead>
<tr>
<th>Model</th>
<th>Mean</th>
<th>Variance</th>
<th>Mean</th>
<th>Variance</th>
<th>Mean</th>
<th>Variance</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**TABLE 2.2**

CHARACTERISTICS OF RESIDUALS AND ONE-MONTH-AHEAD PREDICTION ERRORS FROM SINGLE OUTPUT MODELS
0.285 (S3-1). In other words, 70-80\% of the variance is explained by the models for fluctuations in W1. The ratio of the one step-ahead error variance to the signal variance varies from 0.272 (S1-2) to 0.346 (OD-1), which indicates that the water levels can be predicted on the average within about 70\% accuracy. The mean values of error and residuals are very close to zero which indicate a lack of bias in prediction.

The observed data, the one step-ahead predictions and the prediction errors are shown for the observed and detrended Wisconsin data in Fig. 2.1. The prediction errors are generally small. The variations in the error mean square and the residual mean square values are also shown in Fig. 2.1. The residual mean square values, because they are estimated from larger amounts of information are smaller than the error mean square values. Although some initial fluctuations are apparent in the error and residual mean square values, they tend to stabilize with larger numbers of observations.

The fluctuations in levels in L1 cannot be as accurately predicted as those in W1 as evidenced by the ratios of error variance to the signal variance (table 2.2). These ratios vary from 0.522 (S1-1) to 0.715 (OD-2). The main reasons for this aspect are that the levels in L1 are affected by the fluctuations in Wabash river stages in a complex fashion and also that the levels in L1 are very much subjected to pumping rates which vary. Consequently, the fluctuations in L1 are not as predictable as the fluctuations in L2. The residual variances of all the models for L1 are smaller than the error variances.
FIGURE 2.1  ONE-MONTH-AHEAD PREDICTED (TRIANGLES), OBSERVED (CIRCLES), GROUND WATER LEVELS, PREDICTION ERRORS, ERROR MEAN SQUARE, AND RESIDUAL MEAN SQUARE VALUES, AND HISTOGRAMS OF RESIDUALS
2.16

Because of the very highly correlated structure of the water levels in L2, they can be predicted much more accurately than the levels in L2. The ratio of the prediction error variance to signal variance for all the models in L2 are of the order of 0.01. The residual variances are also of the same order. The error and residual means are very close to zero.

The histograms of residuals from some of the models fitted to Wisconsin data are shown in Fig. 2.1. The results of χ²-test conducted to test the goodness-of-fit of normal distributions to these histograms are given in Table 2.2. These residual histograms can be adequately fitted by normal distribution.

2.6.2 Testing the Residuals for Whiteness

The random input V(k) was assumed to be zero mean, independent of its lagged values and of the lagged values of the outputs Y_1(k), Y_2(k), etc. in order to estimate the parameters of the system equation (Sec. 2.2). If the given process obeys the difference equation model (eq. 2.1), then the residuals which are computed with the aid of the given observations and the estimated model parameters would possess properties similar to that of a random sequence. In the present section, tests for checking the whiteness of the residual sequences are discussed, and the test for detecting the presence of deterministic sinusoidal trends in the residual sequences is discussed in Sec. 2.6.3. We will briefly discuss the tests first and then give the results.

Let us consider the final model such as that given in 2.18 which was selected on the basis of minimum one step-ahead prediction errors and residual variances as mentioned in Sec. 2.5.
\[ Y_1(k) = a^T(N) Z_F(k - 1) + V_F(k) \]  \hspace{1cm} (2.18)

The residuals \( V_F(k) \), are estimated by eq. 2.9 and the whiteness of these residuals is to be tested.

\[ V_F(k) = Y_1(k) - a^T(N) Z_F(k - 1). \]  \hspace{1cm} (2.9)

Let the mean square value of the residuals be \( \rho^0 \).

The F-test, the \( \chi^2 \)-test and the portmanteau lack of fit test were used to test the whiteness of the residuals \( V_F(\cdot) \). In the following discussion the residual sequence \( V_F(n) \), \( V_F(n + 1), \ldots, V_F(N) \) is designated \( V_F \), and the number of residuals \( (N - n) \) is designated \( N_1 \).

(a) The F-test

The investigation of the lack of correlation among the residuals \( V_F(i) \) can be treated as a hypothesis testing problem with the null hypothesis \( H \) being that "the \( V_F(\cdot) \) sequence is uncorrelated" and the alternative hypothesis \( H \) as that "the residuals obey an AR process of order \( m \) or less." These hypotheses are formally stated below.

\( H_0 \): The random variables \( V_F(k) \) are mutually independent with zero mean and unknown variance \( \rho^0 \), whose estimate is denoted by \( \rho \).

\( H_1 \): The sequence \( V_F(k) \) obeys an AR process of order \( m \) or less as in eq. 2.19

\[ V_F(k) = \sum_{j=1}^{m} \delta_j V_F(k - j) + X(i) \]  \hspace{1cm} (2.19)

\[ \tilde{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)^T \]

\[ E[X^2(i)] = \rho_X. \]
2.18

In eq. 2.19, \( \hat{\theta} \) is unknown and \( \{X(\cdot)\} \) is a sequence of independent and normally distributed variables with zero mean and unknown variance \( \rho_X \). The characteristic polynomial corresponding to the m-th order AR process in eq. 2.19 must have all its zeros outside the unit circle. The decision rule for the F-test may be stated as follows.

**Decision Rule**

\[
g_1^*(V_F) = \begin{cases} 
    d_1(V_F) - \lambda_1(\varepsilon) \leq 0 \rightarrow \text{Accept } H_0 \\
    d_1(V_F) - \lambda_1(\varepsilon) < 0 \rightarrow \text{Reject } H_0.
\end{cases}
\]  

(2.20)

In the above decision rule (2.20), \( \lambda_1(\varepsilon) \) is the threshold depending on \( \varepsilon_1 \), the chosen probability of type I error, and \( d_1(V_F) \) is defined below.

\[
d_1(V_F) = \left[ \frac{\hat{r}_{V_F}(0)}{\hat{r}_X(0)} - 1 \right] \frac{[N]}{m} - 1
\]  

(2.21)

\( \hat{r}_{V_F}(0) \) = an estimate of \( \rho_0 \) under hypothesis \( H_0 = \frac{1}{N_1} \sum_{i=1}^{N} V_i^2(1) \)  

(2.22)

\( \hat{r}_X(0) \) = an estimate of \( \rho_X \) under hypothesis \( H_1 = \frac{\det \hat{\Gamma}(m)}{\det \hat{\Gamma}(m-1)} \)  

(2.23)

\[
\hat{\Gamma}(m) = [\hat{r}_{V_F}(i-j), i,j = 1, \ldots, m]
\]

\[
\hat{\Gamma}(m) = \begin{bmatrix}
    \hat{r}_{V_F}(0), & \hat{r}_{V_F}(1), & \ldots, & \hat{r}_{V_F}(m) \\
    \hat{r}_{V_F}(1), & \hat{r}_{V_F}(0), & \ldots, & \hat{r}_{V_F}(m-1) \\
    \vdots & \vdots & \ddots & \vdots \\
    \hat{r}_{V_F}(m), & \hat{r}_{V_F}(m-1), & \ldots, & \hat{r}_{V_F}(0)
\end{bmatrix}
\]  

(2.24)
\[ \hat{r}_{V_F}(k) = \frac{1}{N_1-k} \left[ \sum_{i=k+n}^{N} V_F(i) V_F(i-k) \right] \]  \hspace{1cm} (2.25)

The computation of the decision function \( d_1(\cdot) \) which is used in
the decision rule 2.20 is straightforward.

The random variable \( d_1(V_F) \) may be shown to be \( F(m, N_1) \) distributed
for large \( N \) under the hypothesis \( H_0 \). Consequently, the values of \( \lambda_1(\varepsilon) \)
may be taken from a table of \( F \)-distributions for a prespecified \( \varepsilon \) value.
The \( F \)-test minimizes the maximum probability of occurrence of type II
error, and the probability of type II error in this test may be shown
to be less than \( (1 - \varepsilon) \). The test is invariant to the scale of
measurements \( V_F(i) \) and can be interpreted as a generalized likelihood
ratio test. Details of \( F \)-test are found in Kasyap and Rao (1976).

(b) The \( \chi^2 \)-test (Test No. 2)

An alternative form of the \( F \)-test discussed above has been proposed
by Hannan (1960). The decision rule for the Hannan's \( (\chi^2) \) test is given
below.

**Decision Rule**

\[
d_2(V_F) = \begin{cases} 
    d_2(V_F) - \lambda_2(\varepsilon) \leq 0 \rightarrow \text{Accept } H_0 \\
    d_2(V_F) - \lambda_2(\varepsilon) < 0 \rightarrow \text{Reject } H_0.
\end{cases}
\]  \hspace{1cm} (2.26)

\[
d_2(V_F) = N \left\{ 1 - \frac{\hat{r}_{\chi}(0)}{\hat{r}_{V_F}(0)} \right\} \]  \hspace{1cm} (2.27)
As the decision rule 2.26 compares the ratio \( \hat{\lambda}_2(0)/\hat{\lambda}_2(\varepsilon) \) to the threshold \( \lambda_2(\varepsilon) \) it can be considered as a variant of the F-test. The value of the threshold \( \lambda_2(\varepsilon) \) for various values of \( m \) can be estimated by a procedure similar to that outlined for F-test, except that the random variable \( d_2(V_F) \) is chi-squared distributed with \( m \) degrees of freedom for large \( N \), provided the residuals \( V_F \) obey \( H_0 \). The values of \( \lambda_2(\varepsilon) \) can be obtained from a \( \chi^2 \)-table. It must be emphasized that the \( \chi^2 \)-test does not possess any of the optimality properties of the F-test discussed earlier.

(c) The Portmanteau Test

Box and Pierce (1971) have given a "goodness of fit" test for testing the whiteness of a sequence of the residuals. This test is designed for accepting or rejecting the hypothesis \( H_0 \), which is the same as for the F- and \( \chi^2 \)-tests discussed above. In the Portmanteau test, the counter-hypothesis is not explicitly mentioned. The decision rule in the Portmanteau test is given below.

**Decision Rule**

\[
g_3(V_F) = \begin{cases} 
  d_3(V_F) - \lambda_3(\varepsilon) \leq 0 \Rightarrow \text{Reject } H_0 \\
  d_3(V_F) - \lambda_3(\varepsilon) < 0 \Rightarrow \text{Accept } H_0.
\end{cases}
\]  

(2.27)

where

\[
d_3(V_F) = N_1 \sum_{j=1}^{m_1} d_j^2(V_F)
\]

\[
d_j(V_F) = \frac{\hat{r}_{V_F}(j)}{\hat{r}_{V_F}(0)}
\]

\( m_1 \) = a positive integer, usually about 0.1 \( N \) or less.
The threshold \( \lambda_3(V_F) \) is determined by the chosen value of \( \varepsilon \), which is the probability of type I error. To determine the threshold value, the fact that for large \( N_1 \) the random variable \( d_3(V_F) \) is chi-square distributed with \( m \) degrees of freedom, if \( V_F \) obeys \( H_0 \), is used. For the given value of \( \varepsilon \), the threshold \( \lambda_3 \) can be obtained by using a table of chi-square values and the decision rule 2.27 can be applied (Box and Jenkins, 1970).

The Portmanteau test does not possess the optimality properties of the F-test. Consequently, the probability of error of type II involved in the Portmanteau test can be large. The Portmanteau test has a stronger tendency than the F-test to allow the acceptance of the \( H_0 \) hypothesis even though it is not valid.

2.6.3 The Results of the Whiteness Tests

The results of the F-test and the \( \chi^2 \)-test applied to the models for water level fluctuations in L1, L2 and W1 are given in Table 2.4. The statistics \( d_1(V_F) \) and \( d_2(V_F) \) for different lags and the threshold values \( \lambda_1(\varepsilon) \) and \( \lambda_2(\varepsilon) \) are also shown in table 2.4 for \( \varepsilon = 0.05 \). The acceptance of the null hypothesis is indicated by A and the rejection by R. Acceptance of the null hypothesis implies acceptance of the residual sequence \( V_F(\cdot) \) as being white.

The results of the whiteness test applied to the models fitted to the observed as well as transformed data from L1 indicate that the residuals can be considered white up to 20 lags for the models fitted to the observed and S1 and S2 data series. The residuals from the models fitted to the S3 sequence can be accepted to be white up to
<table>
<thead>
<tr>
<th>Week</th>
<th>Test</th>
<th>n</th>
<th>m</th>
<th>Last</th>
<th>20</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table 2.4** Results of the F and $\chi^2$ Tests on Residuals
15 lags as both the F-test and $\chi^2$-test indicate the rejection of the null hypothesis for higher lags.

The tests of residuals from the models fitted to the data from L2 indicate that they can be considered white up to 20 lags. Once again the whiteness tests on residuals from the models fitted to the S3 sequence gives doubtful results at lags 20 and 25. The results of whiteness tests of residuals from the models fitted to the data from W1 are similar to the results of whiteness tests of residuals from models for L1 and L2, in the sense that the residual sequences from these models can be considered white up to 20 lags.

The results of the Portmanteau test which are given in Table 2.6 indicate that the residuals from models fitted to the observed and transformed data can be considered to be white up to lags 25 or more. The Portmanteau test is weaker than the $\chi^2$-test and F-test and this fact is reflected in these results.

In conclusion, the residual sequences from the models which are selected for modeling the fluctuations in the water levels in the different wells may be considered white up to approximately 20 to 25 lags. These residuals were also found to be uncorrelated with the observed values.

2.6.4 Test for Detecting Periodicities in the Residuals

The residual sequences must be tested for the presence of sinusoidal trend components because the fitted model may not have the relevant sinusoidal trend terms even though they may be required for the correct model of the process. The cumulative periodogram test proposed by Bartlett (1966), is used for detecting the sinusoidal components of the residuals.
<table>
<thead>
<tr>
<th>LAG</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRIT. VALUE</td>
<td>11.070</td>
<td>16.310</td>
<td>25.000</td>
<td>31.410</td>
<td>37.650</td>
<td>43.770</td>
</tr>
</tbody>
</table>

**Wisconsin Data**

| Well - Wi | OD-1 | 3.109 | 8.447 | 8.983 | 12.191 | 17.100 | 23.975 |
|           |      |       |       |       |       |       |       |
|           | OD-2 | 7.268 | 11.417 | 15.554 | 19.268 | 34.313 | 42.172 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-3 | 2.170 | 4.760 | 5.510 | 6.840 | 15.200 | 21.500 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-3 | 2.140 | 3.160 | 4.130 | 6.910 | 13.100 | 15.560 |
|           |      |       |       |       |       |       |       |
|           | SI-2 | 7.020 | 11.200 | 12.400 | 15.200 | 18.100 | 21.100 |

**Indiana Data**

| Well - L1 | OD-1 | 1.071 | 7.873 | 15.426 | 22.335 | 37.773 | 39.354 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-1 | 2.241 | 10.750 | 17.684 | 22.687 | 33.866 | 35.233 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-3 | 3.141 | 11.686 | 19.740 | 24.289 | 32.336 | 43.130 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-3 | 2.290 | 17.750 | 21.040 | 31.020 | 37.450 | 43.060 |
|           |      |       |       |       |       |       |       |
|           | SI-2 | 6.070 | 17.750 | 23.120 | 37.060 | 39.480 | 42.110 |

|           |      |       |       |       |       |       |       |
|           | OD-2 | 5.713 | 8.357 | 15.729 | 27.212 | 31.619 | 37.920 |
|           |      |       |       |       |       |       |       |
|           | SI-1 | 7.466 | 10.849 | 16.361 | 28.917 | 30.946 | 34.920 |
|           |      |       |       |       |       |       |       |
|           |      |       |       |       |       |       |       |
|           | SI-2 | 8.765 | 11.284 | 17.211 | 24.521 | 31.742 | 36.820 |
|           |      |       |       |       |       |       |       |
|           | SI-2 | 7.036 | 11.021 | 12.091 | 29.020 | 30.569 | 41.230 |
|           |      |       |       |       |       |       |       |
|           | SI-3 | 9.345 | 17.294 | 19.251 | 22.651 | 30.622 | 31.099 |
|           |      |       |       |       |       |       |       |

**Table 2.5** Results of the Portemanteau Test
The periodogram \( I(f_k) \) of the residual \( V_F(j) \) is defined in eq. 2.28 for even \( N \).

\[
I(f_k) = \frac{2}{N} \left( \sum_{j=1}^{N} V_F(j) \cos 2\pi f_k j \right)^2 + \left( \sum_{j=1}^{N} V_F(j) \sin 2\pi f_k j \right)^2, \tag{2.28}
\]

where \( f_k = \frac{k}{N}, \ k = 0, 1, 2, \ldots, N-1 \).

The normalized cumulative periodogram \( C_k \) is defined by

\[
C_k = \frac{k}{\sum_{j=1}^{N} I(f_j)/N \Var(V_F)}; \ k = 0, 1, 2, \ldots, \frac{N}{2}.
\]

The plot of \( C_k \) against \( f_k \) is the cumulative periodogram of the data. If the residual sequence \( V_F(i) \) is a zero mean white noise and is free of deterministic sinusoidal terms then the values of \( C_k \) should be tightly scattered around a straight line from \((0,0)\) to \((0.5,1)\) which we will call "the white noise line."

Similar to the whiteness tests, a quantitative decision rule may be obtained for this test also. Let \( H_0 \) be the hypothesis that the given residual sequence is a zero mean white noise sequence whereas \( H_1 \) is the hypothesis that the residual sequence is a sum of the zero mean white noise and sinusoidal trend terms.

Let \( \epsilon \) the probability of type I error

\[= \text{the probability of rejecting } H_0 \text{ when it is true.} \]

The following decision rule has a prespecified value \( \epsilon \):

"Choose \( H_0 \) only if the entire cumulative periodogram lies within \( \pm K_{\epsilon}/\sqrt{Q} \) around the 'white noise' line joining \((0,0)\) to \((0.5,1)\)."
The parameters $K_e$ and $q$ are defined below.

$$q = (N - 2)/2 \quad \text{if } N \text{ is even}$$

$$q = (N - 1)/2 \quad \text{if } N \text{ is odd.}$$

$K_e$ is the threshold listed below (Bartlett, 1966).

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>$0.05$</th>
<th>$0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_e$</td>
<td>$1.36$</td>
<td>$1.63$</td>
</tr>
</tbody>
</table>

The correlograms, power spectral densities and the cumulative periodograms of residuals are shown in Fig. 2.2 for the Wisconsin data. The 95% confidence limits are also shown on the cumulative periodogram plots. The cumulative periodogram of residuals tightly cluster around the white noise line indicating the absence of sinusoidal components in the residual sequences.

The correlograms of residuals from the different models are within the 95% confidence limits or 2 standard error limits given by $\pm 2/\sqrt{N}$, with a probability of 95%.

In the present case, for the results shown in Fig. 2.2 and for the residuals from models fitted to L1 and L2, the autocorrelation coefficients are small and they are within the two standard error limits. The power spectral density plots also indicate that the residual sequence may be considered white.
FIGURE 2.2  CORRELATION CHARACTERISTICS OF RESIDUALS
III. THE MULTIVARIATE MODEL

3.1 The Model Specification

The ground water level, precipitation and river flow sequences can be formulated as a multivariate model. The advantages of a multivariate model are that they incorporate the cross correlation properties of the series and that they can be used for simultaneous simulation of the series involved. In order to use a univariate model such as that discussed in the previous chapter for simulation of ground water levels, models of precipitation and stages are also required. These models must be developed and examined separately. Such auxiliary models are not required in a multivariate model.

Although multivariate autoregressive (AR) or autoregressive moving average (ARMA) models can be considered for modeling the processes, we have considered a multivariate AR model only. This selection is motivated by the fact that the parameter estimation in multivariate ARMA models is considerably more complex than in multivariate AR models. Consequently, it was decided to consider only the AR model. If the AR model is not suitable to adequately characterize the process, then the more complex ARMA models may be considered.

The parameter estimation scheme used in the present study is based upon the assumption of weak stationarity of the series. Consequently, in the present chapter we deal only with standardized series (S1, S2, and S3) of ground water levels, precipitation series, etc.
3.2

Let $\mathbf{y}(k)$ represent an $n$-dimensional time series. The model to be fitted can be written as in 3.1 where $\mathbf{y}(k)$ and $\mathbf{z}(k)$ are $(n \times 1)$ vectors and $A_j$ are $(n \times n)$ matrices.

$$
\mathbf{y}(k) + A_1 \mathbf{y}(k-1) + A_2 \mathbf{y}(k-2) + \cdots + A_p \mathbf{y}(k-p) = \mathbf{z}(k) \quad (3.1)
$$

$$
k = 1, 2, \ldots, N.
$$

$$
\mathbf{y}^T(k) = [Y_1(k), Y_2(k), \ldots, P_1(k), P_2(k), \ldots, S_1(k), S_2(k), \ldots] (3.2)
$$

The details of the vectors $\mathbf{y}(k)$ for the Indiana and Wisconsin data are as follows.

**Indiana Data**

$$
\mathbf{y}(k) = \begin{bmatrix}
Y_1(k) \\
Y_2(k) \\
P_1(k) \\
S_1(k)
\end{bmatrix}
$$

- Levels in L1
- Levels in L2
- Precipitation at Lafayette
- Wabash river stages

**Wisconsin Data**

$$
\mathbf{y}(k) = \begin{bmatrix}
Y_1(k) \\
P_1(k) \\
P_2(k) \\
P_3(k)
\end{bmatrix}
$$

- Levels in W1
- Precipitation at Wisconsin Rapids
- Precipitation at Stevens point
- Precipitation at Marsh field

Equation 3.1, which is the basic multivariate model used in the present study is a multivariate autoregressive model without any moving average or trend terms. The forward autoregressive counterpart of eq. 3.1 is given in eq. 3.3, in which $\bar{A}_j$ are the $(n \times n)$ coefficient matrices and $\mathbf{y}(k)$ and $\mathbf{z}(k)$ are $(n \times 1)$ vectors.

$$
\mathbf{y}(k) + \bar{A}_1 \mathbf{y}(k+1) + \bar{A}_2 \mathbf{y}(k+2) + \cdots + \bar{A}_p \mathbf{y}(k+p) = \mathbf{z}(k). \quad (3.3)
$$
3.2 The Parameter Estimation Scheme

The parameter estimation scheme used in the present study is recursive in parameter estimates and was proposed by Whittle (1963). The estimation scheme is based on fitting both forward and backward autoregressive schemes to the data. Let us assume that a first order AR scheme has been fitted to the data and the resulting backward and forward models are as shown in eqs. 3.4 and 3.5 where $A_1^{(1)}$ and $A_1^{(1)}$ are the initial estimates of the coefficient matrices.

$$\xi(k) = \begin{bmatrix} \xi_1(k) \\ \xi_2(k) \\ \xi_3(k) \\ \xi_n(k) \end{bmatrix} \quad \eta(k) = \begin{bmatrix} \eta_1(k) \\ \eta_2(k) \\ \eta_3(k) \\ \eta_n(k) \end{bmatrix} \quad \xi(k) + \begin{bmatrix} A_1^{(1)} \\ \eta(k-1) = \xi(k) \quad (3.4) \\

\eta(k) + \begin{bmatrix} A_1^{(1)} \\ \eta(k+1) = \eta(k). \quad (3.5) 

Let the one step-ahead prediction error covariance matrix be $V$, and the covariance matrix corresponding to $\eta$ be $V$ as in 3.6

$$V = E[\eta(k) \eta^T(k)] ; \quad \bar{V} = E[\xi(k) \xi^T(k)]. \quad (3.6)$$

Let the lag-sums of $Y(\cdot)$ be denoted by $\Gamma_j$ (eq. 3.7) and the covariance matrices by $R_j$ (eqs. 3.8). The $(n \times n)$ matrices $\Gamma_j$ and $R_j$ are not symmetrical but satisfy eq. 3.9.

$$\Gamma_j = \sum_{k=j+1}^{N} \xi(k) \xi^T(k-j) \quad (3.7)$$

$$R_j = E[\xi(k) \xi^T(k-j)] \quad (3.8)$$

$$\Gamma_j = \Gamma_j^T \quad R_j = R_j^T. \quad (3.9)$$
3.4

The algorithm (AL1) for computing the estimates of the coefficients of the p-th order AR scheme, provided the (p-1)th order AR coefficients available are given below.

\[
A_p^{(p-1)} = \Gamma_0 + A_1^{(p-1)} \Gamma_{-1} + A_2^{(p-1)} \Gamma_{-2} + \ldots + A_{p-1}^{(p-1)} \Gamma_{-p-1} (S_{p-1})^{-1}.
\]

\[
S_{p-1} = \Gamma_0 + A_1^{(p-1)} \Gamma_{-1} + A_2^{(p-1)} \Gamma_{-2} + \ldots + A_{p-1}^{(p-1)} \Gamma_{-p-1}.
\]

\[
\overline{A}_p^{(p)} = \Gamma_0 + A_1^{(p-1)} \overline{\Gamma}_{-1} + \ldots + A_{p-1}^{(p-1)} \overline{\Gamma}_{-p-1} (S_{p-1})^{-1}.
\]

\[
S_{p-1} = \Gamma_0 + A_1^{(p-1)} \overline{\Gamma}_{-1} + \ldots + A_{p-1}^{(p-1)} \overline{\Gamma}_{-p-1}.
\]

When the p-th order AR scheme is fitted to the data the (p-1) AR coefficients will change because of the additional associated coefficient vector \(A_p\). These forward and backward coefficients (\(A_p^{(p)}\) and \(\overline{A}_p^{(p)}\)) are corrected by using eqs. 3.10 and the residual matrices \(S_{p-1}\) and \(\overline{S}_{p-1}\) are updated by using eqs. 3.11.

\[
A_k^{(p)} = A_k^{(p-1)} + \frac{A_p^{(p-1)}}{\overline{A}_p^{(p-1)}} A_k^{(p-1)}
\]

\[
\overline{A}_k^{(p)} = \overline{A}_k^{(p-1)} + A_k^{(p)} A_k^{(p-1)}
\]

\[
S_p = \Gamma_0 + A_1^{(p)} \Gamma_{-1} + A_2^{(p)} \Gamma_{-2} + \ldots + A_p^{(p)} \Gamma_{-p}
\]

\[
\overline{S}_p = \Gamma_0 + A_1^{(p)} \Gamma_{1} + A_2^{(p-1)} \Gamma_{2} + \ldots + A_p^{(p)} \Gamma_{p}
\]
The unbiased estimate of the one step-ahead prediction "error" covariance matrix $V_p$, corresponding to the $p^{th}$ order AR scheme, is given by eq. 3.12.

$$V_p = \frac{1}{N - np} S_p. \quad (3.12)$$

The convergence and other properties of the algorithm All are found in Whittle (1963).

### 3.2.1 Significance of Additional Terms $A_j$

The significance of the $p^{th}$ order AR coefficients can be tested by using the multidimensional test of significance proposed by Anderson (1958). The test statistic $F_0$ is defined in eq. 3.13, which is asymptotically distributed as an $F$-distribution with $n$ and $N - np$ degrees of freedom.

$$F_0 = \left[ \frac{|S_{p-1}| - |S_p|}{|S_p|} \right] \left( \frac{N - np}{n} \right) \sim F[n, N - np] \quad (3.13)$$

For large values of $N$, $F_0$ may be considered to be approximately $\chi^2$ distributed with $n$ degrees of freedom. If the statistic $F_0$ is less than $F[n, N - np]$ (or less than $\chi^2(N)$ for large $N$) at a given significance level (such as 95%), then we can consider the $p^{th}$ term to be insignificant. We will refer to this test as the $F_0$-test.

Another goodness-of-fit test, proposed by Bartlett and Rajalakshman (1953) may also be used to test the significance of the $p^{th}$ order term. Although Bartlett and Rajalakshman (1953) have proposed the so-called G-test and H-test, we have used only the G-test in the present study. The details of the G-test are found in the paper by Bartlett and Rajalakshman.
3.2.2 Prediction Error Variance

The model 3.1 may also be used to obtain k steps-ahead prediction. The k steps-ahead prediction error covariance matrix can be recursively computed. Let \( y(k) \) be represented as a moving average process as in eq. 3.14,

\[
y(k) = \xi(k) = B_1 \xi(k-1) + B_2 \xi(k-2) + \ldots,
\]

where the coefficient matrices \( B_j \) are of dimension \((n \times n)\). The coefficient matrix \( B_k \) and the k steps-ahead prediction error covariance matrix can be estimated by eqs. 3.15 and 3.16.

\[
y(k+1) = y(k) + B_k \xi B_k^T.
\]

3.3 Multivariate Models for Indiana and Wisconsin Data

The multivariate model has fitted to the detrended (S1) and the normalized (S2, S3) data sequences. As precipitation of Neillsville has negligibly small influence on the water levels in L1, as indicated by the analysis of scalar models, it was not included in the multivariate model of the Wisconsin data. The precipitation sequences at Stevens Point, Marshfield and Wisconsin Rapids were, however, included. The precipitation at Lafayette, Wabash river stages, and the levels in L1 and L2 were used in the model for Indiana data.

The estimates of the \( A_1 \) and \( \bar{A}_1 \) matrices were computed by using the single equations of the first order AR model which constitute the
the multivariate first order AR system. The coefficients in these equations were estimated by using the algorithm AL discussed in Chapter II. These estimates and the corresponding residual sequences were used in the algorithm AL1 to obtain the coefficients of the higher order AR systems and to update the coefficients of the lower order systems as discussed in the previous section.

The test statistic $F_0$, defined in eq. 3.13 was also computed for each of the higher order systems ($p = 2, 3, \ldots$). The statistic $F_0$ was used to test the significance of the additional AR term. The test statistic $Z$ corresponding to the $G$-test and the corresponding $\chi^2$ value at 95% level were also obtained.

The matrix coefficients $A_p$ and the corresponding test statistics are shown in Table 3.1 for the models fitted to the detrended (S1) and standardized (S2, S3) data from Indiana and Wisconsin. The test statistic $F_0$ and the $Z$ statistic are given in Table 3.2, along with the critical $F(\cdot)$ and $\chi^2$ values at 95% level. These results are discussed below.

According to the results of the $F_0$-test, a second order AR model for the S1 and the S3 data series and a third order AR model for the S2 series appear to be adequate for the Indiana data. For the Wisconsin data, a fourth order model fits the detrended (S1) series whereas a second order model fits the S2 and S3 series, once again according to the $F_0$-test.

The results of the $G$-test indicate that the second order AR model adequately fits the S1, S2, and S3 series. For the data from Wisconsin, for the S1 and S2 series, second order AR models and for the S3 series
<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>-0.959</td>
<td>-0.147</td>
<td>0.233</td>
</tr>
<tr>
<td>W2</td>
<td>0.729</td>
<td>-0.206</td>
<td>-0.894</td>
</tr>
<tr>
<td>W3</td>
<td>0.239</td>
<td>-0.095</td>
<td>0.026</td>
</tr>
<tr>
<td>W4</td>
<td>0.153</td>
<td>0.225</td>
<td>0.162</td>
</tr>
<tr>
<td>W5</td>
<td>0.198</td>
<td>-0.187</td>
<td>0.181</td>
</tr>
<tr>
<td>W6</td>
<td>0.208</td>
<td>0.014</td>
<td>0.064</td>
</tr>
<tr>
<td>W7</td>
<td>0.161</td>
<td>-0.037</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Table 3.1** Matrix coefficients \( A_p \)
<table>
<thead>
<tr>
<th>$D_{(15)}$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>WISC(55)</td>
<td>$F_g$</td>
<td>23.771</td>
<td>15.177</td>
<td>11.262</td>
<td>10.981</td>
<td>6.470</td>
<td>5.781</td>
<td>4.680</td>
<td>1.442</td>
<td>2.113</td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NC</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WISC(55)</td>
<td>DEC</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>WISC(55)</td>
<td>$F_g$</td>
<td>4.540</td>
<td>2.440</td>
<td>2.240</td>
<td>2.220</td>
<td>2.200</td>
<td>2.180</td>
<td>2.160</td>
<td>2.140</td>
<td>2.120</td>
<td>2.100</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NC</td>
<td>NC</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WISC(55)</td>
<td>DEC</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>IND(51)</td>
<td>$F_g$</td>
<td>47.550</td>
<td>7.891</td>
<td>6.750</td>
<td>5.625</td>
<td>4.510</td>
<td>3.395</td>
<td>2.990</td>
<td>2.690</td>
<td>2.390</td>
<td>2.190</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND(51)</td>
<td>DEC</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND(51)</td>
<td>DEC</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>IND(51)</td>
<td>$F_g$</td>
<td>14.670</td>
<td>9.590</td>
<td>6.320</td>
<td>5.310</td>
<td>4.300</td>
<td>3.290</td>
<td>2.280</td>
<td>1.270</td>
<td>0.260</td>
<td>0.250</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IND(51)</td>
<td>DEC</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>$F_g$: Test statistic</td>
<td>DEC: Decision</td>
<td>INADEQ: Inadequate</td>
<td>ADEQ: Adequate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3.2  $F_g$ AND $G$ TEST STATISTICS AND FINAL MODELS
a third order model would fit adequately. Only for the S2 series of both Lafayette and Wisconsin data do the results obtained by $F_0$-test and G-test coincide. As the G-test is more powerful of the two tests, the results from the G-test are accepted. The final model selected for each of the series is shown at the bottom of Table 3.2.

Some of the statistical properties of the residuals $\xi_1(k)$ obtained from the models listed in Table 3.2 such as the covariance matrix of the residuals, and the statistics of the residuals are shown in Table 3.3. These residuals were further tested for whiteness and absence of periodicities and the results of these tests are discussed below.

3.4 Characteristics of Residuals from the Multivariate Models

The multivariate models can obviously be written as a set of $n$ univariate models, each of which corresponds to one of the variables. For instance the first univariate model in the multi-output model for Wisconsin data is the model for the water level fluctuations in W1. We will refer to the corresponding residual sequence $\xi_1(\cdot)$ as the residual sequence corresponding to W1. Similar designations are given for other residual sequences $\xi_2(\cdot)$, $\xi_3(\cdot)$ and $\xi_4(\cdot)$.

The residual variance of $\xi_1(\cdot)$ from the model fitted to the Wisconsin data is much smaller than the variance of residuals corresponding to other variates (precipitation sequences) irrespective of the type of transformation used. The ratio of residual mean square value to the mean square value of the observations is about 0.23 for the residuals $\xi_1(\cdot)$ and about 0.9 for other variates. These ratios indicate that the model fitted to the Wisconsin data explains about 77% of the
### Table 3.3 Characteristics of Residuals from Multivariate Models and Results from χ²-Test for Normality

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean Deviance</th>
<th>Mean Deviance Standard Error</th>
<th>Mean Deviance Chi-Square</th>
<th>Mean Deviance p-value</th>
<th>Mean Deviance Skewness</th>
<th>Mean Deviance Kurtosis</th>
<th>Mean Deviance Kurtosis Degrees of Freedom</th>
<th>Mean Deviance Kurtosis Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>12.34</td>
<td>0.23</td>
<td>13.45</td>
<td>0.001</td>
<td>1.23</td>
<td>3.45</td>
<td>12.34</td>
<td>0.23</td>
</tr>
<tr>
<td>Model 2</td>
<td>12.35</td>
<td>0.24</td>
<td>13.46</td>
<td>0.002</td>
<td>1.24</td>
<td>3.46</td>
<td>12.35</td>
<td>0.24</td>
</tr>
<tr>
<td>Model 3</td>
<td>12.36</td>
<td>0.25</td>
<td>13.47</td>
<td>0.003</td>
<td>1.25</td>
<td>3.47</td>
<td>12.36</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Note: Table 3.3 presents the characteristics of residuals from multivariate models along with the results from the χ²-test for normality. The table includes mean deviance, mean deviance standard error, mean deviance chi-square, mean deviance p-value, mean deviance skewness, mean deviance kurtosis, mean deviance kurtosis degrees of freedom, and mean deviance kurtosis standard error for each model.
variation in the ground water levels in well W1. The statistical properties of residuals obtained from models fitted to the S1, S2 and S3 data sequences are similar. The model fitted to the S1 sequence, however, gives the smallest value of the ratio of the residual variance to the variance of the observed sequences.

The characteristics of residual sequences $\varepsilon_i(\cdot), i = 1, 2, 3, 4$ obtained from the multi-output models for S1, S2 and S3 sequences of Indiana data are quite similar. The residuals have zero mean in all cases. However, for the model fitted to the S3 sequence the ratios of residual variance to variance of observations are 0.45, 0.025, 0.81 and 0.99 for the sequences L1, L2, Wabash stage and precipitation at Lafayette. Hence the model fitted to S3 accounts for about 50 and 98% of the variation in L1 and L2 respectively, which is greater than the S1 and S2 sequences. Consequently, the transformation S3 appears to give slightly better results for Indiana data.

The skewness and noncentrality parameters of residuals $\varepsilon_i(\cdot)$ obtained from multivariate models for S1, S2 and S3 sequences of Indiana data are summarized in Table 3.3. The skewness coefficient of residuals corresponding to W1 is large and negative for the model of the S1 sequence but very small and negative for the models of the S2 and S3 sequences. The residuals corresponding to other variates in Wisconsin data have a small positive value for the S3 sequence and large values for S1 and S2. The noncentrality parameter which is about 3 for all residuals indicates that the residuals are approximately normal.

The skewness coefficients of residuals from the models for Indiana data are relatively large and positive, for all the sequences S1, S2
and S3. However, the residuals from the S2 sequence have the lowest skewness coefficient. The residuals $\xi_1(\cdot)$ and $\xi_2(\cdot)$ from the model fitted to the S2 sequence have skewness coefficients of 0.66 and 0.20 respectively. The residuals have lower values of noncentrality parameter for this (S2) data.

The histograms of residuals from multivariate S1, S2, and S3 sequences for both Wisconsin and Indiana data are shown in Fig. 3.1. These histograms are single peaked, and symmetrical and appear to be approximately normally distributed. Further, the $\chi^2$-test for normality indicated that the residuals are normally distributed. Consequently, the residuals from these models can be considered to be normally distributed.

3.5 Whiteness Tests on Residuals - Autocorrelation Tests

Residuals from the multivariate models shown in Table 3.2 were tested for whiteness and presence of any periodicities. The assumptions made for parameter estimation require the residuals to be uncorrelated and free from any periodic components. The various whiteness tests and the cumulative periodogram tests for detecting the presence of periodicities are discussed in Sec. 2.6. These tests were conducted on individual residual sequences $\xi_i(\cdot)$ obtained from the multivariate models and the results are discussed below. The cross correlation properties of residual sequences are discussed in the next section.

The correlogram and the power spectral densities of residuals are shown in Fig. 3.2. The approximate 95% confidence limits ($\pm 2/\sqrt{N}$) are also shown on the correlograms in Fig. 3.2. All the correlation
FIGURE 3.1  HISTOGRAMS OF RESIDUALS FROM MULTIVARIATE MODELS
coefficients lie within these confidence limits and, therefore, the residual correlations can be considered to be not significantly different from zero. This is further demonstrated by the power spectral density plots in Fig. 3.2.

The results of and decisions from the F-test, the \( \chi^2 \)-test and the Portmanteau tests on each of the residual sequences obtained from the models fitted to Wisconsin and Indiana data are summarized in Table 3.4. The results from all these tests indicate that the residuals can be considered white.

The cumulative periodogram of residuals together with the 95\% confidence limits are shown in Fig. 3.3. Since the cumulative periodogram is within the confidence limits, the residuals are considered to be free from periodic components.

Therefore, the individual residual sequences from the multivariate models can be considered to be uncorrelated and free from periodicities. In the next section, several cross correlation tests are discussed. These tests are used to check whether the residual sequences are correlated with each other.

3.6 Cross Correlation Tests on Residual Sequences

Several cross-correlation tests were used for testing whether a residual sequence is correlated with other residual sequences. The covariance matrix and correlation matrix of residuals are shown in Table 3.3. The lag-zero and lag-one cross correlation coefficients shown in Table 3.3 are within the two standard error limits \( (2/\sqrt{N}) \) which approximately indicate that the residual sequences are uncorrelated
<table>
<thead>
<tr>
<th>Station</th>
<th>Type of Data</th>
<th>Leg 5</th>
<th>Leg 10</th>
<th>Leg 15</th>
<th>Leg 20</th>
<th>Leg 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>wi</td>
<td></td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
</tr>
<tr>
<td>51</td>
<td>1.039</td>
<td>0.306</td>
<td>0.577</td>
<td>0.297</td>
<td>0.627</td>
<td>0.616</td>
</tr>
<tr>
<td>52</td>
<td>1.111</td>
<td>0.464</td>
<td>0.765</td>
<td>0.479</td>
<td>0.787</td>
<td>0.723</td>
</tr>
<tr>
<td>53</td>
<td>1.060</td>
<td>0.529</td>
<td>0.884</td>
<td>0.626</td>
<td>0.716</td>
<td>0.714</td>
</tr>
<tr>
<td>sp</td>
<td></td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
</tr>
<tr>
<td>51</td>
<td>0.309</td>
<td>0.667</td>
<td>0.766</td>
<td>0.417</td>
<td>0.667</td>
<td>0.766</td>
</tr>
<tr>
<td>52</td>
<td>0.896</td>
<td>2.966</td>
<td>2.557</td>
<td>2.072</td>
<td>4.041</td>
<td>4.041</td>
</tr>
<tr>
<td>53</td>
<td>0.344</td>
<td>1.750</td>
<td>1.750</td>
<td>1.750</td>
<td>1.750</td>
<td>1.750</td>
</tr>
<tr>
<td>sp</td>
<td></td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
</tr>
<tr>
<td>51</td>
<td>0.534</td>
<td>2.706</td>
<td>2.706</td>
<td>0.519</td>
<td>6.886</td>
<td>6.886</td>
</tr>
<tr>
<td>52</td>
<td>0.634</td>
<td>1.230</td>
<td>1.230</td>
<td>1.230</td>
<td>1.230</td>
<td>1.230</td>
</tr>
<tr>
<td>53</td>
<td>0.414</td>
<td>2.092</td>
<td>2.092</td>
<td>0.599</td>
<td>4.461</td>
<td>4.461</td>
</tr>
<tr>
<td>sp</td>
<td></td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
<td>$P$</td>
<td>$x^2$</td>
</tr>
<tr>
<td>51</td>
<td>0.065</td>
<td>1.086</td>
<td>1.086</td>
<td>0.106</td>
<td>9.185</td>
<td>9.185</td>
</tr>
<tr>
<td>52</td>
<td>0.397</td>
<td>2.017</td>
<td>2.017</td>
<td>0.432</td>
<td>6.419</td>
<td>6.419</td>
</tr>
<tr>
<td>53</td>
<td>0.108</td>
<td>0.920</td>
<td>0.920</td>
<td>0.737</td>
<td>2.452</td>
<td>2.452</td>
</tr>
</tbody>
</table>

$F$: F-Test; $x^2$: Chi-square test; $P$: Portmanteau test

### TABLE 3.4 RESULTS OF WHITENESS TESTS ON RESIDUALS FROM MULTIVARIATE MODELS

3.16
FIGURE 3.2  CORRELOGRAMS AND POWER SPECTRAL DENSITIES OF RESIDUALS FROM MULTIVARIATE MODELS
with each other. The tests discussed below, for investigating the
cross correlation properties of the residual sequences, although not
very powerful, are useful indicators of lack of cross correlation
between residual sequences. The details of the tests may be found in

(a) **CROSS CORRELATION TEST**

The cross correlation coefficient \( d_k(\cdot) \) between two residual
sequences \( x_i(\cdot) \) and \( x_j(\cdot) \) is defined as follows for sequences \( x_i(\cdot) \) and \( x_j(\cdot) \).

Let

\[
r_k(x_i, x_j) = \left[ \frac{1}{N-j} \sum_{k'=k+1}^{N} x_i(k') \cdot x_j(k'-j) \right]
\]

(3.17)

\[
d_k(x_i, x_j) = \frac{r_k(x_i, x_j)}{r_0(x_i, x_i) \cdot r_0(x_j, x_j)}; \ i \neq j.
\]

(3.18)

The cross correlation coefficients \( d_k(x_i, x_j) \) between the sequences
\( x_i(\cdot) \), \( x_j(\cdot) \) are shown in Fig. 3.4 for a few pairs of residual
sequences. In Fig. 3.4, the approximate two standard error limit
\( \pm 2/\sqrt{N} \) is also indicated. All the cross correlation coefficients lie
within this error limit which indicates that the residual sequences
\( x_i(\cdot) \) and \( x_j(\cdot) \) are uncorrelated. The results are similar for all the
pairs of residual sequences.

(b) **THE COHERENCE TEST**

The coherence square \( (CO_{xy}) \) at various frequencies between any two
residual sequences can be computed by following the procedure outlined
in Sec. 1.3.4. The coherence square is analogous to the square of the
FIGURE 3.4 CROSS CORRELATION COEFFICIENTS OF RESIDUALS FROM MULTIVARIATE MODELS
correlation coefficient between two samples. Hence, at any frequency, the larger the value of the coherence square, the more closely correlated are the two sequences. The plot of \( C_{xy}(\omega_h) \) (eq. 1.15) against the frequency \( \omega_h \) is called the coherence diagram.

The variation of coherence square with frequency of the residual sequences obtained from the multi-output model for Wisconsin data are shown in Fig. 3.5. The coherence square for all the pairs of residual sequences appear to be approximately uniformly distributed over the entire frequency range. The magnitude of coherence square over any frequency rarely exceeds 0.3, thereby indicating that the components of residual sequences are uncorrelated. Similar results were obtained for all the pairs of residual sequences for Indiana data also.

(c) \textbf{INTEGRATED PHASE SPECTRUM TEST}

The integrated phase spectrum test is similar to the cumulative periodogram test for single residual sequences which was discussed earlier in Sec. 2.6.4. If the residual sequences \( \xi_i(k) \) and \( \xi_j(k) \) are uncorrelated, then the integrated phase spectrum \( F_{\xi_i \xi_j}(f_k) \) (where \( f_k \) is the frequency ranging from 0.0 to 0.5) of the two series should be uniformly distributed in the range 0.0 to 0.5. Hence, when the cumulative phase spectrum is plotted against frequency, the plot of integrated phase spectrum values \( F_{\xi_i \xi_j}(f_k) \) versus \( f_k \) should be tightly clustered around the straight line from (0,0) to (0.5, 1). We can also construct the 95% confidence limits (± 1.37/\( \sqrt{N/2} \)) around this line to test the lack of cross correlation at 5% significance levels.
FIGURE 3.5  COHERENCE DIAGRAM OF RESIDUALS FROM MULTIVARIATE MODELS
The cumulative phase spectrum of some of the sequences of residuals $\xi_i(\cdot)$ and $\xi_j(\cdot)$ obtained from multivariate models are shown in Fig. 3.6 along with the 95% confidence band. The cumulative phase spectrum of $\xi_i(\cdot)$ and $\xi_j(\cdot)$ lies within the 95% confidence band, thereby indicating the absence of significant sinusoidal components in the cross correlations between the residuals $\xi_i(\cdot)$ and $\xi_j(\cdot)$. The results are similar for other pairs of residuals.

(a) **INTEGRATED SAMPLE CO-SPECTRUM TEST**

Let $L_{\xi_i \xi_j}(f_k)$ be the co-spectrum computed from the residual series $\xi_i(\cdot)$ and $\xi_j(\cdot)$. The normalized estimator of the integrated co-spectrum at frequency $f_k$ is defined as $\hat{J}_{\xi_i \xi_j}(f_k)$, and is given below.

$$
\hat{J}_{\xi_i \xi_j}(f_k) = \left[ \frac{2}{N S_{\xi_i} S_{\xi_j}} \right] \sum_{k'=0}^{k} L_{\xi_i \xi_j}(f_{k'}) ; \quad 0 \leq f_k \leq 0.5
$$

$S_{\xi_i}$ and $S_{\xi_j}$ are the standard deviations for the residual sequences $\xi_i(\cdot)$ and $\xi_j(\cdot)$. If the residual series are uncorrelated then the values of $J_{\xi_i \xi_j}(f_k)$ should oscillate around zero for all $f_k$.

The normalized integrated co-spectrum between several residual sequences $\xi_i$ and $\xi_j$ are shown in Fig. 3.7. The values of $J_{\xi_i \xi_j}(f_k)$ are very small (of the order of $10^{-4}$) for all frequencies indicating the lack of cross correlation between $\xi_i(\cdot)$ and $\xi_j(\cdot)$. Similar results were obtained for other pairs of residuals.
FIGURE 3.6 INTEGRATED PHASE SPECTRA OF RESIDUALS FROM MULTIVARIATE MODELS
FIGURE 3.7 INTEGRATED CO-SPECTRA OF RESIDUALS MULTIVARIATE MODELS
IV. SIMULATION OF THE MULTIVARIATE MODEL

4.1 Introduction

The validity of the multivariate models which were developed by using the methods discussed in the previous chapter was further tested by simulation. As discussed in Chapter III, the residuals from the multivariate models were found to be uncorrelated and without periodicities. The histograms of these residuals may be adequately approximated by normal distributions, by using the means and variances of the residual sequences given in table 3.3. The multivariate models for the transformed sequences were simulated by sampling from these normal distributions. The characteristics of these simulated sequences were analyzed to determine the capability of the model to preserve the characteristics of the observed data. This is the first objective of the simulation study. The second objective of the simulation study is to investigate the relative merits of the transformations used.

Matalas (1967) has proposed a multivariate AR model for simulation in hydrology. This model has received considerable attention. Consequently, it was decided to use the model proposed by Matalas to develop a model for the present data. The validity of the Matalas' model to represent the present data may also be determined. This analysis was conducted to determine whether the simpler Matalas' model was adequate to represent the data of the simulation study discussed in this chapter.
4.2

The following statistical characteristics of the observed data sequences are compared with those of the simulated data.

(i) Histograms
(ii) Mean and variance
(iii) The skewness coefficient
(iv) Lag-one correlation coefficient
(v) Lag-zero and lag-one correlation matrices
(vi) Correlograms
(vii) Power spectral densities.

The properties mentioned above can be considered to adequately characterize the observed data and these properties should be preserved by the simulated (or generated) data. In the following discussion we will compare the statistical characteristics i-vii mentioned above of the generated S1, S2, and S3 sequences with the corresponding characteristics of the S1, S2, and S3 sequences obtained from observed data. Also, the sequences S1, S2 and S3 generated by the multivariate model are transformed back by using eqs. 1.1 - 1.3. The statistical characteristics of the back transformed sequences are compared with those of the observed sequences.

The results discussed in the following sections were typical of the simulation results. As giving all the results would have been voluminous, especially the number of figures would have been very large, only some of the results are discussed below.
4.2 Comparison of Statistical Properties of Observed and Generated Data

4.2.1 Histograms

The histograms of generated and back transformed data obtained from the models for S2 and S3 sequences are shown in fig. 4.1. The histograms of the observed data are also shown in fig. 4.1. The match between the histograms of the observed and generated data is acceptable. The other histograms, which are not shown very similar to those shown in fig. 4.1 for both the Indiana and Wisconsin data and for other variates.

4.2.2 The Mean and Variance

The mean and variance of the generated S1, S2 and S3 sequences and the corresponding observed values are summarized in Table 4.1 for both Indiana and Wisconsin Data. The mean values of the generated S1, S2 and S3 sequences are very close to the corresponding observed values for the Wisconsin data. The means of the generated S2 sequences give a better match with the corresponding observed mean values for Indiana data. The means of the generated S1 and S3 sequences are also close to the corresponding observed values.

The variances of the generated and observed data are also tabulated in table 4.1 for the Wisconsin data. The variance of the generated S1 sequence appears to preserve the variance of all the variates in the multivariate model whereas the variances of the variates of the generated S2 and S3 sequences are all different from the corresponding observed values.
FIGURE 4.1  HISTOGRAMS OF OBSERVED AND BACK TRANSFORMED GENERATED DATA
### TABLE 4.1
STATISTICAL CHARACTERISTICS OF THE OBSERVED AND BACK TRANSFORMED GENERATED DATA

<table>
<thead>
<tr>
<th>LAG</th>
<th>Variance</th>
<th>Skewness</th>
<th>Curr. Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 4.2
LAG-ZERO AND LAG-ONE CROSS CORRELATION MATRICES OF TRANSFORMED OBSERVED AND GENERATED DATA

<table>
<thead>
<tr>
<th>LAGZERO/mat</th>
<th>LAGONE/mat</th>
<th>LAGZERO/mat</th>
<th>LAGONE/mat</th>
<th>LAGZERO/mat</th>
<th>LAGONE/mat</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6

For the Indiana data, the variances computed from the models of the S2 sequences are close to the observed variances, whereas those computed from the model S1 are not so close. The variances computed from the model of the S3 sequence are considerably different from the corresponding observed values.

4.2.3 Skewness Coefficients

The skewness coefficients of the generated and observed data are tabulated in table 4.1. For Wisconsin data the S3-model preserves the skewness of the rainfall variates better than the models of S1 and S2 sequences. However, the skewness coefficient of the water levels in the well W1 is not so well preserved as it is about 0.915 for the observed data and 0.644 for the data generated by the S3 model. The skewness coefficients of the generated S1 and S2 sequences are not acceptable, as they are considerably different from the corresponding observed values.

For the Indiana data, the skewness coefficients computed from the data generated from the S2 model are closer to the skewness coefficients of the observed data. The skewness coefficients of the data generated from the models of the S1 and S3 sequences are not close to the skewness coefficients of the observed data.

4.2.4 Lag One Correlation Coefficients

The lag-one correlation coefficients of the observed and backtransformed generated data are summarized in table 4.1. For Wisconsin data, the lag-one correlation coefficient of the water levels in W1 is preserved in all models. The model for the S2 sequence appears to preserve
the lag-one correlation coefficient of all other variates better than the models for the other sequences. Thus transformation S2 can again be considered good in preserving the lag one correlation coefficients of the Wisconsin data.

For Indiana data the lag-one correlation coefficients computed from the generated data from the S1 model are closer to the lag-one correlation coefficients of the observed data. The data generated from the models of the S2 and S3 sequences do not preserve the lag-one correlation coefficient.

4.2.5 Lag-zero and Lag-one Cross Correlation Matrices

The lag-zero and lag-one cross correlation matrices are given in table 4.2. The S1 model for Wisconsin data preserves the lag-zero and lag-one cross correlations better than the values generated by S2 and S3 models. The cross correlation coefficients computed by the data generated by S2 and S3 models are similar.

For Indiana data, the cross correlation coefficients computed by the data from S2 and S3 models are better than those from the S1 model in preserving lag-zero correlation coefficients. Once again the lag-zero cross correlation coefficients computed from the data generated by the models for S2 and S3 sequences are close to each other. The lag-one correlation matrix is not well preserved by the data generated from any of the three models.

4.2.6 Correlogram

The correlogram of the generated and the observed sequences were also compared. The results are given for some of the generated and
back transformed sequences in fig. 4.2. The observed and computed 
correlograms were very close to each other for the precipitation 
variates in Wisconsin data irrespective of the model. The correspondence 
between the correlograms of the observed and generated (back transformed) 
data is better for the S2 and S3 sequences than for the S1 sequence.

Similarly for Lafayette data, transformation S2 gave a better fit 
of the correlograms.

The correlograms of the data generated from the models S2, S2 and 
S3 (which were not back transformed) and the corresponding transformed 
observed data sequences are shown in fig. 4.3 for some cases. It can 
be seen that the correlograms of the data generated from the models 
of the S1, S2 and S3 sequences match the corresponding correlograms 
of the observed sequences, with the exception of the correlogram of 
the sequences for water levels in W1, generated by using the S1 model. 
Similar results were obtained for the models of Indiana data also.

4.2.7 Power Spectral Density

The power spectral densities of the transformed observed and 
generated data are shown in fig. 4.4, for Wisconsin data. The power 
spectral densities of the generated data (which were not back trans- 
formed) are shown in fig. 4.5 along with the correlograms of the 
observed S1, S2 and S3 sequences. The results from the comparison to 
the power spectral densities are, of course, identical to those obtained 
from correlograms.
FIGURE 4.3  CORRELOGRAMS OF TRANSFORMED OBSERVED AND GENERATED DATA
4.2.8 Summary

The results of the simulation study indicate that the data simulated by the models of the S1 and S2 sequences preserve the observed statistical characteristics of the data. The results obtained by the model for the S3 sequence are not as good as those from the other two models. There is some discrepancy in the results from the models of the S1 and S2 sequences. Whether these can be attributed to the transformations is a matter which needs further investigation.

4.3 The Multivariate AR(1) Model

As mentioned earlier, the model proposed by Matalas (1967) was also used to model the ground water level, precipitation and river stage series. The Matalas model is given in eq. 4.1,

\[
y(k) = A \tilde{y}(k-1) + B \tilde{\xi}(k)
\]  

(4.1)

where, \( \tilde{y}(k) = (n \times 1) \) vector of outputs

\( \tilde{\xi}(k) = (n \times 1) \) vector of noise sequence

\( A = (n \times n) \) matrix of coefficients

\( B = (n \times n) \) matrix of coefficients.

In eq. 4.1, \( \tilde{\xi}(k) \) is Gaussian with zero mean and unit variance. Further, in estimating the parameters of eq. 4.1, in order to preserve the skewness coefficient, Matalas recommends log transforming the data with a lower bound.

The coefficient matrices \( A \) and \( B \) are estimated as follows. By post multiplying eq. 4.1 by \( \tilde{y}^T(k-1) \), taking expectation of the resulting equation and assuming that \( E[\tilde{\xi}(k) \tilde{y}^T(k-1)] = 0 \), eq. 4.2 is obtained,
\[ M(1) = AM(0) \]  \hspace{1cm} (4.2)

where,
\[ M(1) = E[y(k) y^T(k-1)] \]
\[ M(0) = E[y(k) y^T(k)]. \]

Similarly by post multiplying eq. 4.1 by \( y^T(k) \) and taking expectation of the resulting equation one can derive eq. 4.3.

\[ BM^T = M(0) - M(1) M(0)^{-1} M(1) \]  \hspace{1cm} (4.3)

The coefficients in the matrix \( A \) are directly obtained by eq. 4.2. In order to obtain the coefficients in the matrix \( B \), eq. 4.3 must be solved. Matalas (1967) suggested a method based on principal component analysis, and Young and Pisano (1968) gave a recursive scheme. In order to arrive at their solution the matrix \( B \) was assumed to be lower triangular by Young and Pisano. The method proposed by Young and Pisano was used in the present investigation.

4.3.1 Results

The parameter estimation scheme discussed above was applied to the data transformed by using different transformations including the log transformation. The parameters \( A \) and \( B \) of eq. 4.1 were estimated for each case and the resulting models were used to generate synthetic data. The estimates \( A \) and \( B \) obtained for the S1 series are shown in table 4.3. The statistics of the generated data were compared with those of the original data.

The statistics of the data generated from the model of the transformed series S1 were closer to the corresponding statistics of the observed data than those obtained from other models. There are
Table 4.3 Estimates A and B Matrices (S1-Series) of the Matalas Model.

\[
A = \begin{bmatrix}
0.854 & -0.055 & 0.025 & 0.029 \\
0.006 & 0.085 & 0.346 & 0.125 \\
-0.178 & 0.099 & 0.175 & 0.256 \\
-0.178 & 0.233 & 0.213 & 0.087 \\
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
0.051 & 0.000 & 0.000 & 0.000 \\
-0.196 & -0.244 & 0.000 & 0.000 \\
-0.250 & 0.219 & -0.373 & 0.000 \\
-0.301 & -0.141 & 0.251 & -0.771 \\
\end{bmatrix}
\]
summarized in table 4.4. From the results shown in table 4.4 the mean, variance and lag-one correlation coefficients of the generated data (from the model of the S1 sequence) are very close to the corresponding observed values. Also, the lag-zero and lag-one cross correlations are preserved fairly well. The skewness coefficient of the generated data is slightly different from that of the observed data.

Although some of the statistics of the process are well preserved by the Matalas model, its adequacy should also be tested. The G-test was used to determine the adequacy of the model and the results obtained from the test are summarized in table 4.5 for Wisconsin data. The results of the G-test for the models of the transformed series S1, S2 and S3 shown in table 4.5 indicate that the first order AR model is inadequate. However, the model for the S2 sequence is better than the models for the S1 and S3 sequences. It may be recalled that similar results of the G-test were obtained in Chapter III also, where the parameters of the first order AR model were estimated by separately considering the scalar equations.

In conclusion, although the Matalas model preserves some of the statistics of the process it cannot be accepted as the valid model of the process.
<table>
<thead>
<tr>
<th></th>
<th>MEAN</th>
<th>VAR</th>
<th>SKEW</th>
<th>CORR</th>
<th>CORR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>STD OBS</td>
<td>MAT OBS</td>
<td>STD OBS</td>
<td>MAT OBS</td>
<td>0.466</td>
</tr>
<tr>
<td>W1</td>
<td>7.912</td>
<td>0.120</td>
<td>5.670</td>
<td>5.724</td>
<td>0.945</td>
</tr>
<tr>
<td>WR</td>
<td>2.488</td>
<td>2.391</td>
<td>3.270</td>
<td>3.362</td>
<td>1.322</td>
</tr>
<tr>
<td>SP</td>
<td>2.500</td>
<td>2.418</td>
<td>5.840</td>
<td>3.633</td>
<td>1.329</td>
</tr>
<tr>
<td>MF</td>
<td>2.482</td>
<td>2.035</td>
<td>5.070</td>
<td>4.543</td>
<td>0.748</td>
</tr>
</tbody>
</table>

**Table 4.4** COMPARISON OF STATISTICAL CHARACTERISTICS OF BACK TRANSFORMED GENERATED (BY MATALAS MODEL) AND OBSERVED DATA

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>0.207</td>
<td>0.207</td>
<td>0.152</td>
<td>0.216</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR</td>
<td>0.346</td>
<td>0.256</td>
<td>0.319</td>
<td>0.215</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>0.191</td>
<td>0.252</td>
<td>0.534</td>
<td>0.300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>1.000</td>
<td>0.378</td>
<td>0.378</td>
<td>0.319</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W1</td>
<td>0.448</td>
<td>0.446</td>
<td>0.448</td>
<td>0.263</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WR</td>
<td>0.480</td>
<td>0.483</td>
<td>0.535</td>
<td>0.483</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>0.674</td>
<td>0.689</td>
<td>0.674</td>
<td>0.294</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF</td>
<td>1.000</td>
<td>0.689</td>
<td>0.674</td>
<td>0.294</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.18

Table 4.5
G-test on residuals from Matalas model

<table>
<thead>
<tr>
<th>Wisdomin Data</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Statistic</td>
<td>371.2</td>
<td>339.7</td>
<td>347.2</td>
</tr>
<tr>
<td>Critical Value</td>
<td>320.0</td>
<td>320.0</td>
<td>320.0</td>
</tr>
<tr>
<td>Decision</td>
<td>Inadequate</td>
<td>Inadequate</td>
<td>Inadequate</td>
</tr>
</tbody>
</table>
CHAPTER V
DISCUSSION AND CONCLUSIONS

The single output models investigated in the present study can be used to determine the relative effects of the external inputs on the water levels at a particular location in the aquifer. For example, the water level fluctuations in the well W1 is affected by the precipitation at different locations such as Wisconsin Rapids, Marshfield, etc. to varying degrees. Similarly the water levels in L1 are affected by the water levels in L2 but not the other way around. Similar conclusions can be drawn from the multivariate models also.

Conclusions such as those mentioned above are useful in defining inputs to digital deterministic models of aquifer systems. Apart from the spatial variability inherent in the precipitation process at the ground level, which must be included in the inputs to the deterministic ground water models, the rate of infiltration of precipitation into the ground water storage also varies spatially.

The multiple input-single output models such as those discussed in Chapter II are useful in estimation of the effects of precipitation on the ground water levels in a small region. They are also useful in estimating the interaction between the ground water levels in two neighboring wells and the interaction between the water levels in a stream and water levels in the wells in the adjacent areas.
5.2

The feasibility of development of multivariate models for the ground water levels and other related variables has also been demonstrated. In view of the nature of the variables involved, simple models such as the multivariate AR(1) model may not be adequate to represent the data. It is noteworthy in this regard that the validation of multivariate models by testing the characteristics of the residuals has not received much attention in hydrology. Unless the residuals from the models are found to be uncorrelated and without periodicities, the models cannot be accepted.

The transformation S2 appears to be the best in the sense that the synthetic data generated from the model for the S2 sequence appears to preserve the statistical characteristics of the observed data better than the other two transformations. However, some of the individual characteristics are better preserved by transformations S1 and S2.

In conclusion, the ground water levels in an aquifer and the related processes can be modeled by stochastic models. These models can be used for simulation and forecasting water levels. The computational effort involved in the development of these models is quite small. These models are of substantial use in investigating the aquifer characteristics.
CHAPTER VI

REFERENCES


CHAPTER VII
COMPUTER PROGRAMS

Two of the programs developed as a part of the study are listed in the following pages. The first program "Multivariate Model" is designed to estimate the parameter matrices $A_p$ by the Whittle's Algorithm. The $F_0$- and G-test statistics, and statistics of the residuals are also computed in the program.

The second program is based on Young and Pisano's method of parameter estimation in Matalas Model. The program can also be used for generating synthetic data. The statistics of observed and generated data and of residuals are also computed.