



This is to certify that the

thesis entitled

COMPARISON OF FIRST-ORDER ERROR ANALYSIS WITH
THE MONTE CARLO METHOD FOR A LAKE ONTARIO TIME-
DEPENDENT PHOSPHORUS CONCENTRATION PREDICTION MODEL

presented by

Ralph Emerson Ancil

has been accepted towards fulfillment
of the requirements for

M.S. degree in Resource Development

A handwritten signature in cursive script that reads "Carl W. Ramm". The signature is written in black ink and is positioned above a horizontal line.

Major professor

Date 10. Nov. 81.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 00837 3973

L



RETURNING MATERIALS:
Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

274

COMPARISON OF FIRST-ORDER ERROR ANALYSIS WITH THE
MONTE CARLO METHOD FOR A LAKE ONTARIO TIME-
DEPENDENT PHOSPHORUS CONCENTRATION
PREDICTION MODEL

By

Ralph Emerson Ancil

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Resource Development

1981

ABSTRACT

COMPARISON OF FIRST-ORDER ERROR ANALYSIS WITH THE
MONTE CARLO METHOD FOR A LAKE ONTARIO TIME-
DEPENDENT PHOSPHORUS CONCENTRATION
PREDICTION MODEL

By

Ralph Emerson Ancil

9/17/80

The problem of quantifying model prediction error, or uncertainty, is the subject of much ongoing research. This study compares two methods of prediction error quantification for a Lake Ontario eutrophication model, namely, the Monte Carlo method and first-order analysis. The Monte Carlo method accounts for uncertainty by characterizing each known error source by a frequency distribution. Values are obtained randomly from the distribution for each term which simulates the inherent variability. First-order analysis defines uncertainty by its first non-zero moment, the variance. Uncertainties are propagated through the model by using the first-order terms of the Taylor series expansion. The variances are then combined to yield the total prediction error. The comparison includes using the same parameter values; testing sensitivities; and running the Monte Carlo simulation with all-normal stochastic inputs. The results show that for similar input values, similar predictions but different errors are obtained.

ACKNOWLEDGMENTS

This research was made possible in part by funds from the National Oceanic and Atmospheric Administration, grant #03-78-b01-109.

I would like to acknowledge the fine work done by Barb Visser, Shelley Saub, and especially Sue Watt, for typing the final draft of this thesis, and to Paul Schneider for his excellent graphics.

Thanks are also extended to V. David Lee, my friend and colleague, for providing me with the information of his research which is used in this study. And thanks are also extended to Dr. Daniel Chappelle and Dr. Carl Ramm, two of my committee members, for their helpful comments and criticisms.

I especially wish to acknowledge Dr. Kenneth H. Reckhow for his guidance and patience during the course of my studies. My association with him has been an invaluable contribution to my education.

Finally, the indispensable assistance, patience, and encouragement of my dearest friend, my wife Clarissa, is acknowledged with deepest gratitude. It is to her that this thesis is dedicated.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF SYMBOLS	vii
INTRODUCTION	1
Chapter	
I. THE LAKE PHOSPHORUS MODEL	4
II. THE MONTE CARLO METHOD AND ITS APPLICATION TO THE LAKE ONTARIO MODEL	9
The Monte Carlo Method	9
Definition	9
Distribution Selection	10
Application to the Model	12
Loadings	13
Settling Velocity	15
Areal Water Loading	15
Hydraulic Retention Time	16
Other Parameters	16
III. RESULTS OF THE MONTE CARLO SIMULATION	17
Major Observations	17
Base Run	17
All-Normal Run	18
Minor Observations	22
Base Run Fluxes	22
Base Run Water Load	23
Higher Iterations Per Year	23
Flux Sensitivities	24
Land Use Fraction Sensitivities	25
Summary	25

Chapter	Page
IV. FIRST-ORDER ERROR ANALYSIS	27
V. RESULTS FOR FIRST-ORDER ANALYSIS	36
The Model Prediction	36
The Error Equation	41
Major Observations	41
Minor Observations	48
Summary	51
VI. COMPARISON OF FIRST-ORDER WITH MONTE CARLO	53
The Numerical Results	53
The Procedures	57
Theoretical Considerations	57
Practical Considerations	62
Summary	63
FINAL COMMENTS	65
APPENDICES	
A. MONTE CARLO APPENDICES	67
B. FIRST-ORDER ERROR APPENDICES	88
BIBLIOGRAPHY	114

LIST OF TABLES

Table	Page
3.1 Statistical Summary of Major Monte Carlo Runs . . .	19
5.1 Base Run	37
5.2 Modified Base Run	44
6.1 Comparison of Numerical Results	54
6.2 Comparison of Values	56
A.1 Statistical Summary of 100 Iteration Base Run--Model Terms	70
A.2 Statistical Summary of Base Run Fluxes	71
A.3 Statistical Summary of Base Run Water Load	72
A.4 Statistical Summary of Higher Iterations	73
A.5 Statistical Summary of Flux Sensitivities--Phosphorus Concentration Predictions	76
A.6 Statistical Summary of Land Use Shifts	80

LIST OF FIGURES

Figure	Page
3.1 40 Years of Phosphorus Concentration Predictions at 100 Iterations Per Year	20
3.2 40 Years of Phosphorus Concentration Predictions Using All Normal Distributions	21
5.1 Theoretical Behavior of Predictions and Errors . . .	38
5.2 Base Run Predictions	38
5.3 Parameter Error	42
5.4 Input Variable Error	42
5.5 Base Run--Total Error	43
A.1 40 Years of Phosphorus Concentration Predictions at 500 Iterations Per Year	74
A.2 Concentration Predictions Using Larger Random Seed for 1000 Iterations Per Year	75
A.3 40 Years of Phosphorus Concentration Predictions Using 1/2 Variability for Point Source Flux	77
A.4 40 Years of Phosphorus Concentration Predictions with all Phosphorus Removal for Point Flux	78
A.5 40 Years of Phosphorus Concentration Predictions Using Eastern Erie Concentration Data	79
A.6 40 Years of Phosphorus Concentration Predictions: Land Use Shifts to All One Type	81

LIST OF SYMBOLS

\bar{x}	= Inter-year prediction mean
\bar{s}	= Inter-year standard deviation
$CV_{\bar{x}}$	= Inter-year coefficient of variation
$R_{\bar{x}}$	= Inter-year mode to mean ratio
$\bar{C}V$	= Intra-year coefficient of variation
SS	= Steady-state model
k	= Decay factor
P_0	= Initial phosphorus concentration
P_{i-1}	= Previous year's phosphorus concentration
L	= Mean loading
q_s	= Mean surface overflow
v_s	= Mean apparent setting velocity
z	= Mean depth
τ	= Mean hydraulic retention
SSE	= Steady-state error
FSSE	= Fractional steady-state error
ME	= Model error
D_i	= i the diminishing term
F	= FSSE + ME
E_i	= ith error (as variance)
S_t^2	= Variance of prediction
$r(s_i, x_j)$	= Correlation coefficient of x_i with x_j .
$S^2(x_i)$	= Variance of x_i

INTRODUCTION

With the increasing burden on the earth's water resources, the resource manager whether in the private or government sector, is called upon to balance competing interests, evaluate different sources of information and to anticipate the consequences or impact of proposed uses and activities on the resource base. To do this, the manager requires some knowledge of the natural sciences to understand the biological, physical and chemical processes that shape the environment and to comprehend its fragile and interdependent character. Yet the water resource manager must also be familiar with the social sciences since it is in this sphere that decisions concerning the application of knowledge gleaned from the natural sciences are made. Questions of water allocation, efficiency and distribution are economic and political. Thus the planner stands at the interface of the natural and social disciplines. Furthermore, these two realms of thought may suggest conflicting answers to water development or management problems. What may be "good" from a strictly scientific viewpoint may not be "good" economically or politically. Also, uncertainty may arise regarding projects or activities, the nature and effect of which may be subtle and long-term. It is not surprising then that to make sound decisions, the water resource manager requires detailed, accurate information, as well as the

ability to anticipate the possible consequences of various courses of action, that is, to forecast or predict impacts.

There are several different ways prediction can be accomplished. One common way is a qualitative study of the problem resulting in a projection of possible scenarios. Another is simply to extrapolate from past experience if the present problem is analogous to previous ones. The conceptual model offers another approach of envisioning complex interactions among a system's components. This latter method is often closely followed by a quantitative model which expresses the component mathematically and yields a quantitative prediction. The complexity of these models ranges from simple input-output models (Vollenweider, 1969, 1975) to sophisticated treatments of biological systems (Thomann et al., 1977).

However, since the modeling process involves a simplification of reality, the predictions necessarily reflect a level of uncertainty or error. Any reduction in this uncertainty will likely increase the cost of the modeling process in both money and time. For example, increasing the complexity of the model may reduce its prediction uncertainty but will also increase the data requirements. The costs of data acquisition, model calibration and validation, and computation will increase. The modeler must, therefore, determine what level of uncertainty is acceptable under the circumstances and consistent with the available resources of time and money. Furthermore, the modeler's personal and professional reputation may be risked on the reliability of the results. Clearly, then, to the water resources manager who uses quantitative models, the quality of

the results is of great importance. This concern has recently led some researchers to investigate ways of quantifying model uncertainty and thereby provide this needed information (Reckhow, 1979; Chapra and Reckhow, 1979).

The purpose of this paper is to compare two such methods for prediction error quantification and to evaluate their relative advantages and disadvantages for a time-dependent phosphorus concentration prediction model of Lake Ontario.

The thesis proceeds in the following order. The reader is acquainted first with the model and its derivation. This is followed by a discussion of the Monte Carlo procedure and the model predictions of one simulation are analyzed. Then the alternative method of first-order error analysis is considered along with a presentation of its model predictions. Finally, the results of these two different approaches are compared and their relative strengths and weaknesses are discussed. The appendices are referenced in the text to supplement the information of the main text or to provide mathematical proofs, only the results of which are used in the texts.

CHAPTER I

THE LAKE PHOSPHORUS MODEL

The lake phosphorus model used in this paper is a "black box" model which means it treats the physical phenomena empirically, rather than theoretically. In other words, no attempt is made to individually account for all the processes that may be involved in producing the final product, instead, all the processes are aggregated into a few parameters which may or may not have a physical basis. To the water resources planner, the empirical approach has the advantage of relative mathematical simplicity over the more sophisticated theoretical models and is, therefore, more compatible with available financial resources, as well as the planner's mathematical ability. A survey of this kind of lake phosphorus model has been made by Uttormark and Hutchins (1980) and by Reckhow (1979).

The underlying assumption of this model is that given a lake's size, mean depth, phosphorus loading, and flushing rate, phosphorus concentrations can be predicted. The model was derived by first calculating a phosphorus mass balance for lakes. In this calculation it is further assumed that phosphorus losses occur through the outlet and through internal losses such as sedimentation and that the net internal loss is directly proportional to the amount of phosphorus in

the lake. A phosphorus mass balance for a simple lake system is given by:

$$V \frac{dP}{dt} = M - \sigma PV - QP \quad 1.1$$

where,

P = lake phosphorus concentration (mg/l)

V = lake volume (10^6m^3)

M = annual mass rate of phosphorus inflow to lake (10^3kg/yr)

Q = annual volume rate of water outflow from lake ($10^6 \text{m}^3/\text{yr}$)

σ = sedimentation coefficient (per yr).

Another expression for the phosphorus mass balance is given by:

$$V \frac{dP}{dt} = M - v_s PA - QP \quad 1.2$$

where,

A = lake surface (bottom) area (km^2)

v_s = apparent setting velocity (m/yr).

Equation 1.2 differs from Equation 1.1 in that the sediment sink term (σPV) is treated as an areal sink term ($v_s PA$) and hence the rate of phosphorus deposition to the sediments is a function of the bottom (surface) area.

Equation 1.1 has the steady-state solution ($\frac{dP}{dt} = 0$):

$$P = \frac{L}{z\left(\frac{1}{\tau} + \sigma\right)} \quad 1.3$$

where,

L = annual areal phosphorus loading ($\text{g}/\text{m}^2\text{-yr}$)

z = lake mean depth (m)

σ = hydraulic retention time (yr)

τ = hydraulic retention time (yrs).

The time-dependent solution is:

$$P_i = \frac{L}{z + \frac{z}{\tau}} \left(1 - e^{-\left(\frac{1}{\tau} + \sigma\right)\Delta t}\right) + P_{i-1} e^{-\left(\frac{1}{\tau} + \sigma\right)\Delta t} \quad 1.4$$

where,

P_i = ith phosphorus concentration prediction (mg/l)

P_{i-1} = previous phosphorus concentration prediction (mg/l)

Δt = time interval (yrs).

Let $z = q_s$ = surface overflow rate (m/yr) and Equation 1.3 may be restated as:

$$P = \frac{L}{q_s + z} \quad 1.5$$

Equation 1.2 had the steady-state solution:

$$P = \frac{L}{v_s + \frac{z}{\tau}} = \frac{L}{v_s + q_s} \quad 1.6$$

Note that the difference between these two expressions (Equations 1.5 and 1.6) is in the term for the settling velocity. In Equation 1.5 the settling velocity is assumed to be depth-dependent, whereas in Equation 1.6 it is assumed to be constant.

A third steady-state phosphorus model is based on the fraction of influent phosphorus retained in the lake and is given by:

$$R_p = \frac{M - QP_o}{M} \quad 1.7$$

where,

R_p = retained influent phosphorus

M = $L \cdot A$

P_o = average outflow phosphorus concentration (mg/l)

thus

$$R_p = \frac{LA}{LA} - \frac{QP_o}{LA} = 1 - \left(\frac{Q}{A}\right) \frac{P_o}{L} \quad 1.8$$

and since $Q/A = z/\tau$

$$R_p = 1 - \frac{(z/\tau) P_o}{L} \quad 1.9$$

If $P_o = P$ (lake and outflow concentration are equal), then the model becomes:

$$P_o = P = \frac{L\tau}{z} (1 - R_p) \quad 1.10$$

And, finally, by setting Equation 1.3 equal to Equation 1.10, it can be shown that

$$R_p = \frac{\tau \sigma}{1 + \tau} = \frac{1}{1 + 1/\tau\sigma} \quad 1.11$$

These then are the basic models for lake phosphorus loadings. The question becomes which one to use. In considering the choice of an appropriate model, one needs to consider the following (Reckhow, 1979):

1. Consider the data base used to fit the model. Does it contain lakes similar to yours (similar, in size, depth, climate, etc.)?
2. Has the model been used successfully on similar lakes?
3. Are the model's limitations clearly documented?

In this paper the time dependent solution of Equation 1.6 is used to model phosphorus behavior in Lake Ontario and is given by:

$$P_i = \frac{L}{v_s + q_s} (1 - e^{-(1/\tau + v_s/z)\Delta t}) + P_{i-1} e^{-(1/\tau + v_s/z)\Delta t} \quad 1.12$$

Although specification in cross-sectional modeling is usually not an issue (Reckhow, 1979), this model is chosen in order to conform to previous work done by Chapra (1977).

CHAPTER II

THE MONTE CARLO METHOD AND ITS APPLICATION TO THE LAKE ONTARIO MODEL

The Monte Carlo Method

Definition

The Monte Carlo method is a procedure for internalizing the variability, and thus the uncertainty, of model parameters. By characterizing the variables as distributions rather than as single values (constants) the errors or variability associated with the variables is incorporated in the model. By repeatedly obtaining random values from these distributions, the model prediction also becomes a distribution of values reflecting the combined uncertainties of the independent variables. Exactly how the distributions are described varies from one case to another. In a normal distribution, for example, the arithmetic mean and the standard deviation are adequate descriptors. For the log-normal case, however, the geometric mean may be preferred in addition to the standard deviation or some measure of non-normality such as the coefficient of skewness. There are, however, sufficient statistics for this instance, too (Benjamin and Cornell, 1970).

Distribution Selection

Clearly the most critical aspect of the Monte Carlo method is the selection of the distributions which characterize the variables. There are both qualitative and quantitative ways to help the modeler make the appropriate distribution choice.

Qualitative basis.--The qualitative basis for the selection of distribution depends upon the extent to which the qualitative nature of the process is known. One such way is based on the behavioral similarity of the one process to some other known process (McGrath and Irving, 1973). Such a similarity allows for the assumption that the two processes may be characterized by the same family of distributions.

Another qualitative approach is based on the understanding of the underlying theory. For example, the failure of electrical components has been widely believed to follow exponential or Weibull distributions (Weibull, 1951). The deviations of shots from a bull's-eye is described by a Maxwell distribution in three dimensions and a Rayleigh distribution in two dimensions (Kendall and Stuart, 1958).

Other qualitative considerations which may help in distribution selection include a knowledge of whether or not the variable is discrete or continuous, symmetric, bounded, and so on.

Quantitative basis.--The amount of data is most important in distribution selection for it determines the quality of the empirical determination. Frequently, a small data base will allow the modeler

to make an appropriate choice of distributions. But if not, additional data must be obtained to properly select the distribution, especially if the model is highly sensitive to distribution choice. This, of course, must be balanced against the cost necessary to procure the additional data.

There are a number of techniques available to the modeler to analyze the data. The frequency histogram is one such technique which involves plotting the frequency with which a class of values occurs in the sample data. This provides a visual aid to determining the possible shape of the distribution. Of course, the fewer the data used, the less likely low probable events will be represented and the more misleading the histogram can be. In that case the histogram should be used cautiously and supplemented with some other technique.

Probability paper can also be used to help select the appropriate distribution. This paper comes with scaled ordinates such that the cumulative distribution function of the probability law plots as a straight line. Thus comparison of the proposed distribution with the data is reduced to a comparison between the cumulative frequency polygon of the data and a straight line. The closer the data are to the straight line, the better the fit. Probability paper exists for a number of distributions including the triangular, exponential and the normal (Benjamin and Cornell, 1970).

Parameter estimation is another common method for use with parametric distributions. A parametric distribution is one which is represented as a function with one or more parameters. Simple

parametric distributions have only one or two parameters such as the normal distribution. Complex parametric distributions, on the other hand, lack well-defined physical interpretations and are less easily expressed as a function and usually involve three to five parameters. Examples of these include the Weibull, Johnson and Pearson families of distributions.

After selecting a distribution there are a number of techniques which can be used to test the goodness-of-fit of the choice. This is particularly critical for those variables for which the Monte Carlo result is highly sensitive to distribution choice. Two of the most popular tests are the Chi-square and the Kolmogorov-Smirnov tests. In general, the more data used with the tests, the greater the likelihood of rejecting an inadequate distribution choice. Likewise, the less data used, the less likely of rejecting an inadequate choice. However, with smaller data sets the modeler can use the W-test for the normal distribution and the WE-test for the exponential distribution.

Application to the Model

Lee (1980) performed the Monte Carlo simulation on the time-dependent form of this model. For the sake of completeness and to facilitate comparisons, the characterization of the distributions and the manner in which the values are obtained from those distributions are briefly described below for each term in the model. The data used in the simulation are gathered from the literature, that is, from secondary sources. (For greater detail and for the computer algorithm, the reader is referred to Lee, 1980.)

Loadings

The loading term (L) is the single most complicated expression of the model. It is a combination of several terms, namely, diffuse source flux (DIFFLX); atmospheric source flux (ATMFLX); point source flux (PNTFLX); and Lake Erie source flux (EREFLX). These terms are added together and divided by the lake surface area (LSA) to give the total load to Lake Ontario:

$$L = (\text{DIFFLX} + \text{ATMFLX} + \text{PNTFLX} + \text{EREFLX})/\text{LSA}.$$

The flux terms are themselves composed of several terms which are considered below.

Diffuse flux (DIFFLX).--The diffuse source flux is the sum of three land use runoff concentrations: agricultural, urban and forested. Each concentration is represented by a log-normal distribution characterized by its mean and standard deviation derived from data taken from the literature (Beaulac, 1980). For more details concerning the selection of these data, see Reckhow et al. (1981).

The sum of these three concentrations is then multiplied by a randomly obtained value from the normal distribution for the Lake Ontario tributary flows (excluding the Niagara/Welland Canal flow). The distribution is characterized by a mean and standard deviation calculated from the flow of several tributaries (Chapra, 1981). The cross-correlations among tributary flows are sufficiently high to allow their treatment with a single distribution (see Appendix A.1).

The land use fractions are changed over the 40 year interval by increasing or decreasing the relative size of each land use fraction according to projected land use information (IJC, 1977). The projected difference for each land use fraction for the first 20 year period (1980 - 2000) is divided by 20 to yield an average annual change, which, when progressively summed over the 20 year interval accounts for the total change in the land use fractions. The same method holds for the second 20 year interval.

Atmospheric flux (ATMFLX).--The atmospheric source flux is the sum of the phosphorus contributions in the atmosphere from these three land uses. The three values are randomly obtained from three log-normal distributions defined by their means and standard deviations. The data used reflect total bulk loads (Reckhow et al., 1980). The relative weighting for each land use is determined by using an average land use fraction which is itself calculated according to the average relative size of each land use fraction over the entire 40 year period.

Point flux (PNTFLX).--Point source flux is the sum of phosphorus contributions attributable to primary, secondary and phosphorus removal wastewater treatments. Each treatment is represented by a normal distribution defined by a mean determined from survey data (DePinto et al., 1980). The variability of these point flux contributions is estimated by standard deviations about the mean concentrations for each treatment type (Reckhow, 1978).

Lake Erie flux (EREFLX).--The Lake Erie flux is calculated by multiplying a randomly obtained value from a normal concentration distribution by a lag-one autocorrelated flow value. This concentration distribution is defined by a mean and standard deviation from annual data (Chapra, 1981). The flow data are represented by a lag-one autocorrelated Markov flow model. The model uses 80% of the previous year's flow, 20% of the historical mean and a random component based on the standard deviation and correlation coefficient of the historical data to calculate the lag-one flow value (see Appendix A.2).

Settling Velocity

The settling velocity term is procured by randomly selecting values from a normal distribution defined by its mean and standard deviation calculated from Lake Ontario data (Chapra, 1981). This term also contains the model standard error which is included in the standard deviation; its value is about 2.7 g/m^3 . It is an estimate of the lack of model fit due to the assumption of a constant settling velocity.

Areal Water Loading

The areal water loading term is defined as

$$q_s = Q/A$$

where,

Q is the sum of the Ontario tributary flows, the Lake Erie flow, and the atmospheric water loading.

The calculations for the Ontario tributary flows are discussed above under Diffuse Flux and the Lake Erie flow is considered under Lake Erie Flux. The atmospheric water load from precipitation is calculated from historical rainfall data (World Weather Records), obtained from the meteorological records for the Lake Ontario basin. Pseudo-random values are generated from a normal distribution defined by the mean and the standard deviation derived from the data.

Hydraulic Retention Time

The hydraulic retention time (τ) is represented by a normal distribution defined by its mean and standard deviation calculated from lake volume and outflow data provided by Chapra (1981). The twenty years of outflow data were each divided by the respective year's lake volume to arrive at a distribution of hydraulic retention times.

Other Parameters

The time interval, Δt , is set to one year, for all iterations since only annual data are used. In terms of many planning considerations this is a shortcoming of the model, however, in terms of uncertainty analysis it is not a problem.

The previous phosphorus prediction, P_{i-1} , is randomly selected from a normal concentration distribution and is used to calculate the present year's, P_i , phosphorus prediction. It is initialized at .0200 mg/l.

The mean depth is set at 89 meters (Snodgrass, 1974), and is used for all runs.

CHAPTER III

RESULTS OF THE MONTE CARLO SIMULATION

Major Observations

The results of the Monte Carlo simulation indicate that the parameter values that are randomly generated from distributions characterized by literature data are, of course, accurately characterized in the simulation. The model parameters, which are calculated also agree closely with the findings of other Lake Ontario studies. The dependent variable, in-lake phosphorus concentration, is also reasonably well represented in the simulation. The model predictions agree with major Lake Ontario studies. Some highlights of these results are given below. (For a complete description, see Lee, 1980).

Base Run

The fifteen parameters of the Monte Carlo simulation include six log-normal distributions and nine normal distributions. The base run for the simulation is a 40-year, 100 iterations per year test (see Table 3.1 and Figure 3.1). The phosphorus concentration prediction mean for the forty year period is .0182 milligrams/liter (mg/l), with a standard deviation of .0005 mg/l and a coefficient of variation of 2.75 percent. This coefficient measures the year-to-year variation

in the annual phosphorus concentration predictions. The 14.3 percent mean annual coefficient of variation and the .0026 mg/l mean standard deviation estimate within-year variability which reflect the combined input errors and variabilities. (For additional results, see Table A.1).

All-Normal Run

The 100 iterations per year run using all-normal distributions is given in Table 3.1 and Figure 3.2. The average annual standard deviation dropped to .0016 mg/l and its corresponding coefficient of variation dropped to .0840. The inter-year standard deviation spanning the 40-year test, increased substantially to .0021 mg/l. Its corresponding coefficient of variation increased to .0620. The mean increased only slightly from .0182 to .0194 mg/l.

Lee suggests that this lower variability and slightly higher prediction may be due to the manner in which the normal variates are chosen. Unlike the log-normal case, negative values can be generated by the normal distributions. However, since negative concentrations and fluxes are difficult to interpret, only positive values are used. This results in two different sets of summary statistics. Since adequate data for comparison are not available, it is unclear whether or not this revision to all-normal distributions is more realistic.

It should be noted, though, that the simulation may be sensitive to changes in the random number seed. For example, increasing the magnitude of the random seed from 123457.D0 (double precision) to 987225.D0 results in an average annual concentration prediction about

TABLE 3.1.--Statistical Summary of Major Monte Carlo Runs

Base Run	\bar{x} = .0182	\overline{CV}	= .1440
(100 iterations per year)	$S_{\bar{x}}$ = .0005	\bar{s}	= .0027
	$CV_{\bar{x}}$ = .0275	y-intercept	= .0186
	$R_{\bar{x}}$ = .9989	Slope	= -.00002
		r	= -.4586
All Normal Run	\bar{x} = .0194	\overline{CV}	= .0840
	$s_{\bar{x}}$ = .0021	\bar{s}	= .0016
	$CV_{\bar{x}}$ = .0620	y-intercept	= .0207
	$R_{\bar{x}}$ = .9940	Slope	= -.00006
		r	= -.6367

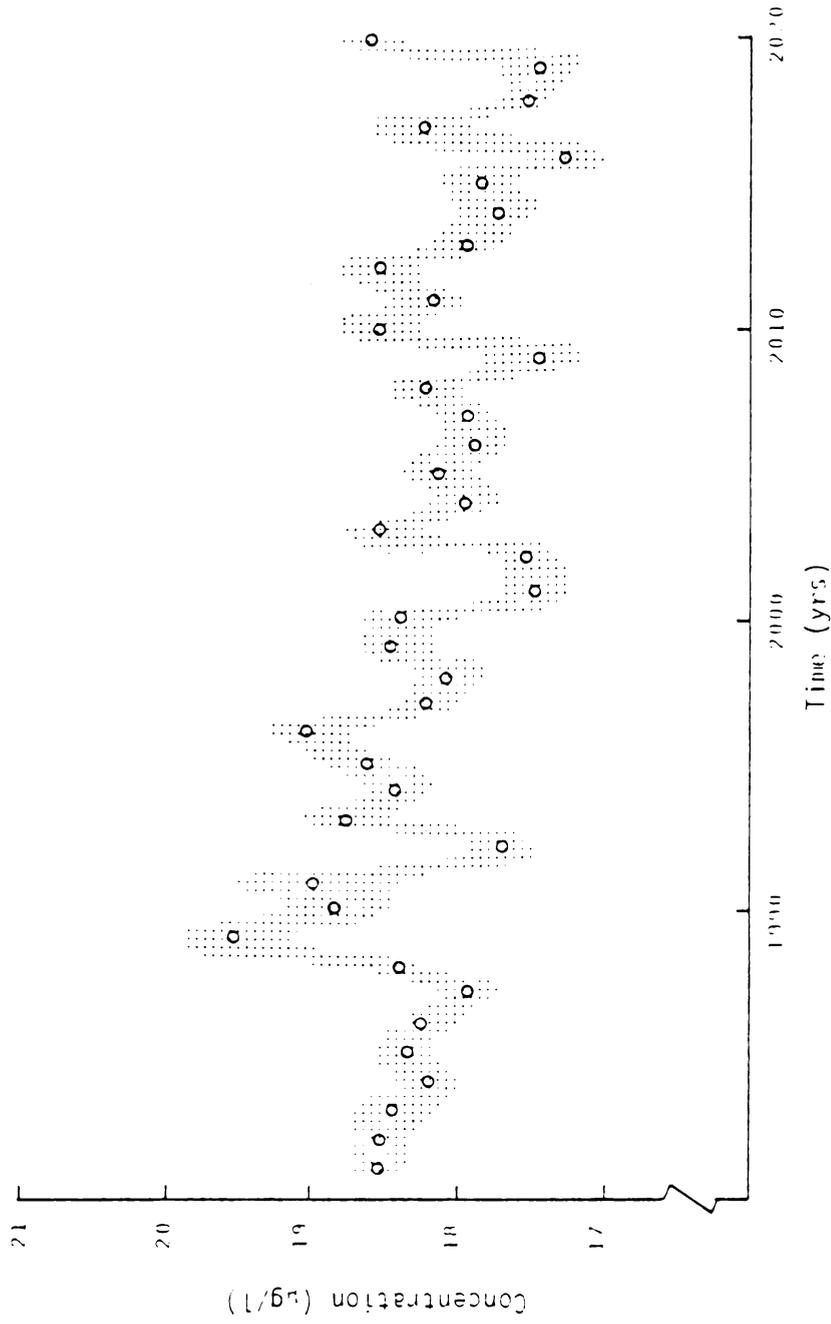


Figure 3.1.--40 Years of Phosphorus Concentration Predictions at 100 Iterations Per Year.
Source: Lee (1980).

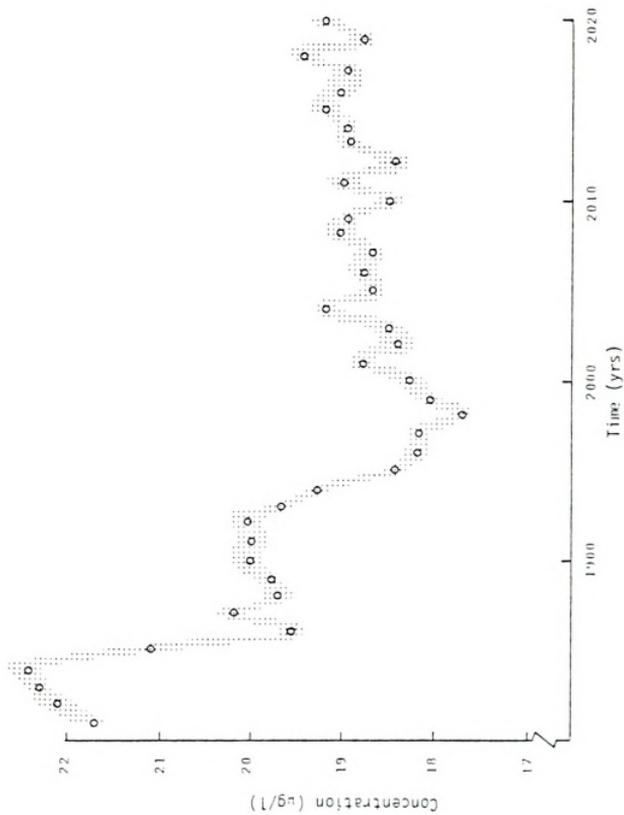


Figure 3.2.--40 Years of Phosphorus Concentration Predictions Using All Normal Distributions.
Source: Lee (1980).

9% higher than for the smaller seed (.0211 mg/l vs. .0194 mg/l). The average annual standard deviations, however, are about the same (.0017 vs. .0016); the average annual coefficient of variation for the larger seed is somewhat smaller (.0802 vs. .0840). The tendency of the prediction to increase or decrease over time may also be altered. In this case, it changes from a negative to a positive slope.

These changes suggest, according to Lee, that normal distributions are less characteristic of the parameters previously described by log-normal distributions. But the interrelationships among the parameters are so involved that such a suggestion requires considerable additional research to validate including more experiments to determine whether or not these results are consistently obtained. Hence, the problem of apparent sensitivity is not yet firmly established. For further details, see Lee, 1980.

Minor Observations

Base Run Fluxes

The phosphorus fluxes generally exhibited low variability for the base run. The Lake Erie flux, which comprises 42 percent of the total phosphorus input to Lake Ontario, had the lowest variability with a coefficient of variation of 1.1 percent. The second largest source of flux was diffuse flux, which constitutes 32.5 percent of the total and has a coefficient of variation of 8.8 percent which, though low, is the most variable of all the fluxes. Point source flux is third in influence with a coefficient of variation of only 2.5 percent. Atmospheric flux, constituting 8 percent of the total, has a

variability of 3.7 percent. (For further details, see Table A.2.)

Base Run Water Load

Water loading is comprised of the atmospheric, Ontario basin, tributary and Lake Erie sources. Atmospheric water loading, which is the smallest component, has a coefficient of variation of 2.0 percent. The variability of the load from the Ontario basin tributaries is approximately the same at 2.1 percent. The remaining 81 percent of the water load comes from Lake Erie which has the greatest influence on the overall system. It has a coefficient of variation of .3 percent. (For further details, see Table A.3.)

Higher Iterations Per Year

The 500 iterations per year simulation may also be sensitive to changes in the random number seed (123457.D0 to 987543.D0). However, the summary statistics of each of these runs do not significantly differ (.0191 mg/l vs. .0187 mg/l), and the mean annual variability is only slightly lower (14.3% vs. 14.4%). In both instances, the concentration appears to change very little over the 40-year period (see Figure A.1 and Table A.4).

The 1,000 iterations per year run, though giving similar results, displayed the problem of periodicity (see Figure A.2). The graph gives an almost sinusoidal impression. Although several aspects of the simulation were examined, no answer to this problem was found.

Flux Sensitivities

Reducing the point source flux variability gave no great deviations from the base run. Both runs gave the same average annual concentration prediction (.0182 mg/l) as well as the same average annual coefficient of variation (14.3%). See Table A.5 and Figure A.3.

Reducing the variabilities of the treatment facility effluent has very little effect on the total point source flux. The upgrading to all phosphorus removal treatment, though, has a noticeable impact: it reduces the mean annual point flux value by 27 percent. See Table A.5 and Figure A.4.

The concentration values used to characterize the flux from Lake Erie are those taken from Niagara River-Welland Canal data. Since flux from Lake Erie is the largest single source, it is quite predictable that changing this term gives distinct changes in the final concentration distribution. The mean for the eastern Lake Erie concentration data is 23 percent lower than that used for the Niagara-Welland data (.0170 mg/l vs. .0220 mg/l), and results in an average annual concentration prediction 1.7 mg/l less than the 18.2 mg/l mean value for the Niagara-Welland run. Also, the eastern Lake Erie data is 15 percent more variable than the Niagara-Welland data. The mean annual coefficient of variation for the entire period increases by 2 percent (see Table A.5 and Figure A.5).

Land Use Fraction Sensitivities

Altering the balance of the land use fractions has virtually no effect on the concentration predictions over the 40-year test period. The only slight change appears in the annual coefficient of variation which decreases slightly from 14.3 percent to 14.1 percent when the fractions are doubled, and down to 13.8 percent when they are tripled.

Shifting land use to all urban, or all agriculture or all forest by the end of the 40-year period did have an impact on the prediction. The results clearly show the difference between disturbed and undisturbed ecosystems as far as their contributions to lake phosphorus concentration are concerned. The two disturbed systems, urban and agriculture, contribute to increasing in-lake concentrations as their respective fractions increase. Increasing the undisturbed system fraction, forest, results in a decrease of the concentration predictions of .7 mg/l every 5 years. The average annual variability of these predictions is a very low 10.4 percent (see Table A.6 and Figure A.6).

Summary

The Monte Carlo results can be briefly summarized as follows. The base run exhibited low variability and therefore gave a small prediction error. The method was sensitive to the choice of distributions. For example, the switch to all normal distributions increased the prediction slightly, lowered the intra-year standard deviation (or variability), but increased the intra-year standard deviation (or

variability). The method may be sensitive to changes in the random number seed, especially with the all-normal run. It was not highly sensitive to increased per year iterations. The fluxes, too, were relatively insensitive to changes in variability. However, the use of a different data set lowered the prediction mean and increased the error. With respect to the land use fractions, the model was sensitive only to differences between disturbed and undisturbed land uses.

CHAPTER IV

FIRST-ORDER ERROR ANALYSIS

An alternative method for estimating the error associated with model predictions is known as first-order error analysis. This method is characterized by two basic features: single moment treatment of the uncertain, or random, component and the first-order analysis of functional relationships among the variables.

The first characteristic implies that the random variable X is described by its first and second moments, m_x^1 and m_x^2 , which are given by:

$$\int_{-\infty}^{+\infty} x^n g(x) dx = E[x^n] = m_x^n \quad 4.1$$

where,

$g(x)$ is the probability density function and $E[x^n]$ is the expectation of x^n (Benjamin and Cornell, 1970).

Thus the first moment of X , m_x^1 , with respect to the origin is the mean. The second moment of X , m_x^2 , is the variance.

When the moments are found with respect to the mean, they are called central moments and are defined as:

$$\int_{-\infty}^{+\infty} (x-\bar{x})^n g(x) dx = E[(x-\bar{x})^n] = \mu_x^n. \quad 4.2$$

In this case the first central moment, M_x , is zero and the variance is the first non-zero moment (Benjamin and Cornell, 1970).

The second characteristic implies that only the first-order or linear terms in a Taylor series expansion will be retained when dealing with system or functional relationships among random variables. In the case of a function, $y = f(x)$, of a random variable x , the question arises as to how to determine the moments of the dependent variable. In this regard, it should be observed that if the coefficient of variation is small, x is very likely to lie close to m_x and thus a Taylor series expansion of $f(x)$ about m_x is suggested as an approximation to the moments of y (Benjamin and Cornell, 1970).

For a single variable relationship the Taylor series expansion may be written as:

$$f(x) = f(x_0) + \left. \frac{df(x)}{dx} \right|_{x_0} (x-x_0) + \left. \frac{1}{2} \frac{d^2f(x)}{dx^2} \right|_{x_0} (x-x_0)^2 + \dots \quad 4.3$$

If the error $(x-x_0)$ is small and if the higher-order derivations are not large, i.e., if the relationship is not substantially non-linear in the area of interest, the above equation reduces to:

$$f(x) \cong f(x_0) + \left. \frac{df(x)}{dx} \right|_{x_0} (x-x_0). \quad 4.4$$

Taking expectations of both sides yields the first-order approximation to the first moment of Y , the mean:

$$E[f(x)] \cong f(E[x]) = f(x_0). \quad 4.5$$

The second term drops out since the $E[x-x_0] = 0$.

Similarly, finding the variance of both sides yields the first-order approximation for the variance of $f(x)$:

$$\text{Var} [f(x)] \cong \text{Var} \left[f(x_0) + \left. \frac{df(x)}{dx} \right|_{x_0} (x-x_0) \right] \quad 4.6$$

and after carrying out the multiplication and rearranging terms gives:

$$\text{Var} [f(x)] \cong \text{Var} \left[f(x_0) - x_0 \left. \frac{df(x)}{dx} \right|_{x_0} + \left. \frac{df(x)}{dx} \right|_{x_0} x \right]. \quad 4.7$$

Since the first two terms are constants (hence, zero variance), they do not affect the variance of $f(x)$ and thus,

$$\text{Var} [f(x)] \cong \text{Var} \left[\left. \frac{df(x)}{dx} \right|_{x_0} \cdot x \right] \cong \left[\left. \frac{df(x)}{dx} \right|_{x_0} \right]^2 \text{Var} [x]. \quad 4.8$$

A similar result holds for the multi-variate case starting with a Taylor series expansion of partial derivatives and finding the first-order approximation of the expected value of

$Y = f(x_1, x_2, \dots, x_n)$ which gives

$$\begin{aligned} E[Y] &= E[f(x_1, x_2, \dots, x_n)] \\ &= f(E[x_1, x_2, \dots, x_n]) \\ &= f(x_{10}, x_{20}, \dots, x_{n0}). \end{aligned} \quad 4.9$$

where,

x_{i0} is the mean of the i th random variable.

Because the second term drops out again, $E(x_i - x_{i0}) = 0$, there is no way of accounting for the covariance of the variables unless the second-order approximation is determined:

$$E[Y] \cong f(x_{10}, x_{20}, \dots, x_{n0}) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \Big|_{x_0} \quad 4.10$$

where,

$\frac{\partial^2 f}{\partial x_i \partial x_j} \Big|_{x_0}$ is the second partial derivative of $f(x_1, x_2, \dots, x_n)$ with respect to x_i and x_j evaluated at their means $x_{10}, x_{20}, \dots, x_{n0}$.

If the coefficients of variation of the random variables, x_i , and the non-linearity in the function are not large, this second term is negligible (Benjamin and Cornell, 1970).

The first-order approximation of the variance is:

$$\text{Var} [Y] \cong \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i} \Big|_{x_{i0}} \frac{\partial f}{\partial x_j} \Big|_{x_{j0}} \text{Cov}[x_i, x_j]. \quad 4.11$$

However, if the x_i 's are uncorrelated, then:

$$\text{Var} [Y] \cong \sum_{i=1}^n \left[\frac{\partial f}{\partial x_i} \Big|_{x_{i0}} \right]^2 \cdot \text{Var} [x_i]. \quad 4.12$$

This equation shows that each of the n random variables contributes to the dispersion of Y in a manner proportional to its own variance and proportional to the factor, $(\partial f / \partial x_i \Big|_{x_{i0}})^2$, which relates changes in x to changes in Y , a sensitivity factor (Benjamin and Cornell, 1970).

The total prediction error includes the model error, the contributions to the uncertainty from the parameter error and the independent variable error. These are combined as follows:

$$S_t^2 = S_\mu^2 + S_p^2 + S_v^2 \quad 4.13$$

where,

S_t = total error of the prediction

S_μ = standard error of the estimate for the model

S_p = parameter error contribution

S_v = input variable error contribution.

The standard error of the estimate was approximated by a non-linear regression routine and includes errors in the variables. It is referred to hereafter as the model error and is treated as a constant in the error equation. However, since this routine was initialized from an estimate based on the steady-state form of the model, but is applied to the time-dependent form, a non-quantified source of error is

introduced. (See Appendix B-1 for a discussion of the non-linear regression routine.) The parameter error and the input variable error are treated by first-order analysis using the equation with the covariance term (Equation 4.11) since some of the x_i 's are correlated. This equation may be rewritten as:

$$\text{Var} [Y] \cong \sum_{i=1}^n \left(\frac{\partial P}{\partial x_i}\right)^2 S_i^2 + \sum_{i=1}^n \sum_{j=1+1}^n 2 \cdot \frac{\partial P}{\partial x_i} \frac{\partial P}{\partial x_j} \cdot S_i S_j R_{ij}. \quad 4.14$$

The parameter, v_s , is not correlated theoretically with any other variable and thus its error contribution reduces to

$$S_p^2 = \left(\frac{\partial P}{\partial v_s}\right)^2 S_i^2. \quad 4.15$$

The mean and variance for v_s were also estimated from the non-linear regression routine mentioned above (see Appendix B-1) since no data exist for v_s . Recalling the model (Equation 1.12), this gives:

$$\frac{\partial P}{\partial v_s} = \frac{(e^{-v_s/z-1/\tau}-1)L}{(v_s + q_s)^2} + \left(p_{i-1} - \frac{L}{v_s + q_s}\right) (-1/ze^{-v_s/z} - 1/\tau). \quad 4.16$$

This quantity is squared and then scaled by the variance of v_s .

The variables L , q_s , and τ are correlated. Their correlation coefficients were derived from linear regression using the fourteen years of data from 1960 to 1973 for these variables. These data are



taken from the work of Steven Chapra (Chapra, 1981). The means and variances were also calculated from these data. The input variable P_{i-1} is not theoretically correlated with any other variable in the dependent form (Equation 1.12), though linear regression with the other variables yields a relatively high correlation coefficient. This is believed to be due to the small amount of data and the low variability of the data. Therefore, the correlation coefficient was assumed to be zero. The total input variable error is

$$\begin{aligned}
 S_v^2 = & \left(\frac{\partial P}{\partial L}\right)^2 \cdot S^2(L) + \left(\frac{\partial P}{\partial q_s}\right)^2 \cdot S^2(q_s) + \left(\frac{\partial P}{\partial \tau}\right)^2 \cdot S^2(\tau) \\
 & + \left(\frac{\partial P}{\partial P_{i-1}}\right)^2 \cdot S^2(P_{i-1}) + 2 \cdot \frac{\partial P}{\partial q_s} \cdot \frac{\partial P}{\partial L} \cdot S(q_s) \cdot S(L) \cdot r(q_s, L) \\
 & + 2 \cdot \frac{\partial P}{\partial L} \cdot \frac{\partial P}{\partial \tau} \cdot S(L) \cdot S(\tau) \cdot r(L, \tau) \\
 & + 2 \cdot \frac{\partial P}{\partial q_s} \cdot \frac{\partial P}{\partial \tau} \cdot S(q_s) \cdot S(\tau) \cdot r(q_s, \tau). \quad 4.17
 \end{aligned}$$

(The term z is the mean depth and is set equal to 89 meters.) The partial derivatives for the above equation are:

$$\frac{\partial P}{\partial L} = \frac{1 - e^{-v_s/z} - 1/\tau}{v_s + q_s} \quad 4.18$$

$$\frac{\partial P}{\partial q_s} = \frac{L \cdot (e^{-v_s/z} - 1/\tau - 1)}{(v_s + q_s)^2} \quad 4.19$$

$$\frac{\partial P}{\partial \tau} = \left(P_{i-1} - \frac{L}{v_s + q_s} \right) \cdot \tau^{-2} \cdot e^{-v_s/z - 1/\tau} \quad 4.20$$

$$\frac{\partial P}{\partial P_{i-1}} = e^{-v_s/z - 1/\tau} \quad 4.21$$

The full equation for the total error of the prediction is:

$$\begin{aligned} S_t^2 = & \left[\frac{(e^{-v_s/z - 1/\tau} - 1)L}{(v_s + q_s)^2} + \left(P_{i-1} - \frac{L}{v_s + q_s} \right) (-1/ze^{-v_s/z - 1/\tau}) \right]^2 \\ & \cdot S^2(v_s) \\ & + \left[\frac{L(e^{-v_s/z - 1/\tau} - 1)}{(v_s + q_s)^2} \right]^2 \cdot S^2(q_s) + \left[\frac{1 - e^{-v_s/z - 1/\tau}}{v_s + q_s} \right]^2 \cdot S^2(L) \\ & + \left[\left(P_{i-1} - \frac{L}{v_s + q_s} \right) \tau^{-2} e^{-v_s/z - 1/\tau} \right]^2 \cdot S^2(\tau) \\ & + (e^{-v_s/z - 1/\tau})^2 \cdot S^2(P_{i-1}) \quad 4.22 \\ & + \frac{2L(e^{-v_s/z - 1/\tau} - 1)}{(v_s + q_s)^2} \frac{(1 - e^{-v_s/z - 1/\tau})}{(v_s + q_s)} \\ & \cdot S(q_s) \cdot S(L) \cdot r(q_s, L) \\ & + \frac{2(1 - e^{-v_s/z - 1/\tau})}{v_s + q_s} \cdot \left(P_{i-1} - \frac{L}{v_s + q_s} \right) \tau^{-2} e^{-v_s/z - 1/\tau} \end{aligned}$$

$$\begin{aligned}
& \cdot S(L) \cdot S(\tau) \cdot r(L, \tau) \\
& + \frac{2 L (e^{-v_s/z-1/\tau}-1)}{(v_s + q_s)^2} \cdot \left(P_{i-1} - \frac{L}{v_s + q_s} \right) \tau^{-2} e^{-v_s/z-1/\tau} \\
& \cdot S(q_s) \cdot S(\tau) \cdot r(q_s, \tau) + ME.
\end{aligned}$$

The values used for the base run first-order analysis are summarized below:

Error type	Mean	S	S ²
<u>Parameter</u>			
v _s	19.1910	1.1963	1.4311
<u>Input Variables</u>			
q _s	10.660	1.4608	2.1339
L	.6352	.0811	.0066
τ	7.9402	1.0421	1.0860
P _{i-1}	.0206	.0027	.00000729
<u>Model</u>			
ME		.0032	.00001024

The correlation coefficients are:

$$\begin{aligned}
q_s \text{ with } L: & \quad .6822 \\
q_s \text{ with } \tau: & \quad -.9902 \\
L \text{ with } \tau: & \quad -.7078.
\end{aligned}$$



CHAPTER V

RESULTS FOR FIRST-ORDER ANALYSIS

The Model Prediction

The results of forty iterations representing forty years of phosphorus concentration predictions are presented in Table 5.1 and Figure 5.2. The model is initialized at .0206 mg/l and the first iteration predicts .0208 mg/l. From there the predictions grow to .0213 mg/l after which there is no change.

The decay factor is the key to understanding these results. The equation is:

$$P_i = \frac{L}{v_s + q_s} (1 - e^{-v_s/z - 1/\tau}) + P_{i-1} e^{-v_s/z - 1/\tau}. \quad 5.1$$

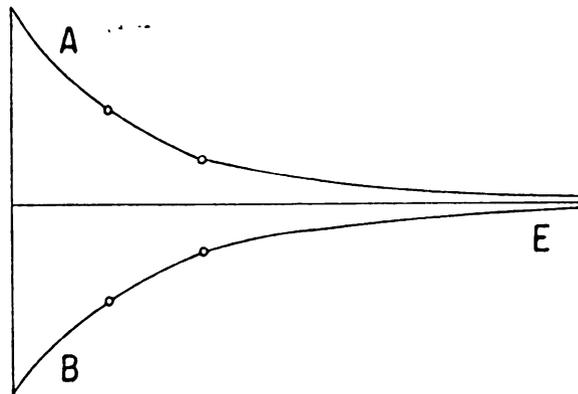
The term $e^{-v_s/z - 1/\tau}$ is the decay factor. In the second term, the decay factor reduces the previous prediction by a fraction of a fraction. The base run values for v_s , z , and τ yield a decay factor of about .71. The prediction, then, for each iteration is 71 percent of the previous value, while the first term remains constant at approximately 29 percent of the steady-state value. Thus the first term constitutes an increasingly larger proportion of the prediction while the second term continuously diminishes. The effect is to move the prediction ever closer to the steady-state value at which point the model ceases to be dynamic (see Appendix B-2).

TABLE 5.1.--Base Run

Prediction (.0206)*	Error (.0027)*	Coefficient of Variation (.1311)*
1. .0208	.0038	.1827
2. .0209	.0042	.2010
3. .0210	.0044	.2095
4. .0211	.0045	.2133
5. .0212	.0046	.2170
6. .0212	.0046	.2170
7. .0212	.0046	.2170
8. .0212	.0046	.2170
9. .0212	.0046	.2170
10. .0213	.0046	.2160
.		
.		
.		
40. .0213	.0046	.2160
\bar{x} .0213	.0046	.2147
S .0001	.0001	.0058
CV .0053	.0313	.0271

*Initial values.

Behavior of Predictions and Errors



A = initialized above equilibrium value
B = initialized below equilibrium value
E = equilibrium value

Figure 5.1.--Theoretical Behavior of Predictions and Errors.

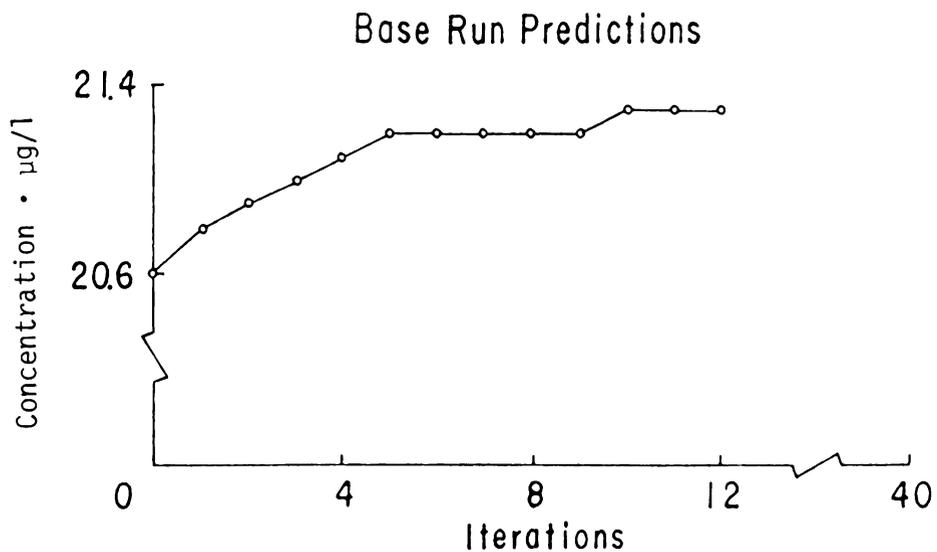


Figure 5.2.--Base Run Predictions.

The behavior of the model predictions is easier to see by using a different form of the model, namely,

$$P_i = SS + (P_0 - SS)k^i \quad 5.2$$

where,

$$SS = \frac{L}{v_s + q_s}, \text{ the steady state}$$

P_0 = initial phosphorus concentration value

$$k = 1/e^{v_s/z} + 1/\tau$$

i = ith iteration.

(See Appendix B-2 for this derivation.)

There are three points to note here. First, the predictions asymptotically approach the steady-state, SS and

$$\lim_{i \rightarrow \infty} P_i = SS.$$

(See Appendix B-3.)

Second, the predictions grow or decay depending upon the difference between the initial and steady-state values. A negative difference, that is, an initial value below the steady-state level, results in growth and a positive difference, that is, an initial value above the steady-state level, results in decay (see Figure 5.1). For example, consider the above equation (Equation 5.2) with the following values,

$$P_i = .0200 + (P_0 - .0200)(.71)^i$$

where

$$SS = .0200$$

$$k = .71$$

If P_0 is initialized at a value below .0200, say at .0180, the difference between P_0 and SS will be negative and

$$\begin{aligned} P_i &= .0200 + (.0180 - .0200)(.71)^i \\ &= .0200 - .0020 (.71)^i. \end{aligned}$$

The amount subtracted from .0200 decreases with each iteration i , thus causing the prediction to grow. If, however, P_0 is initialized at a value above .0200, say at .0220, the difference between P_0 and SS will be positive and

$$\begin{aligned} P_i &= .0200 + (.0220 - .0200)(.71)^i \\ &= .0200 + .0020 (.71)^i. \end{aligned}$$

The amount added to .0200 decreases with each iteration i , thus causing the prediction to decay.

Third, this equation can be solved for i and thereby the number of iterations required to attain the steady-state value, within any desired level of precision, can be determined. This amounts to calculating the number of iterations needed to attain a desired phosphorus concentration. Thus,

$$i = \ln \frac{P_i - SS}{P_0 - SS} \cdot \frac{1}{-(V_s/z + 1/\tau)} \quad 5.3$$

(See Appendix B-4).

The Error Equation

Major Observations

The forty error estimates corresponding to the forty predictions are given in the second column of Tables 5.1. Figures 5.3 and 5.4 give the graphs of the parameter and the input variable errors. The error is initialized at $.0027^2$ and jumps to $.0038^2$ at the first iteration. It then grows steadily to $.0046$ after which it remains constant (see Figure 5.5). These results are more easily accounted for by restating the error equation in a more manageable form in terms of its basic components.

The first component of the error equation to note is the summation of all those terms which are essentially the same as the steady-state error terms. Indeed, these terms differ only by the amount $(1 - k^2)$. These terms are referred to hereafter as the fractional steady-state error and notated as FSSE.

The second component to note is the summation of all those terms which are scaled by the expression,

$$P_{i-1} - \frac{L}{v_s + q_s} \cdot$$

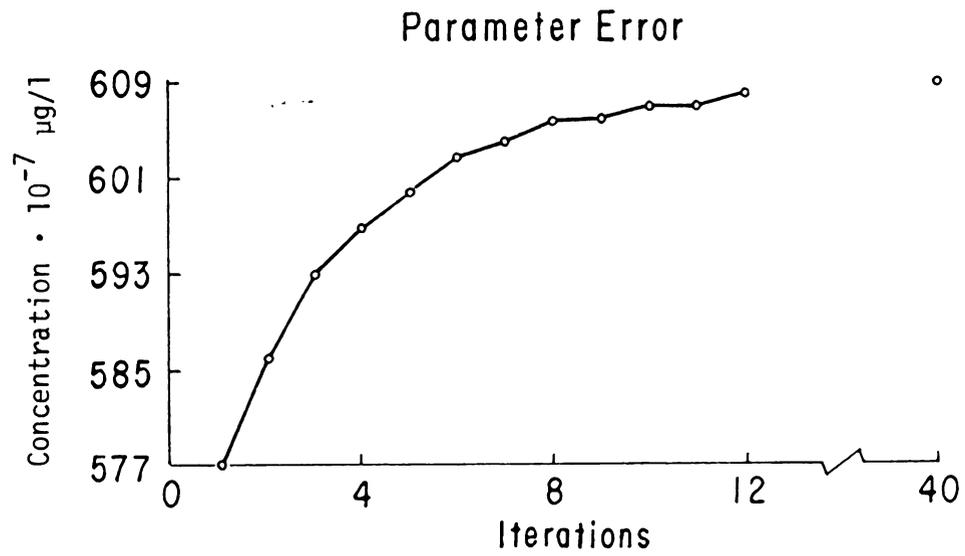


Figure 5.3.--Parameter Error.

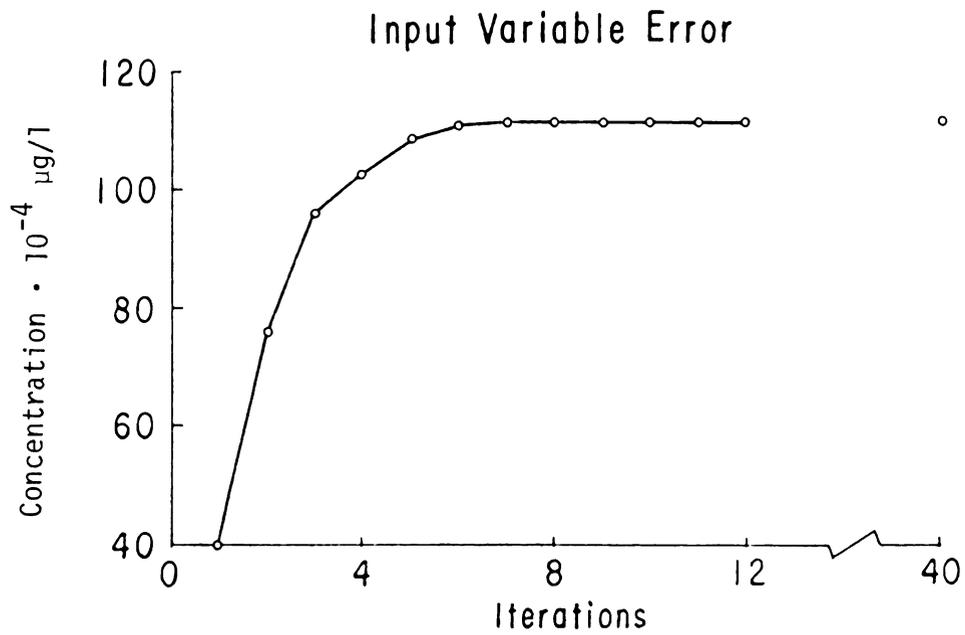
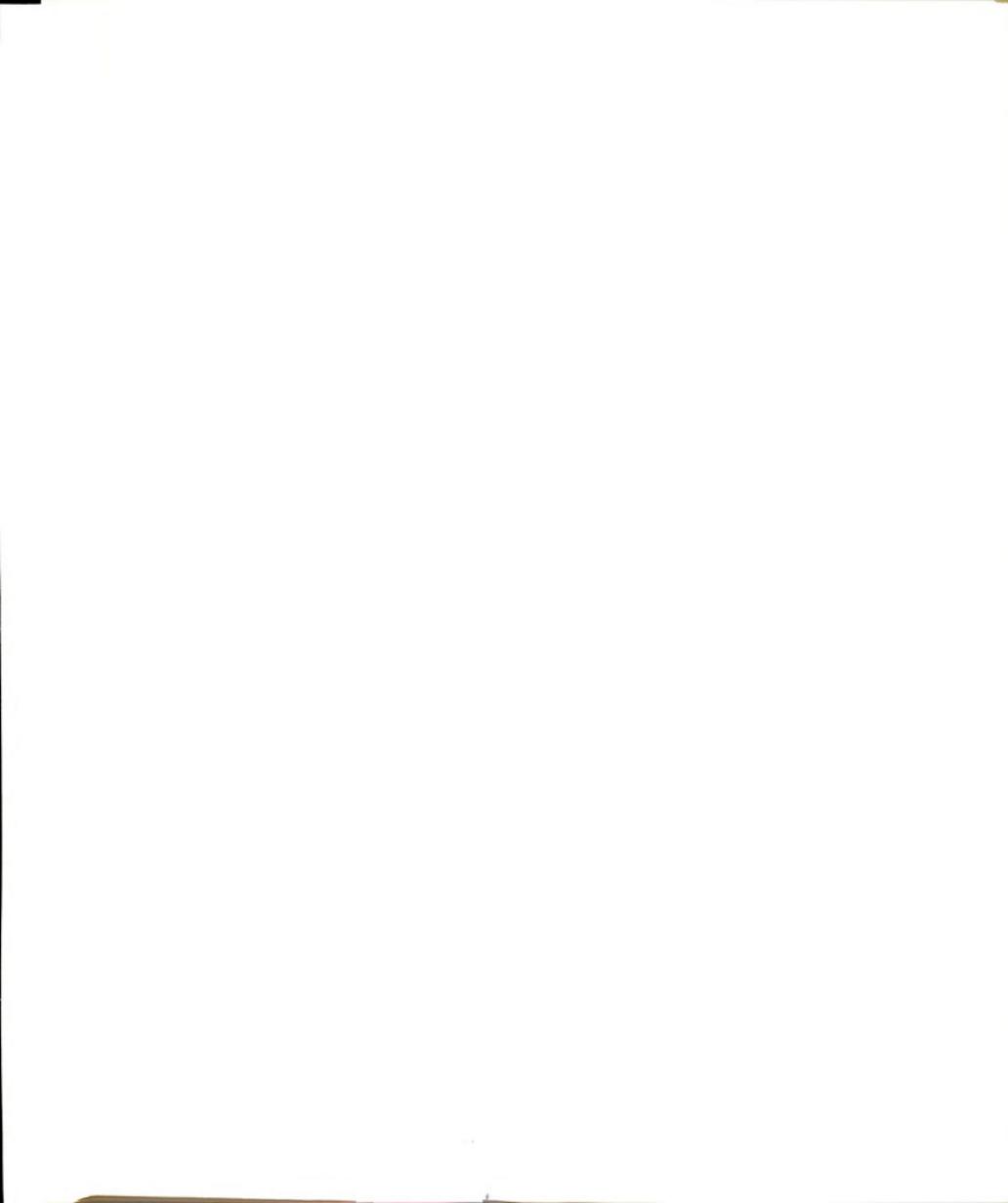


Figure 5.4.--Input Variable Error.



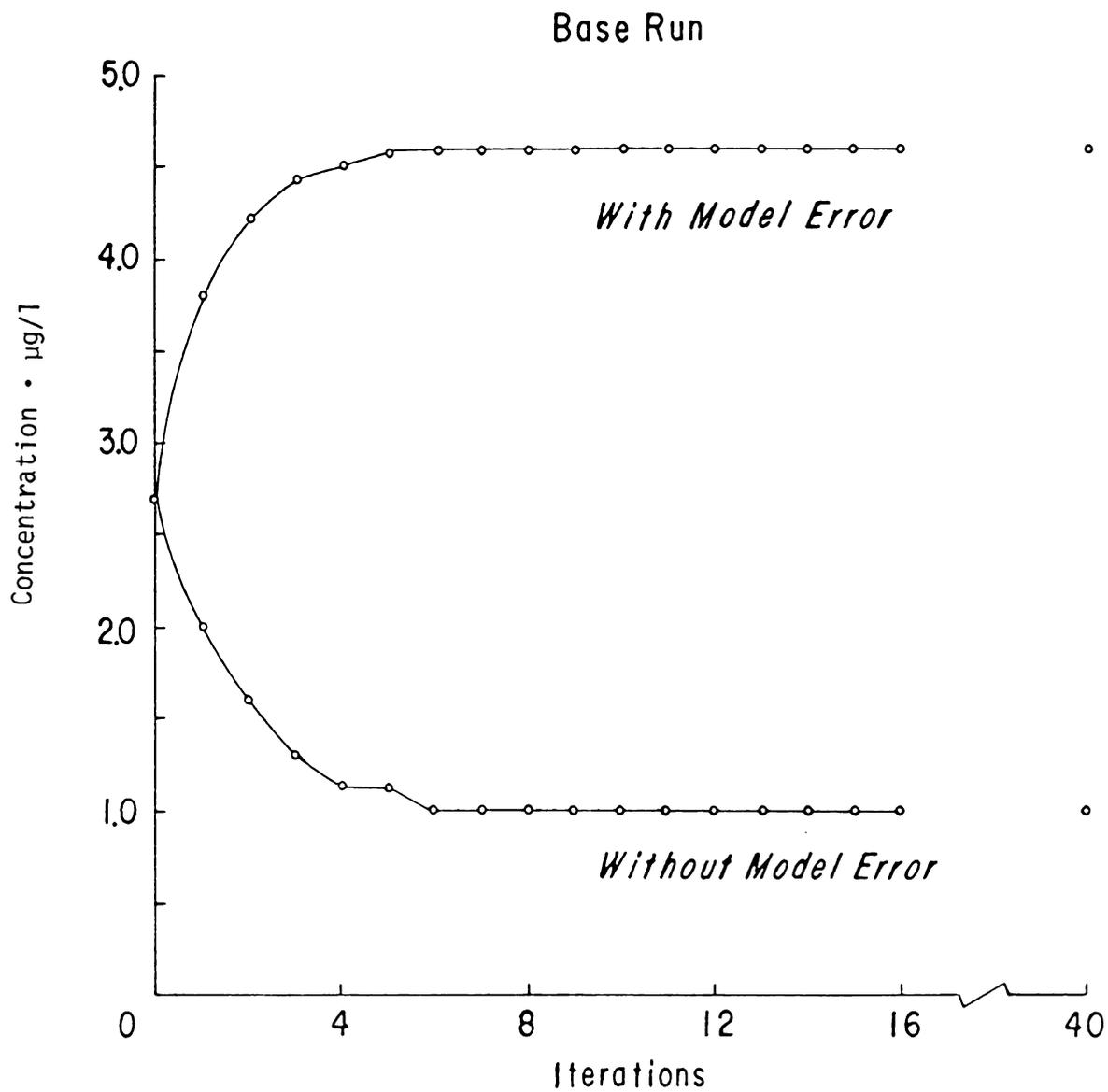


Figure 5.5.--Base Run--Total Error.

TABLE 5.2.--Modified Base Run (without model error)

Prediction	Error	Coefficient of Variation
(.0206)*	(.0027)*	(.1311)*
1. .0208	.0020	.0962
2. .0209	.0016	.0766
3. .0210	.0013	.0619
4. .0211	.0011	.0521
5. .0212	.0011	.0472
6. .0212	.0010	.0472
7. .0212	.0010	.0472
8. .0212	.0010	.0472
9. .0212	.0010	.0472
10. .0213	.0010	.0469
.		
.		
.		
40. .0213	.0010	.0469
\bar{x} .0213	.0011	.0495
S .0001	.0002	.0092
CV .0053	.1774	.1860

*Initial values.

This is the difference between the previous prediction and the steady-state value. This difference diminishes with each iteration and hence is referred to as the "diminishing" factor and notated as D_i where "i" is the ith iteration.

The third element of the error equation may be thought of as a "reduction" term consisting of the decay factor k^2 and the previous error,

$$\left(\frac{\partial P_i}{\partial P_{i-1}} \right)^2 S^2 (P_{i-1}) = k^2 S^2 (P_{i-1}). \quad 5.3$$

This term retains a fraction of the previous error, namely, k^2 , and loses an amount equal to $1-k^2$. As with the model, the decay factor plays the central role.

A fourth component to note is the model error, ME, which is a constant in the error equation.

Using the above defined abbreviations and letting $S^2(P_{i-1}) = E_{i-1}$ and $S_t^2 = E_i$, the error equation becomes,

$$E_i = \text{FSSE} + \text{ME} + D_i + k^2 E_{i-1}. \quad 5.4$$

The base run error estimates apparently approach an equilibrium value at which point the error equation is no longer dynamic. This means that the amount lost by the reduction term each iteration must be exactly compensated for by the other terms. Then at infinity,

$$E_\infty = k^2 E_\infty + \text{FSSE} + \text{ME} + D_\infty \quad 5.5$$



If $R = \text{FSEE} + \text{ME} + D_\infty$, then

$$E_\infty = k^2 E_\infty + R. \quad 5.6$$

Since the sum of the remaining terms must equal the amount lost to have equilibrium, then

$$R = (1-k^2) E \text{ and}$$

$$E_\infty = \frac{R}{1-k^2}$$

Substituting for E_∞ in Equation 5.6 gives,

$$E_\infty = \frac{k^2 R}{1-k^2} + R = \frac{k^2}{1-k^2} + 1 \quad R = \frac{R}{1-k^2}. \quad 5.7$$

(For a rigorous mathematical treatment, see Appendix B-5).

Since at infinity the diminishing term, D_i , has vanished altogether,

$$R = \text{FSSE} + \text{ME} \text{ and}$$

$$E_\infty = \frac{\text{FSSE} + \text{ME}}{1-k^2}. \quad 5.8$$

Thus the error converges to the sum of the fractional steady-state error and the model error divided by the constant $1-k^2$ (see Appendix B-6).

The component form of the error equation, E_i , requires a knowledge of the previous error, E_{i-1} . This equation has a more

general form which allows for calculation independent of the previous error by using the generator form for an infinite series (see Appendix B-7),

$$E_i = \frac{F}{1-k^2} (1-k^{2i}) + E_0 k^{2i} + D_i \frac{(1-k^{2i})}{1-k^2} \quad 5.9$$

where,

E_i = ith error or variance

F = FSSE + ME

E_0 = the initial error

k = decay factor

D_i = ith diminishing term

and

$$\begin{aligned} E_i &= \left(E_0 - \frac{F}{1-k^2} \right) k^{2i} + \frac{F}{1-k^2} + D_i \frac{(1-k^{2i})}{1-k^2} \\ &= \frac{F}{1-k^2} + \left(E_0 - \frac{F}{1-k^2} \right) k^{2i} + D_i \frac{(1-k^{2i})}{1-k^2} . \end{aligned} \quad 5.10$$

In this form it is easier to see just how the error grow or decays. It depends upon the difference between the initial and final errors. If the difference is positive, the error will decay toward the final value $F/(1-k^2)$ and if it is negative, the error will grow toward the final value. The error is based upon the final value, plus or minus an ever shrinking fraction of this difference.



Minor Observations

Diminishing factor.--It was found that the D_i term contributes a negligible amount to the base run total error. Its omission allows for a simpler equation which gives excellent approximations

$$E_i = (E_0 - \frac{F}{1-k^2}) k^{2i} + \frac{F}{1-k^2} . \quad 5.11$$

This also allows for a simpler program and hence is easier to model (see Appendix B-6).

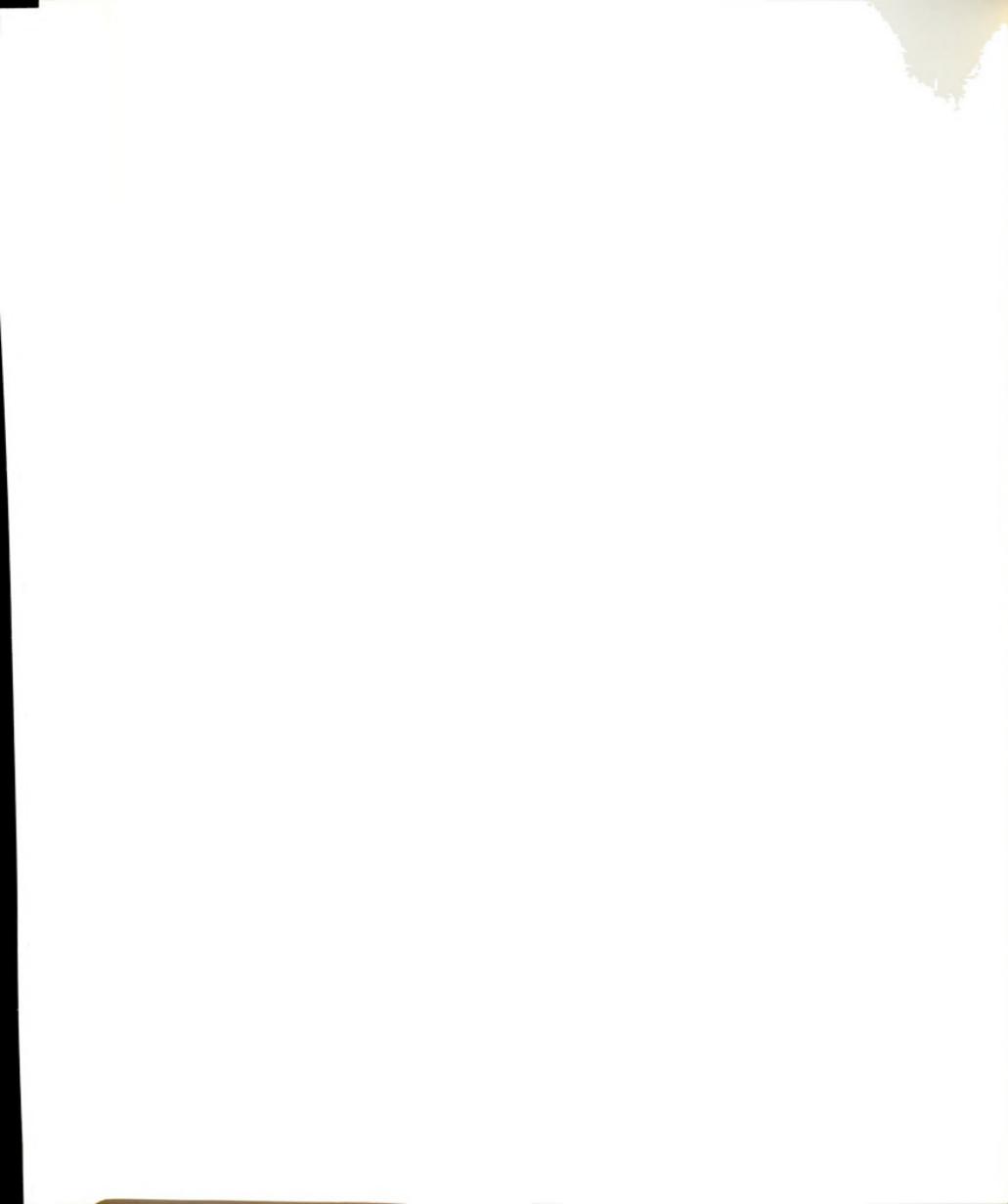
Calculation of i.--Using the above equation, one can calculate how many iterations are required to attain the equilibrium value within any desired level of precision. This amounts to calculating the number of iterations needed to attain a desired error level,

$$i = \ln \frac{E_i - F/(1-k^2)}{E_0 - F/(1-k^2)} \quad 5.12$$

$$\frac{-2 (v_s/z + 1/\tau)}$$

(see Appendix B-7).

Initial value.--In reviewing the base run errors, one notes the jump from the initial value of .0027 to .0038. This is accounted for by observing that the first iteration will produce an error at least equal to the FSSE. The greater the difference between FSSE and E_0 , the greater the increment at the first iteration.



Model error.--The base run was made with an independent model error term, but to facilitate comparison with the Monte Carlo simulation, the program was run without it (see Figure 5.5 and Table 5.2).

Calculation of the final value without the model error is straightforward. One need only remove the constant ME, yielding,

$$E_{\infty} = \frac{FSSE}{1-k^2} . \quad 5.13$$

For example, in the base run, the final value is $215.918 \cdot 10^{-7}$ (in variance form). The final value without the model error is,

$$\begin{aligned} E_{\infty} &= \frac{215.918 \cdot 10^{-7}(1-k^2) - .0032^2}{1-k^2} \\ &= \frac{4.48 \cdot 10^{-7}}{1-k^2} \\ &= 9.050 \cdot 10^{-7} \end{aligned}$$

The standard deviation is .0010.

Symmetry.--Figure 5.5 shows that the base run and modified base run standard deviations are approximately symmetrical. This is because the initial standard deviation is about halfway between the two different equilibrium values. Moving this initial value closer to one equilibrium or the other will cause the error to converge to that equilibrium value more quickly.



Coefficient of variation.--Because both the model prediction and the error converge, the coefficient of variation, CV, also approaches a final value, namely,

$$\begin{aligned} CV_{\infty} &= \frac{E_{\infty}}{P_{\infty}} \\ &= \frac{F/(1-k^2)}{SS} . \end{aligned} \quad 5.14$$

where,

E_{∞} = the variance at infinity

P_{∞} = the prediction at infinity, that is, the steady-state, SS.

F = the sum of the fractional steady-state error and the model error (see Appendix B-8).

However, the coefficient need not approach this final value in the same well-behaved manner that the error and the prediction do, that is, it can have values within the iterations which exceed the final value. This is due to the differential rates of convergence of the prediction and error equations and due to the manner in which they are initialized.

Relation to steady-state error.--Some relationships of the time-dependent error to the steady-state error are given below:

a. $\frac{FSSE}{1-k^2} > FSSE$

$$b. \text{ FSSE} < \text{SSE}$$

$$c. \frac{\text{FSSE}}{1-k^2} < \text{SSE}$$

$$d. \frac{\text{ME}}{1-k^2} > \text{ME}$$

$$e. \frac{\text{FSSE}}{1-k^2} + \frac{\text{ME}}{1-k^2} > \text{SSE} + \text{ME}$$

when

$$\frac{\text{ME}}{\text{SSE}} > 2(1/k - 1).$$

The derivations are presented in Appendix B-9.

Summary

The phosphorus prediction model is relatively simple and its analysis is straightforward. The equation limits to the steady-state and hence the predictions asymptotically approach it. The predictions grow or decay depending upon whether the initial value is less or greater than the final value. By using the generator form of the equation, these characteristics can be determined without actually running the model and the programming needs are simplified. Also, the number of iterations required to attain a particular phosphorus concentration can be determined.

The error equation, though more complicated than the model, has several similarities to it. The errors approach a final equilibrium



value which is a fraction of the steady-state error and they grow or decay depending upon whether the initial error is smaller or greater than the final error. As with the model, these characteristics can be determined by using the generator form of the error equation.

Also, the number of iterations needed to attain a particular error can be calculated. Additionally, other features of possible interest can be calculated such as the steady-state error, the effect of removing the model error, and the coefficients of variation.

CHAPTER VI

COMPARISON OF FIRST-ORDER WITH MONTE CARLO

The Numerical Results

The 100 iterations per year base run phosphorus concentration prediction for the Monte Carlo simulation averaged .0182 mg/l over the 40 years. The error of the prediction is .0026 mg/l mean standard deviation which measures within-year variability. The coefficient of variation is about 14.3 percent. The inter-year standard deviation of the mean is .0005 mg/l and the correspondent coefficient of variation is 2.75 percent (see Table 6.1).

The base run phosphorus prediction concentration for the first-order analysis, using only the mean, averaged .0213 mg/l over the 40 iterations. It ranged as low as .0208 mg/l to .0213 mg/l. The mean standard deviation associated with this prediction which is an estimate of the combined uncertainties of all the model parameters is .0046. The error ranged from .0038 to .0046. The coefficient of variation ranged from .1827 to .2170 and its 40 year average is .2147. The inter-year standard deviation is .0001 and its correspondent coefficient of variation is .0053.

The two means for the two different base runs are substantially different: .0182 to .0213. Both the first-order error and the first-order coefficient of variation are larger than the respective Monte



TABLE 6.1.--Comparison of Numerical Results

Monte Carlo		First-Order
<u>Base Run</u>		<u>Base Run</u>
\bar{x}	.0182	.0213
\bar{s}	.0026	.0046
$\bar{c}\bar{v}$.1430	.2147
R	.9989	.99996
$S_{\bar{x}}$.0005	.0001132
$CV_{\bar{x}}$.0275	.0053
<u>All Normal</u>		<u>Monte Carlo Values Without Model Error</u>
\bar{x}	.0194	.0182
\bar{s}	.0016	.0002
$\bar{c}\bar{v}$.0840	.0117
R	.9940	.9998
$S_{\bar{x}}$.0021	.0000
$CV_{\bar{x}}$.0620	.0000
<u>500 Iterations</u>		<u>Without Model Error</u>
\bar{x}	.01872	.0213
\bar{s}	.0027	.0011
$\bar{c}\bar{v}$.1440	.0495
$R_{\bar{x}}$.1966	.9996
$S_{\bar{x}}$.00090	.0001132
$CV_{\bar{x}}$.0479	.0053
<u>1000 Iterations</u>		<u>Monte Carlo Values with Model Error</u>
\bar{x}	.01899	.0182
\bar{s}	.0027	.0047
$\bar{c}\bar{v}$.1430	.2582
$R_{\bar{x}}$.9963	1.0000
$S_{\bar{x}}$.00095	.0000
$CV_{\bar{x}}$.0499	.0000

Carlo values (see Table 6.1). For Monte Carlo the interval, mean plus or minus the standard deviation, is .0156 - .0218, whereas for first-order the interval is .0167 - .0259.

These differences in variability can be accounted for by considering the values used in calculating the predictions. The first-order analysis uses generally higher variances for v_s , q_s , L , and P_{i-1} . A comparison of these different means and standard deviations is given in Table 6.2.

In the Monte Carlo Simulation, the model error was not included as an independent term. Rather, it was incorporated in the parameter error, that is in the standard deviation of v_s . Hence, to facilitate a comparison, two modifications of the first-order program were made. First, the program was run without the independent model error term, but with the base run values. These results are compared in Table 6.1. Clearly, the removal of the model error significantly drops the first-order error from .0046 to .0011. (The prediction, of course, remains unchanged.) This error is substantially smaller than the Monte Carlo base run error of .0026.

The second refinement for comparison was to run the first-order program with all the values used for the Monte Carlo runs (Table 6.2), as well as without the independent model error term. (This includes use of the parameter error, that is, the standard deviation of v_s which incorporates the model error.) Again, a substantial drop occurs for both the prediction and the error. The prediction, .0182, is exactly the same (to within four decimal places) as

TABLE 6.2.--Comparison of Values

Monte Carlo Values		First-Order Values	
v_s :	$\bar{x} = 16.0367$	v_s :	$\bar{x} = 19.1910$
	$s = .4503$		$s = 1.1963$
	$CV = .0281$		$CV = .0623$
	$R = .9988$		$R = .9942$
q_s :	$\bar{x} = 12.0348$	q_s :	$\bar{x} = 10.6650$
	$s = .0460$		$s = 1.4608$
	$CV = .0038$		$CV = .1370$
	$R = .99998$		$R = .9725$
L :	$\bar{x} = .5108$	L :	$\bar{x} = .6352$
	$s = .0142$		$s = .0812$
	$CV = .0279$		$CV = .1279$
	$R = .9988$		$R = .97595$
τ :	$\bar{x} = 7.5950$	τ :	$\bar{x} = 7.9402$
	$s = 1.065$		$s = 1.0421$
	$CV = .1402$		$CV = .1312$
	$R = .9712$		$R = .9747$
P_{i-1} :	$\bar{x} = .0182$	P_{i-1} :	$\bar{x} = .0206$
	$x = .0005$		$s = .0027$
	$CV = .0275$		$CV = .1311$
	$R = .9989$		$R = .9748$

the Monte Carlo base run and quite close to the Monte Carlo all-normal prediction of .0194. The error decayed to about one eighth the Monte Carlo value and similarly the coefficient of variation was lowered to approximately 14 percent of the Monte Carlo value.

Montgomery et al. (1980), in a comparison of these same methods for the steady-state form of this model, found that both techniques gave similar prediction errors for similar input values.

Finally, increasing the iterations to 500 to 1,000 per year gave no significant differences in predictions or errors from the Monte Carlo base run. For example, the 500 iteration mean annual prediction is .01872 and its standard deviation .0027. Compare with base run values in Table 6.1.

The Procedures

Theoretical Considerations

The first-order analysis generates a series of predictions and errors which converge to their final values. The Monte Carlo method, on the other hand, generates a distribution of values from which predictions and standard deviations are derived for the phosphorus concentration.

One important question in comparing the two methods is whether or not the errors track the same central tendency and dispersion statistics. Scavia et al. (1981) in a similar comparison found that the two variances were indeed qualitatively different. In modeling various aquatic plants and animal species of Saginaw Bay, the first-order predictions generated means which tracked the typical species, whereas



the Monte Carlo method tracked the composite average of the total population. Consequently, the first-order variance represented the variability around the typical species while the Monte Carlo variance reflected the variability of the total population. Also, the first-order means tracked the Monte Carlo medians as the distributions became skewed (see also Burges and Lettenmaier, 1975). The Monte Carlo values, then, were qualitatively more realistic even though quantitatively they were generally similar to the first-order values.

In this study, however, there is only one subject modeled and, hence, both methods track the same variability, i.e., the variability of the phosphorus concentrations. Furthermore, the output distributions are not highly skewed allowing for no conclusion regarding whether or not the first-order mean is the Monte Carlo median. Both the means and the errors are qualitatively the same.

Another important issue in comparing the methods is whether their assumptions are violated. This is especially important for first-order analysis since its assumptions are stricter. It is assumed that the linearization adequately approximates the true mean within the region of interest. The non-linearity within this region must be sufficiently small to allow for such an estimation. Another way of stating this is to say that the method assumes the distributions are not highly non-normal and, hence, that the first and second moments adequately describe the distribution.

In a comparison of sensitivity analysis with Monte Carlo simulation, Gardner et al. (1981) found that for complex, non-linear



models, sensitivity analysis was a poor approximation because the region of interest displayed high variability. Sensitivity analysis is similar to first-order analysis in that both depend upon a linearization by use of a Taylor series expansion to approximate the true model. Unlike sensitivity analysis which examines small perturbations in the parameters by means of the partial derivatives, first-order analysis incorporates variances and covariances and therefore includes the effects of simultaneous parameter variability. Gardner et al. found this to be the major failure of sensitivity analysis to their stream ecosystem model. The variability of their parameter values ranged up to 3,000 percent, while in this study variability is no more than 30 percent and in most cases much less than that.

Since the predictions and errors in the linearization method converge to a final value, they can never generate bell-shaped curves and hence they are not normal distributions. Indeed, the predictions and the errors resemble a cumulative distribution function and in a sense, they are the accumulated summations of the previous calculations (see Appendix B-6). Because they asymptotically approach a final value, they suggest an asymptotic distribution such as the Type I extreme value or Gumbel (Gumbel, 1958) distribution. In this case, one wishes to know the limiting distribution of the largest of n values of predictions as n gets large. This implies that the largest of n observations is also the mean of the n observations. Such a distribution has been used before in modeling hydrologic phenomena such as peak annual flows (Benjamin and Cornell, 1970). This, however, requires further investigation.



Another point to be considered is whether or not the deviation from normality is sufficiently small such that a normal distribution approximates the true distribution. The coefficients of variation in the first-order results are generally lower than the Monte Carlo values. One measure of non-normality is the mode to mean ratio given by,

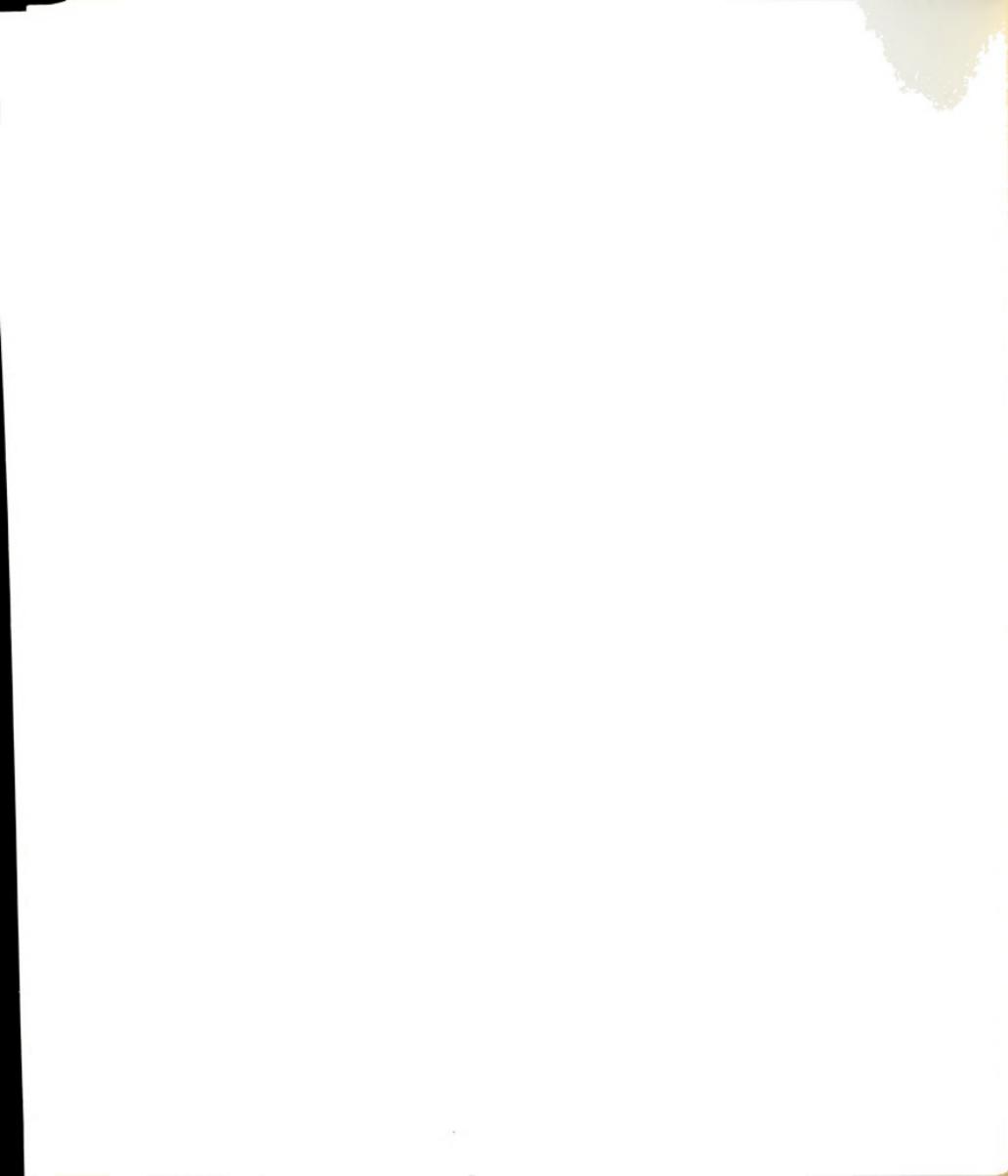
$$R = (1 + CV)^{-1.5}$$

where,

CV = coefficient of variation.

Most of the R values for the first-order analysis were 99 percent or more of normal by this measure.

Another point of comparison lies in the different sensitivities of the two methods. The Monte Carlo procedure is sensitive to the change in distributions, and possibly to the change in the size of the random number seed. The method is not as sensitive to changes in point source flux variability as present information indicates (Lee, 1980), but it does respond to the upgrading of all treatment levels and especially the switch in concentration values for the Lake Erie flux which is consistent with the steady-state results (Montgomery et al., 1980) obtained for a similar test. Its sensitivity to land use changes is restricted primarily to differences between disturbed and undisturbed uses which is all that is reflected in the data. The method is not sensitive to increased iterations per year.



First-order analysis, on the other hand, is highly sensitive to the removal of the model error as well as the changes in the means and variances as the switch from base run to Monte Carlo values indicates.

Further study of this model with the Monte Carlo procedure should include three main points. First, the reason for the increase in average annual prediction concentration, and in the inter-year variability when all normal distributions are used is not clear. Are these increases due to the inadequacy of the normal distribution to characterize the parameters or are they primarily due to the manner in which the variates are chosen? Closely related to this and in need of clarification is the reason for the apparent possible sensitivity of the all normal run to changes in the random number seed. Future research should focus on whether or not these results are consistently obtained, and if so, why. Second, the relative insensitivity of the model to changes in point source variability requires explanation. Third, the model error should be incorporated as an independent error term.

An additional consideration for further research is the problem of periodicity of the 500 and especially 1000 iterations per year runs. The reason for this is unknown.

For the first-order error analysis, future research should focus on two main areas. First, the proper output distribution needs to be determined and its impact on the assumptions of first-order clarified. Second, higher order terms of the Taylor series expansion



should be examined to determine whether there is a significant change in error estimates.

Secondarily, it would be interesting to examine more closely the behavior of the coefficient of variation.

Practical Considerations

Generally, the initial investigator time for first-order analysis involves the derivation and verification of the partial derivatives, a process which can be laborious and lengthy, depending upon model complexity. In some cases there may be no analytical solution to the partial derivatives and one can only approximate the solution through finite differences (Garen, 1979; Coleman and DeCoursey, 1976; and Mein and Brown, 1978).

With the Monte Carlo method, the amount of time required for the initial investigation depends upon the difficulty of properly characterizing the probability density functions. The more easily these can be described, the less investigator time is required. The computation costs, however, especially on large frame computers, can be quite high, depending upon the complexity of the model.

Frequently, as in this particular study, the probability density functions are easily described for the Monte Carlo method. The major cost was computational. In the first-order analysis, the derivation and verification of the partial derivatives was also not difficult, but the programming was much simpler and hence the computation costs were much lower. It should be noted, though, that the Monte Carlo method was significantly more complex in its treatment of

the loading term L which contributed to its larger cost of computation. Nonetheless, the predictions and error estimates of the first-order method can be run on a pocket calculator which is more available and less expensive than is a computer.

Summary

The comparison of these two procedures on this model indicates that the predictions are generally similar but that the errors can differ substantially. Although the Monte Carlo all-normal mean annual error (intra-year variability) is smaller than its base run counterpart, it is significantly larger than the corresponding first-order value even when the same data are used. The exact reason for this is unclear. The first order predictions and errors have non-normal distributions whose exact characterizations require additional investigation. However, it was found that both errors are qualitatively the same.

The comparison also indicates some practical differences in the two methods. The first-order prediction and error equations, for example, can be easily programmed into a calculator which reduces computation cost and may increase its utility. The Monte Carlo procedure, however, requires less initial investigator time since the distributions are easily characterized.

The modeler should remember, then, that the first order method replaces a non-linear model with a linear approximation. This approximation is good only within the region of low variability where the function is not highly non-linear. The philosophy behind this approach



is the belief that an approximation of the whole is better than a detailed account of only a part (Cornell, 1969). Detail on one level is necessarily sacrificed for generality at another. A highly refined, complicated model may not be a good candidate for this technique. For example, Gardner et al. (1981) chose a very complex, non-linear model which required the numerical approximation of 48 sensitivity equations. Most likely this is not susceptible of first-order analysis. The Monte Carlo method, though, has no such restrictions and treats the non-linear model as it is. However, it is critical that the distributions be correctly characterized. In the face of inadequate data, this is not always easy.

FINAL COMMENTS

With the increasing dependency upon quantitative models, the need for estimating the uncertainty or reliability of model predictions will also grow. It is hoped therefore that research in error estimation methods will continue and that this study will have contributed in some measure to this body of knowledge.

However, it is also hoped that as resource managers make greater use of mathematical models, that the broader, non-quantitative aspects of the management decision-making process not be overlooked. It is all too easy in this age of specialization and narrow quantification to neglect issues which require for their solution philosophical insight, rather than numerical approximations. In the context of manifold, complex questions that the water resource manager faces today, it must be remembered that modeling, though useful, is no substitute for sound judgment, and next to a perspicacious understanding of the larger social and scientific issues, must always remain ancillary.

APPENDICES

APPENDIX A

MONTE CARLO APPENDICES

APPENDIX A

MONTE CARLO APPENDICES

A.1 Correlation Coefficients

Correlation coefficients are used to measure the strength of the relationship between two variables. In linear regression analysis, it indicates the goodness-of-fit of the regression line to the data. The correlation coefficient used in linear regression is the Pearson Product-Moment Correlation Coefficient and is defined as:

$$r = \frac{\sum_{i=1}^n X_i Y_i - \left(\frac{\sum_{i=1}^n X_i}{n} \right) \left(\frac{\sum_{i=1}^n Y_i}{n} \right)}{\left\{ \left[\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \right] \left[\sum_{i=1}^n Y_i^2 - \frac{\left(\sum_{i=1}^n Y_i \right)^2}{n} \right] \right\}^{1/2}}$$

In correlating the annual mean flows for the Monte Carlo simulation data were taken from the calendar years 1935-1969 for the Genesee, Oswego, Black, Conaseraga, Fish and Moose Rivers. The correlation coefficients are presented below.

Pearson Product-Moment Correlation Coefficients					
	Moose	Fish	Conaseraga	Black	Oswego
Genesee	.51	.69	.90	.71	.82
Oswego	.60	.74	.84	.73	
Black	.79	.95	.62		
Conaseraga	.44	.62			
Fish	.72				



A.2 Markov Flow Model

Within the Monte Carlo simulation, a Markov flow model is used to generate the historical flow pattern for the Niagara River/Welland Canal flow. Such a model has both a deterministic and a random component. It assumes that the current flow depends on the previous year's flow plus a stochastic term. The equation is:

$$q_i = \bar{x} (1 - r_1) + r_1 q_{i-1} + t_i \cdot s \cdot \sqrt{(1 - r_1^2)}$$

where: q_i = i th year's flow

\bar{x} = historical mean

r_1 = lag-one auto correlation coefficient

s = standard deviation

t_i = i th standard normal random variant.

It is assumed that the flows are normally distributed. Data were taken from surveys of 23 calendar years of flow data (Chapra, 1981).

A statistical analysis of the flows yields the following results:

$$\text{mean} = 1.961 \cdot 10^{11} \text{ m}^3/\text{yr}$$

$$\text{SS} = 1.051 \cdot 10^{10} \text{ m}^3/\text{yr}$$

The lag-one autocorrelation coefficient is $r_1 = .7998$. Hence, the equation is:

$$q_i = 1.961 \cdot 10^9 (1 - .7998) + .7997 q_{i-1} + t_i \cdot 1.051 \cdot 10^8 \cdot$$

$$\sqrt{(1 - .7998^2)} .$$



Table A.1: Statistical Summary of 100 Iteration Base Run - Model Terms

Phosphorus Concentration Prediction (mg/l)	\bar{x}	= .0182	\overline{CV}	= .1430
	$s_{\bar{x}}$	= .0005	\bar{s}	= .0026
	$CV_{\bar{x}}$	= .0275	y-intercept	= .0186
	$R_{\bar{x}}$	= .9989	slope	= .0194
			r	= -.4586
Settling Velocity Prediction (m/yr.)	\bar{x}	= 16.0367	y-intercept	= 16.1854
	$s_{\bar{x}}$	= .4503	slope	= -.0073
	$CV_{\bar{x}}$	= .0281	r	= -.1884
	$R_{\bar{x}}$	= .9988		
Areal Water Load Prediction (m/yr.)	\bar{x}	= 12.0348	y-intercept	= 12.0414
	$s_{\bar{x}}$	= .0460	slope	= -.0003
	$CV_{\bar{x}}$	= .0038	r	= -.0783
	$R_{\bar{x}}$	= 1.0000		
Phosphorus Loading Prediction (g/m ² -yr.)	\bar{x}	= .5108	y-intercept	= .5247
	$s_{\bar{x}}$	= .0142	slope	= -.0007
	$CV_{\bar{x}}$	= .0279	r	= -.5594
	$R_{\bar{x}}$	= .9988		

Source: Lee (1980)



Table A.2: Statistical Summary of Base Run Fluxes (g/yr)*

Lake Erie Flux	$\bar{x} = 4.064 \times 10^9$	$CV_{\bar{x}} = .0108$
	$S_{\bar{x}} = .044 \times 10^9$	$R_{\bar{x}} = .9998$
Diffuse Source Flux	$\bar{x} = 3.161 \times 10^9$	$CV_{\bar{x}} = .0882$
	$S_{\bar{x}} = .276 \times 10^9$	$R_{\bar{x}} = .9885$
Point Source Flux	$\bar{x} = 1.703 \times 10^9$	$CV_{\bar{x}} = .0252$
	$S_{\bar{x}} = .043 \times 10^9$	$R_{\bar{x}} = .9990$
Atmospheric Flux	$\bar{x} = 7.762 \times 10^9$	$CV_{\bar{x}} = .0370$
	$S_{\bar{x}} = .287 \times 10^8$	$R_{\bar{x}} = .9980$
Total Flux	$\bar{x} = 9.702 \times 10^9$	$CV_{\bar{x}} = .0278$
	$S_{\bar{x}} = .268 \times 10^9$	$R_{\bar{x}} = .9989$

Source: Lee (1980)

*Table corresponding to p. 23.



Table A.3: Statistical Summary of Base Run Water Load (m³/yr.)*

Atmospheric Water Loading	$\bar{x} = 1.659 \times 10^{10}$	$CV_{\bar{x}} = .0199$
	$S_{\bar{x}} = .033 \times 10^{10}$	$R_{\bar{x}} = .9994$
Water Load From Lake Ontario Basin Tributaries	$\bar{x} = 2.721 \times 10^{10}$	$CV_{\bar{x}} = .0206$
	$S_{\bar{x}} = .056 \times 10^{10}$	$R_{\bar{x}} = .9994$
Water Load From Lake Erie	$\bar{x} = 1.849 \times 10^{11}$	$CV_{\bar{x}} = .0027$
	$S_{\bar{x}} = .005 \times 10^{11}$	$R_{\bar{x}} = 1.000$
Total Water Load	$\bar{x} = 2.287 \times 10^{11}$	$CV_{\bar{x}} = .0039$
	$S_{\bar{x}} = .009 \times 10^{11}$	$R_{\bar{x}} = 1.0000$

Source: Lee (1980)

*Table corresponding to p. 23.



Table A.4: Statistical Summary of Higher Iterations

Number of Iterations				
500	\bar{x}	= .01872	$\bar{C}\bar{V}$	= .1440
	$S_{\bar{x}}$	= .00090	\bar{s}	= .0027
	$CV_{\bar{x}}$	= .0479	y-intercept	= .01878
	$R_{\bar{x}}$	= .9966	slope	= -.000003
			r	= -.0414
1000	\bar{x}	= .01899	$\bar{C}\bar{V}$	= .1430
	$S_{\bar{x}}$	= .00095	\bar{s}	= .0027
	$CV_{\bar{x}}$	= .0499	y-intercept	= .0186
	$R_{\bar{x}}$	= .9963	slope	= .00002
			r	= .2282

Source: Lee (1980)



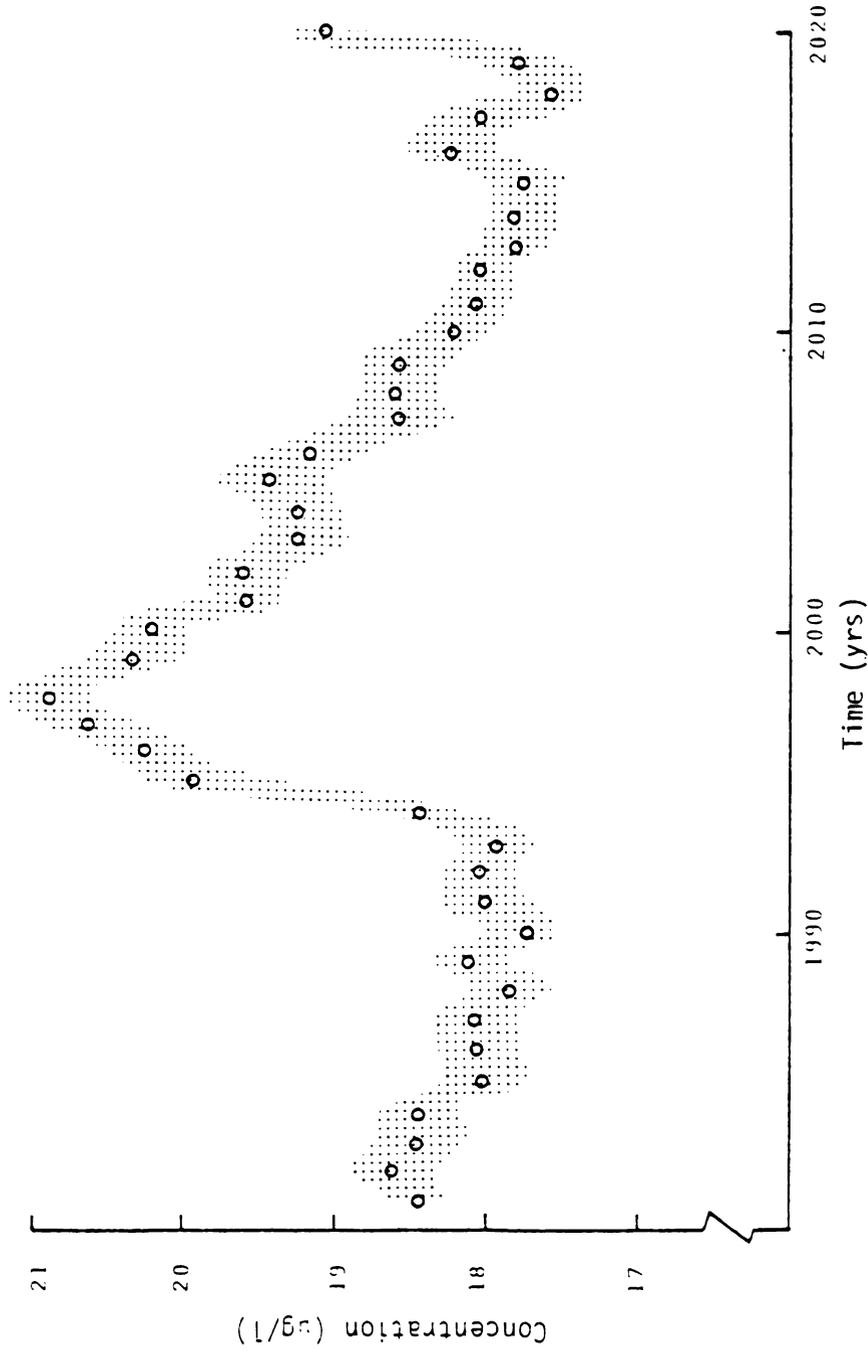


Figure A.1.-40 Years of Phosphorus Concentration Predictions at 500 Iterations Per Year.

Source: Lee (1980)



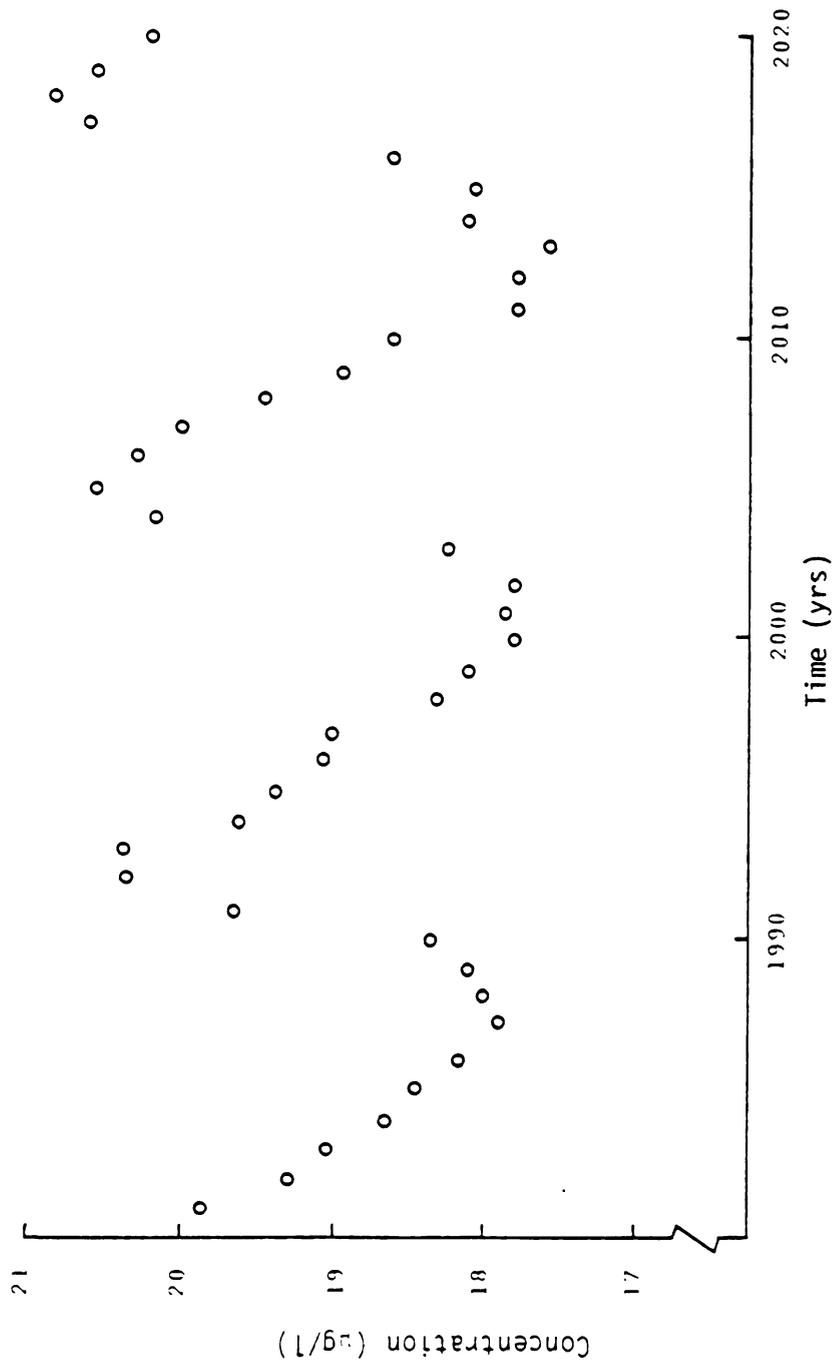


Figure A.2.-- Concentration Predictions Using Larger Random Seed For 1000 Iterations Per Year.

Source: Lee (1980)

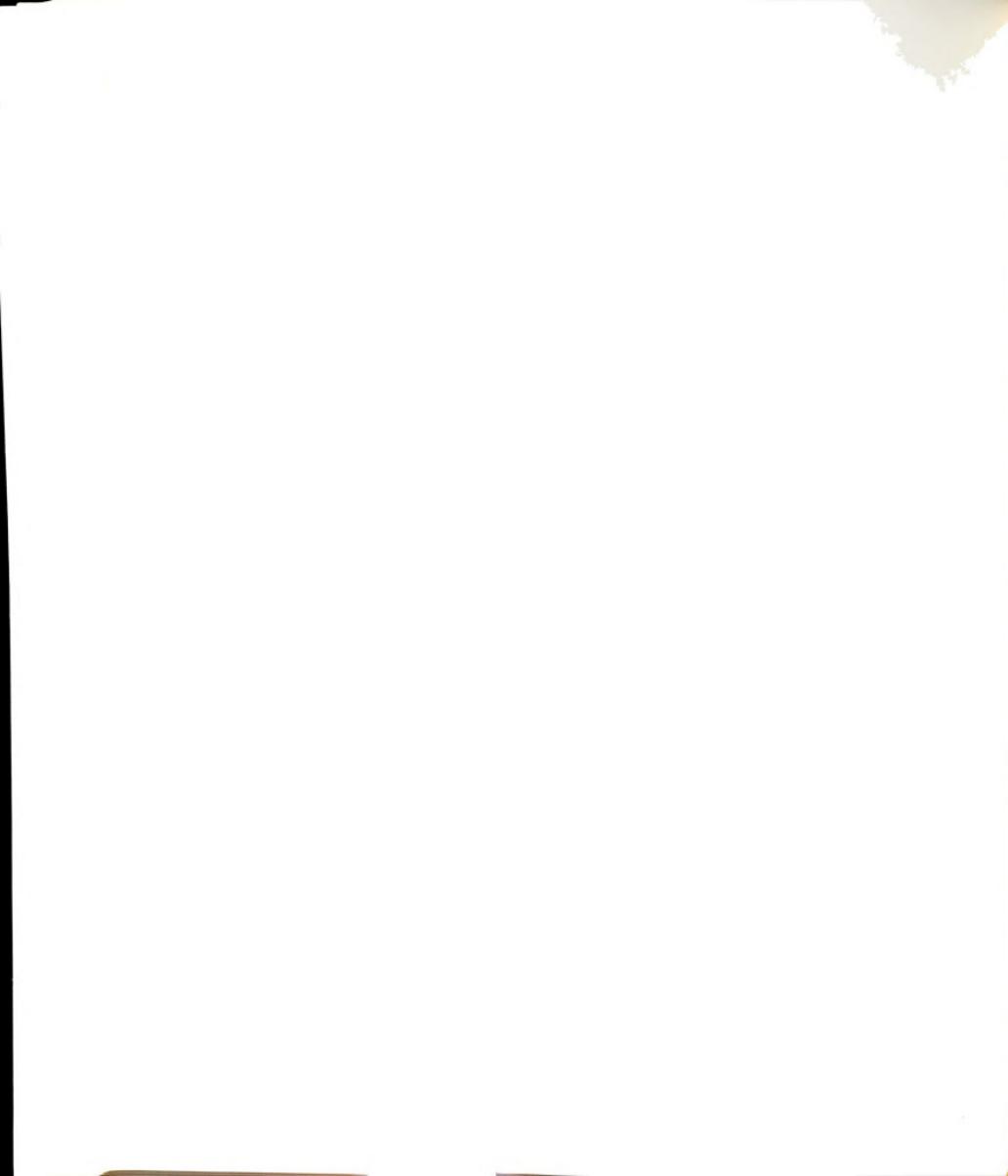


Table A.5: Statistical Summary of Flux Sensitivities--Phosphorus Concentration Predictions (g/m³)

Cut Point Source Flux Variability by Half	\bar{x} = .0182	\overline{CV} = .0143
	$s_{\bar{x}}$ = .0005	\bar{s} = .0026
	$CV_{\bar{x}}$ = .027	y-intercept = .0186
	$R_{\bar{x}}$ = .999	slope = -.0002
		r = -.4339
Upgrade All Point Source Treatment to Phosphorus Removal	\bar{x} = .01736	\overline{CV} = .1494
	$s_{\bar{x}}$ = .00051	\bar{s} = .0026
	$CV_{\bar{x}}$ = .02911	y-intercept = .01779
	$R_{\bar{x}}$ = .9987	slope = -.00002
		r = -.4808
Use Eastern Erie Concentration Value for Calculation of Flux From Lake Erie	\bar{x} = .0165	\overline{CV} = .1636
	$s_{\bar{x}}$ = .0005	\bar{s} = .0027
	$CV_{\bar{x}}$ = .0298	y-intercept = .0169
	$R_{\bar{x}}$ = .9987	slope = -.00002
		r = -.4383

Source: Lee (1980)

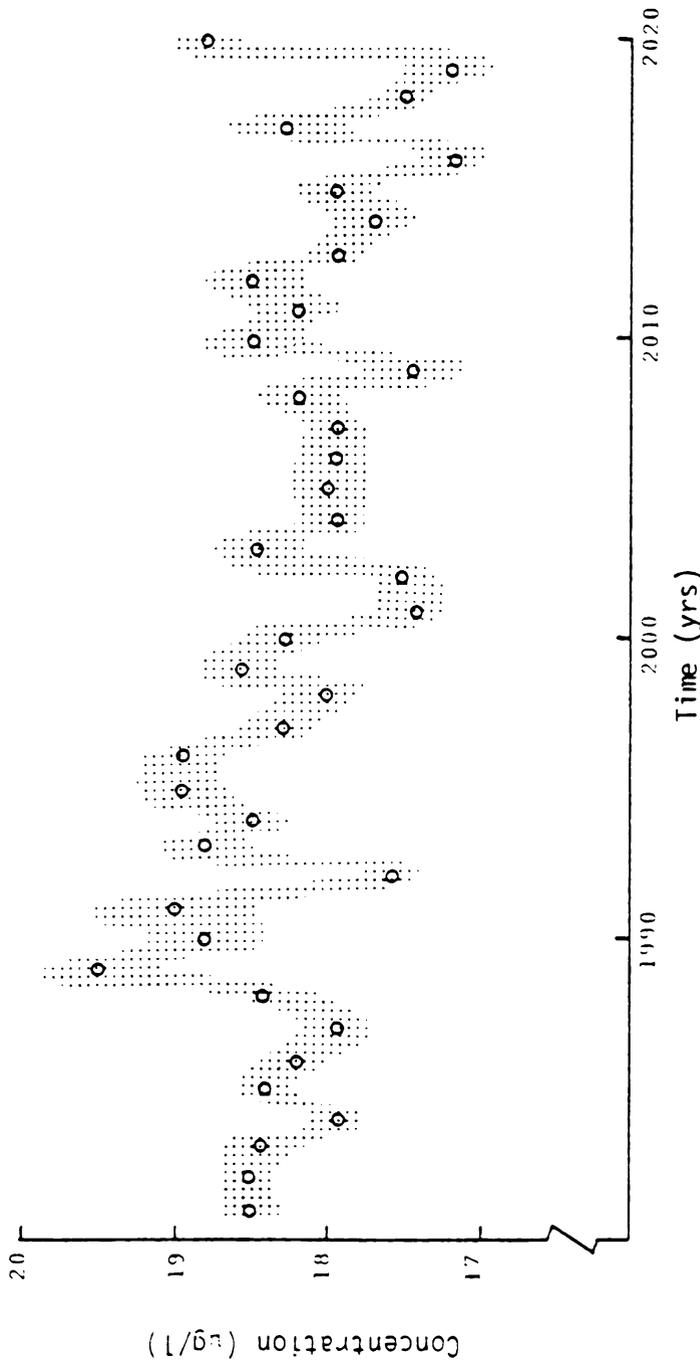


Figure A. 3. --40 Years of Phosphorus Concentration Predictions Using 1/2 Variability for Point Source Flux.

Source: Lee (1980)

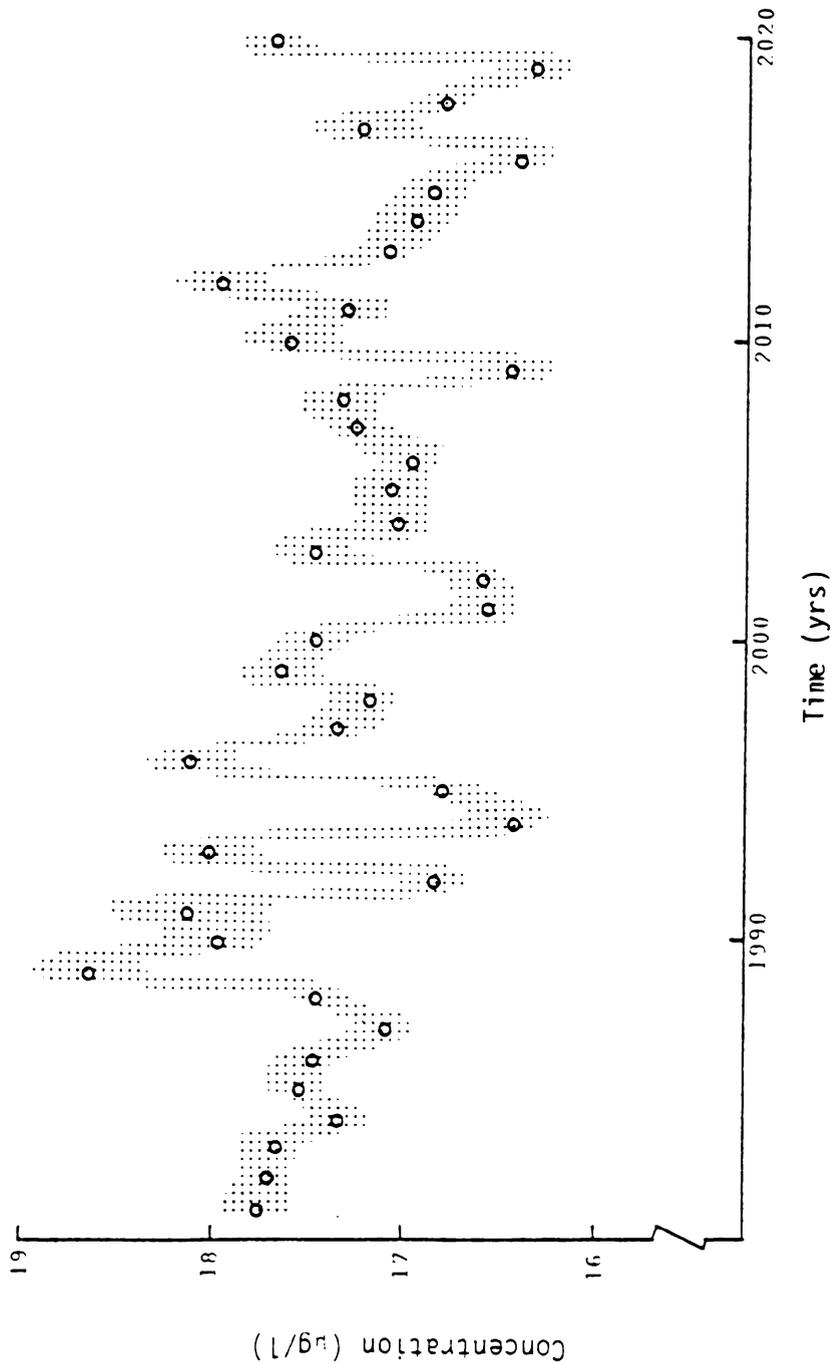


Figure A.4.--40 Years of Phosphorus Concentration Predictions with all Phosphorus Removal for Point Flux.

Source: Lee (1980)

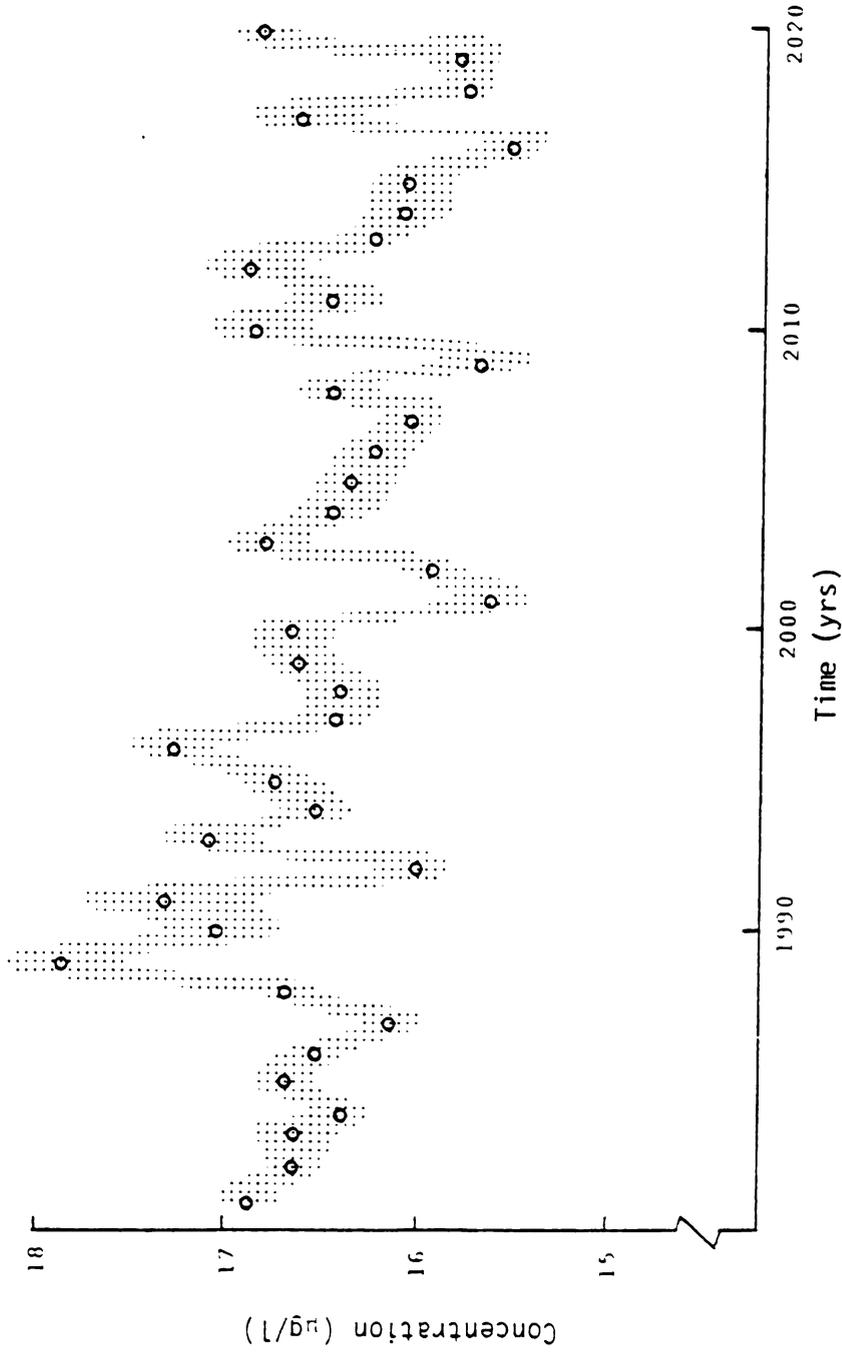


Figure A.5:---40 Years of Phosphorus Concentration Predictions Using Eastern Erie Concentration Data.

Source: Lee (1980).



Table A.6: Statistical Summary of Land Use Shifts

Constant Rate Shift to All Urban	\bar{x} = .02199	$\bar{C}\bar{V}$	= .1363
	$S_{\bar{x}}$ = .00128	\bar{s}	= .00301
	$CV_{\bar{x}}$ = .0582	y-intercept	= .02025
	$R_{\bar{x}}$ = .9949	slope	= .00008
		r	= .77249
Constant Rate Shift to All Agriculture	\bar{x} = .02257	$\bar{C}\bar{V}$	= .2244
	$S_{\bar{x}}$ = .00257	\bar{s}	= .00519
	$CV_{\bar{x}}$ = .1139	y-intercept	= .01833
	$R_{\bar{x}}$ = .9809	slope	= .00021
		r	= .9399
Constant Rate Shift to All Forest	\bar{x} = .01577	$\bar{C}\bar{V}$	= .1039
	$S_{\bar{x}}$ = .00163	\bar{s}	= .00162
	$CV_{\bar{x}}$ = .1034	y-intercept	= .01858
	$R_{\bar{x}}$ = .9842	slope	= -.00014
		r	= -.98295



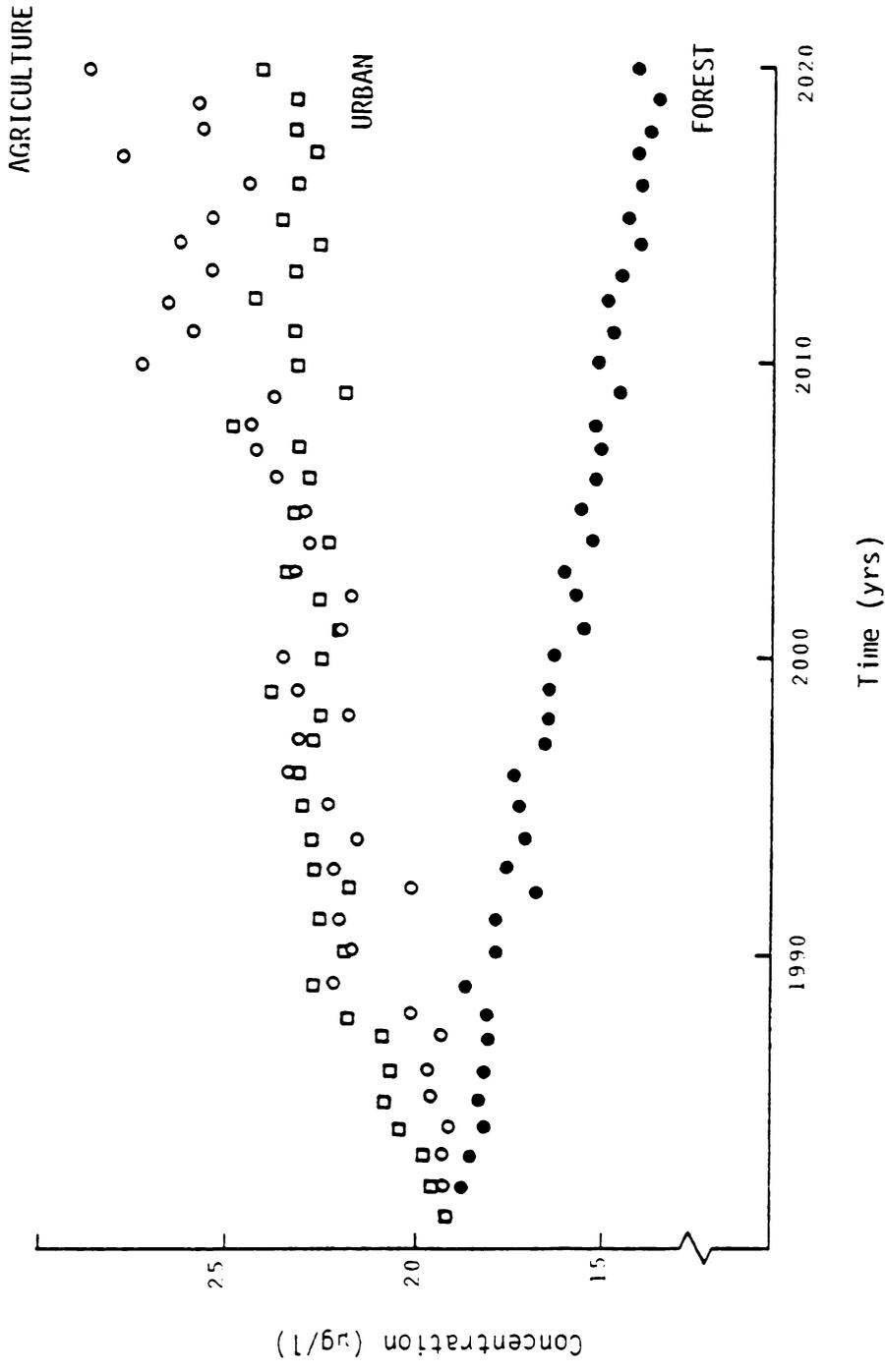


Figure A.6.-- 40 yrs of Phosphorus Concentration Predictions: Land Use Shifts to All One Type.

Source: Lee (1980)

A.3 Calculator Program

Since programmable calculators are becoming more widespread, it was deemed appropriate to write a program using the Monte Carlo method for calculators. The program was written for a Texas Instruments model 58C using the master library module. (This does not necessarily constitute an endorsement of this product.) Generally, the calculator program is similar to the computer program used by Lee (1980). There are some relatively minor differences, though (see Reckhow et al., 1981, Appendix K, for a discussion of these differences). It was found that, although acceptable predictions are obtained, the time required for the computation of one prediction is too long to be practical (one to one-and-a-half minutes). On the other hand, the monetary cost is negligible. Nevertheless, this illustrates the capacity of calculators to handle relatively sophisticated programs and their potential use for a wide variety of modeling problems.

The program is presented here for the interested reader. The subroutines are considered first; then the main program is outlined and explained using calculator keyboard notation.



The Subroutines

Subroutine A (written as SBR A) generates a standard normal random variate and is the random number generator for this program. It used the uniform random number generator from the master library module to generate two uniform random numbers to be used according to the following equation:

$$R = \sqrt{-2 \cdot \ln \mu_1} \cdot \text{Cos} (\mu_2 2 \pi)$$

Where: R = the standard normal random number

μ_1 = the first generated uniform random number on the interval (0,1).

μ_2 = the second generated uniform random number on the interval (0,1).

The subroutine is written as:

2nd LBL A -	- identifies subroutine
((2nd Pgm 15 SBR 2nd D.MS	- uses program 15's uniform random number generator
ln	- takes the natural log of the uniform random number
x 2 +/-)	- multiplies previous value by -2
\sqrt{x})	- takes the square root of the previous value
x (2nd Pgm. 15 SBR 2nd D.MS x 2 x π)	- multiplies second uniform random number 2 π

2nd Rad	- places calculator in radian mode
2nd Cos	- takes the cosine of the value
Inv SBR	- returns program pointer to main program

Subroutine B (SBR B) randomly selects from the Ontario flow distribution (Ontflo) by using SBR A. The random number is multiplied by the standard deviation and the mean is then added. This parameter is then tested to see if it is negative. If it is negative, the program pointer returns to the beginning of the subroutine and starts again; if not, it returns to the main program.

The subroutine is written as:

2nd LBL B	- identified subroutine
(SBR A x 6.24EE9 + 2.75 EE10	- Random number is multiplied by standard deviation and the mean is added
Inv 2nd x ≥ t 309) Inv SBR	- tests parameter sign; if negative it recalculates by going to the start, step 309; otherwise it returns to main program

Subroutines C - E are treated in the same manner; and are written as:

2nd C BL C	
(SBRA x 1.23 EE9 + 3.35 EE9 Inv 2nd x ≥ t <u>335</u>) Inv SBR	
2nd L BL D	
(SBRA x .008 + .032 Inv 2nd x ≥ t <u>360</u> Inv SBR	
2nd L BL E	
(SBRA x 4.67 + 16.02 Inv 2nd x ≥ t <u>381</u> SBR	



THE MAIN PROGRAM

<u>Program Step</u>	<u>Function</u>
2nd Op 20	- Increments counter (R_{00} by 1
(((SBR A x .8 - 1.02 Inv ln	- Randomly selects a log normal urban concentration value (Urbcon)
x (.063 + Rc1 x Rc1 00)	- Multiplies the Urbcon value by the changed land use fraction
= SBR A x .62 - 1.91) Inv ln	- Adds a randomly selected log nor- mal agricultural concentration value (Agrcon)
x (.303 + Rc1 02 x Rc1 00)	- Multiplies Agrcon value by the changed land use fraction
+ (SBR x .2 - 4.27 Inv ln	- Adds a randomly selected log nor- mal forest concentration value (Forcon)
x (.634 + Rc1 03 x Rc1 00)	- Multiplies Forcon value by the changed land use fraction
x SBR B sto 04	- Randomly selects a normal flow value for Lake Ontario and stores it in data register 04 (Ontflo)
+ ((SBR Z + .766 - 2.33) Inv ln x .1	- Randomly selects a log normal urban atmospheric flux value and multiplies it by its land use weighting
+ (SBR A x 2.02 - 3.2) Inv ln x .6	- Randomly selects a log normal forest atmospheric flux value and multiplies it by its land weighting

-----Loadings-----

-----Diffuse Flux-----

-----Atmospheric Flux-----



1st lag --Factor--	Erie Markov --Flux--Model	Point Flux	Atmospheric Flux	Program Step	Function
				+ (SBR A x .773 - 3.356) Inv ln x .3	- Randomly selects a log normal agricultural atmospheric flux value and multiplies it by its land use weighting
				x 1.9 EE10	- Multiplies the sum of the atmospheric flux values by the lake surface area (LSA)
				+ SBR C	- Adds a randomly selected normal value for point flux
				+ (SBR A x 1.046 EE10 + .7998 x Rcl 05 + 3.704 EE10) sto 05	- Markov flow model value is stored in data register 05
				x SBR D)	- Randomly selects a normal concentration value from Lake Erie (Niagra-Welland) (Erecon)
				÷ (SBR E sto 06	- Divides L by the settling velocity, v_s , which is randomly selected from a normal distribution and stored in register 06
				x 1.9 EE 10	- Multiplies v_s by LSA to assure proper units
				+ Rcl 04	- Adds <u>Ontflo</u> from R_{05}
				+Rcl 05	- Adds <u>Ereflo</u> from R_{05}
				+ 1.7 EE 10)	- Adds atmospheric load
				x (1 - Rcl 06 ÷ 89 +/- - .132) Inv ln	- Divides v_s by mean depth, z, and subtracts the inverse of the mean hydraulic retention time, τ . This value is used as a power of "e" and subtracted from 1.



	<u>Program Step</u>	<u>Function</u>
2nd lag --Factor--	+ Rc1 08 x (Rc1 06 ÷ 89 +/- -.132) Inv ln	- The same power of "e" is used here as above but is multiplied by the previous prediction and is added to the first term.
Phosphorus --Prediction--	= sto 08 R/S RST	- Stores the phosphorus prediction in R ₀₈ and stops. (Pressing R/S (Run/Stop) starts program with RST (reset) and program pointer returns to the beginning.)

APPENDIX B

FIRST-ORDER ERROR APPENDICES

APPENDIX B

FIRST-ORDER ERROR APPENDICES

B.1 Calculation of the Parameter and Model Errors

The variances for the parameter, v_s , and the model were established using the computer facilities of Michigan State University. The program is found in the SPSS 6000 Supplement Computer Laboratories Michigan State University (1978). The routine is taken from a package from Northwestern University's Vogelback Computing Center which is itself based on a package originally programmed at the Madison Academic Computing Center of the University of Wisconsin.

The program offers two methods of non-linear regression following the presentation of Draper and Smith (1966): The Marquardt and Gauss techniques. The latter technique was used here. This method requires an initial estimate of the parameters; then a Taylor's series expansion of the function is written near the initial value, ignoring higher order terms. This linearizes the function in the parameters. The initial value chosen was the arithmetic mean of 14 calculations based on 14 years of data for L , q_s , and P using the equation,

$$v_s = \frac{L}{P} - q_s.$$

The routine fits the parameters by a sum-of-squares function such that the error is minimized. The final answer is the one with the

smallest error. An exploration routine was used to examine the neighborhood of the final answer to insure that it was not just a local peak.

Assuming the function is linear in its parameters at their final values, the final answer with upper and lower limits approximating 95% confidence limits is given:

$$b_k \pm 2e_k$$

where

$$e_k = \sqrt{\frac{a_{ii} \cdot S(b)}{\text{NCASE-NB}}} \quad \text{where}$$

$S(b)$ = sum-of-squares

NCASE = number of cases

NB = number of parameters

a_{ii} = matrix Coefficients

The value e_k , of the k th parameter, is a function of the sum-of-squares scaled by the normalizing matrix elements and divided by the degrees of freedom. This value is used as an approximation of the standard deviation, and hence the variance, of v_s . The values are:

Lower Limit	Final Value	Upper Limit
16.798489	19.191036	21.583584

or about

$$19.1910 \pm 2(1.1963)$$

where the standard deviation of v_s is approximately 1.1963.



This routine also gives an error estimate for the model which is the mean square residual. This is similar to the equation above for e_k but without the matrix elements a_{ij} . For this model the estimated value is about .0032 which is used as the standard deviation of the model.

B.2 Alternate Form of the Prediction Equation

We need to show that the model is equivalent to:

$$P_i = SS + (P_0 - SS)k^i$$

where: $k = 1/e^{v_s/z} + 1/\tau$

$SS = L/(v_s + q_s)$, the steady-state

$P_0 =$ Initial value

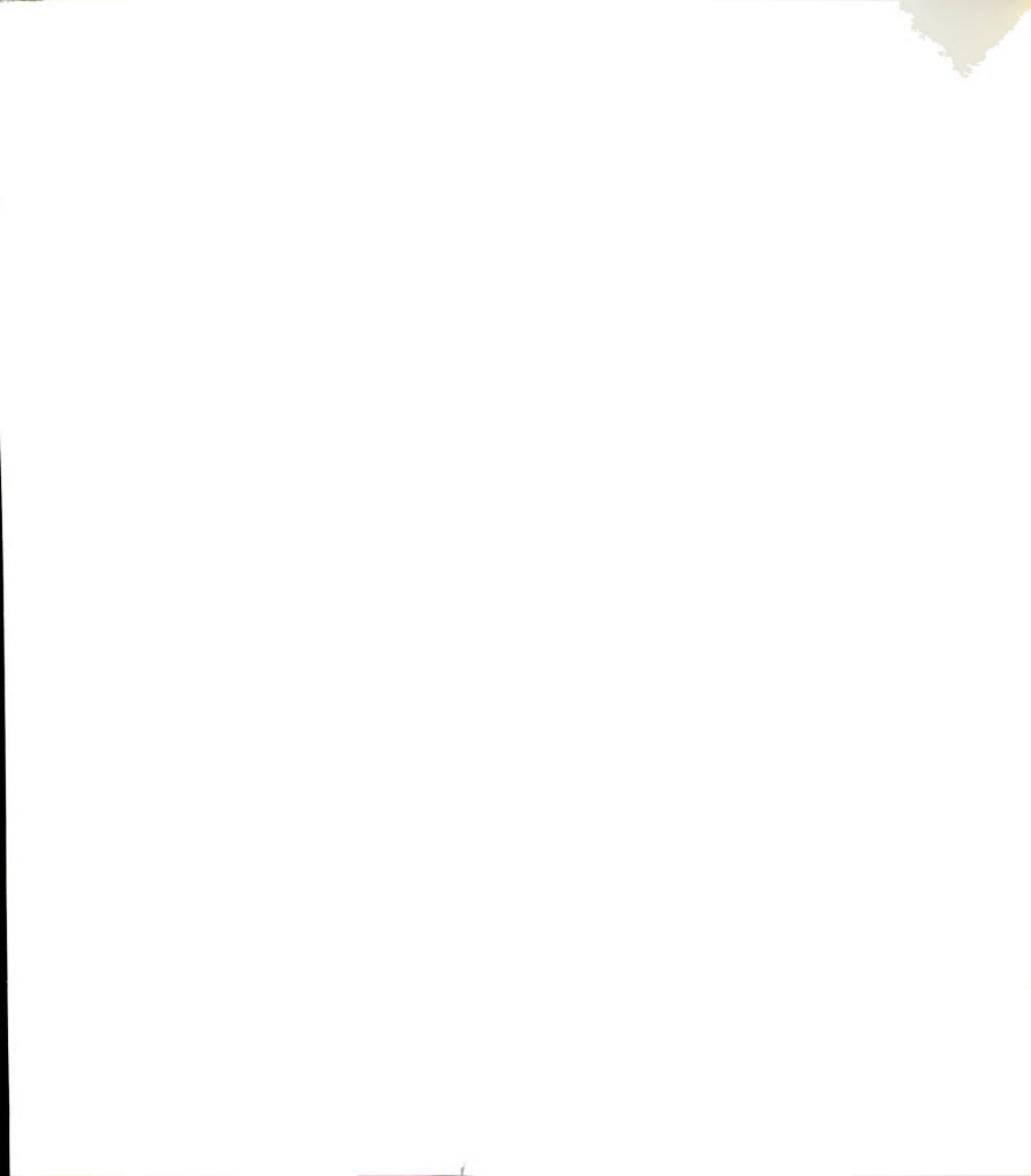
$i =$ number of iterations.

The model is:

$$\begin{aligned} P_i &= \frac{L}{v_s + q_s} (1-k) + P_{i-1}(1-k) \\ &= SS(1-k) + P_{i-1}(k) \\ &= SS - SSk + P_{i-1}(k) \\ &= SS + (P_{i-1} - SS)k. \end{aligned}$$

We observe the following sequences:

$$\begin{aligned} P_1 &= SS + (P_0 - SS)k \\ P_2 &= SS + (P_1 - SS)k \\ &= SS + (SS + (P_0 - SS)k - SS)k \\ &= SS + (P_0 - SS)k^2 \end{aligned}$$



$$\begin{aligned}
 P_3 &= SS + (P_2 - SS)k \\
 &= SS + (SS + (P_0 - SS)k^2 - SS)k \\
 &= SS + (P_0 - SS)k^3
 \end{aligned}$$

.

.

.

$$P_i = SS + (P_0 - SS)k^i.$$

B.3 The Limit of the Prediction Equation

We wish to show that the time-dependent model converges to the steady-state. We have:

$$P_i = SS + (P_0 - SS)k^i.$$

Then:

$$\begin{aligned}
 \text{Limit}_{i \rightarrow \infty} P_i &= \text{Limit}_{i \rightarrow \infty} (SS + (P_0 - SS) k^i) \\
 &= SS + (P_0 - SS) \text{Limit}_{i \rightarrow \infty} k^i.
 \end{aligned}$$

Because both P_0 and SS are constants and because $0 < k < 1$ the

$\text{Limit}_{i \rightarrow \infty} k^i = 0$ and hence

$\text{Limit}_{i \rightarrow \infty} P_i = SS.$



B.4 The Solution for the i th Iterations

We wish to know how many iterations, i , are required to reach convergence or to be within a certain distance of convergence.

Since:

$$P_i = SS + (P_o - SS)k^i$$

then:

$$P_i - SS = (P_o - SS)k^i \quad \text{and}$$

$$\begin{aligned} \frac{P_i - SS}{P_o - SS} &= k^i \\ &= (1/e^{v_s/z + 1/\tau})^i \end{aligned}$$

$$\ln \left(\frac{P_i - SS}{P_o - SS} \right) = i \cdot -(v_s/z + 1/\tau)$$

B.5 The Error Equation

In order to see how the error equation behaves, it is helpful to rewrite it in a more tractable form.

Let $k = 1/e^{v_s/z + 1/\tau}$, then



$$\begin{aligned}
S_t^2 &= \left[(k-1)^2 \left(\frac{L}{(v_s+q_s)^2} \right)^2 + 2(k-1) \frac{L}{(v_s+q_s)^2} \left(P_{i-1} - \frac{L}{v_s+q_s} \right) (-1/z) \right. \\
&\quad \left. \cdot (k) + \left(P_{i-1} - \frac{L}{v_s+q_s} \right)^2 (-1/z)^2 k^2 \right] S^2(v_s) \\
&+ \left[\left(\frac{L}{(v_s+q_s)^2} \right) (k-1)^2 \right] S^2(q_s) + \left[\left(\frac{1}{(v_s+q_s)} \right)^2 (1-k)^2 \right] S^2(L) \\
&+ \left[\left(P_{i-1} - \frac{L}{(v_s+q_s)^2} \right)^2 \tau^{-4} k^2 \right] S^2(\tau) + k^2 \cdot S^2(P_{i-1}) \\
&+ \frac{2 \cdot L \cdot (k-1)(1-k)}{(v_s+q_s)^2} S(q_s) \cdot S(L) \cdot r(q_s, L) \\
&+ 2 \left(P_{i-1} - \frac{L}{v_s+q_s} \right)^2 \tau^{-2} \cdot k \cdot \frac{L}{(v_s+q_s)^2} (k-1) S(\tau) S(q_s) r(\tau, q_s) \\
&+ \frac{2(1-k)}{v_s+q_s} \left(P_{i-1} - \frac{L}{v_s+q_s} \right) \tau^{-2} k S(L) S(\tau) r(L, \tau) \\
&+ \text{ME}
\end{aligned}$$

where ME = model error.

After combining terms with common factors, the following arrangement is noted,

1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900

$$\begin{aligned}
S_t^2 &= (k-1)^2 \left[\frac{L^2}{(v_s+q_s)^4} S^2(v_s) + \frac{L^2}{(v_s+q_s)^4} S^2(q_s) \right. \\
&+ \left. \frac{S^2(L)}{(v_s+q_s)^2} + \frac{-2L}{(v_s+q_s)^3} S(q_s) S(L) r(q_s, L) \right] \\
&+ \left(P_{i-1} - \frac{L}{v_s+q_s} \right) \left[\frac{2L(k-1)}{(v_s+q_s)^2} \cdot \left(\frac{-k}{z} \right) S^2(v_s) \right. \\
&+ \frac{2k}{\tau^2} \frac{L}{(v_s+q_s)^2} (k-1) S(\tau) S(q_s) r(\tau, q_s) \\
&+ \left. \frac{2(1-k)}{(v_s+q_s)} \frac{k}{\tau^2} S(L) S(\tau) r(L, \tau) \right] \\
&+ \left(P_{i-1} - \frac{L}{v_s+q_s} \right)^2 \left[\frac{k^2}{z^2} S^2(v_s) + \frac{k^2}{\tau^4} S^2(\tau) \right] \\
&+ k^2 S^2(P_{i-1}) + ME.
\end{aligned}$$

The terms in the first set of brackets are the steady-state error terms (referred to hereafter as SSE). They are scaled by the decay factor $(k-1)^2$ and the result is a fraction of the steady-state error which is notated hereafter as FSSE. The next two sets of brackets are scaled by the difference between the previous model prediction and the steady-state prediction. Since the phosphorus prediction model converges to the steady-state, the difference, $P_{i-1} - \frac{L}{v_s+q_s}$, converges to zero. Hence these terms diminish with each iteration and are therefore referred to as the "diminishing" factors notated as D_i 's.

The term $k^2 \cdot S^2 (P_{i-1})$ reduces the previous error by an amount equal to $1-k^2$ and retains only k^2 of the previous error. This is referred to as the "reduction" term.

The abbreviation ME is the model error and is a constant.

Combining the FSSE and ME to form one constant, and also combining the constants of the D_i 's gives a simpler equation. It should be observed that the diminishing factors can be rewritten according to Appendix B.4. Since

$$P_i - \frac{L}{v_s + q_s} = \left(P_0 - \frac{L}{v_s + q_s} \right) k^i \quad \text{then}$$

$$P_{i-1} - \frac{L}{v_s + q_s} = \left(P_0 - \frac{L}{v_s + q_s} \right) k^{i-1}$$

The diminishing factors can therefore be written in terms of a constant, the difference between the initialized value P_0 and the steady-state value, which decays according to k^{i-1} .

The error equation for the base run can be rewritten as:

$$S_t^2 = (.0000106880) + [(-.0006754555)k^{i-1}]$$

$$\cdot [-.0000013731] + [(-.0006754555)^2 k^{2i-2}]$$

$$\cdot [.0002292119] + k^2 \cdot S^2(P_{i-1})$$

This can be further simplified to:

$$S_t^2 = (.0000106880) + (.0000000009)k^{i-1}$$

$$+ (.0000000001)k^{2i-2} + k^2 \cdot S^2(P_{i-1})$$

More generally

$$S_t^2 = F + (A + Bk^{i-1})k^{i-1} + k^2 \cdot S^2(P_{i-1})$$

where

$$F = \text{FSSE} + \text{ME}$$

$$A = \text{coefficient for } k^{i-1}$$

$$B = \text{coefficient for } k^{2i-2}$$

or

$$S_t^2 = F + D_i + k^2 \cdot S^2 (P_{i-1})$$

where

$$D_i = (A + Bk^{i-1})k^{i-1}$$

Using this method of calculation for the base run the first 4 iterations yield:

	<u>D_i Error $\times 10^{-10}$</u>	<u>Total Error $\times 10^{-10}$</u>
1.	10	143705*
2.	7	179458
3.	5	197511
4.	3	206626

A two part program can be written which first calculates the appropriate constants and then the second part calculates the actual iterations using the simplified error equation above.

*The slight difference of this value with the other base run value, a difference of $1 \cdot 10^{-10}$, is believed to be due to machine calculation limitation. It occurs only at the first iteration, is retained for the second two iterations and is lost by the fourth iteration.

B.6 The Limit of the Error Equation

Let:

F = the fractional steady-state error (FSSE) and the model error (ME).

D_i = the i th diminishing factor

SS = the steady-state value

$k = 1/e^{Vs/z} + 1/\tau$

$S_t^2 = E_i$

then the error equation can be rewritten as:

$$E_i = F + D_i + k^2 E_{i-1}$$

We observe the following sequences:

$$E_1 = F + D_1 + k^2 E_0$$

$$\begin{aligned} E_2 &= F + D_2 + k^2 E_1 \\ &= F + D_2 + k^2 (F + D_1 + k^2 E_0) \\ &= (F + D_2) + k^2 (F + D_1) + k^4 E_0 \end{aligned}$$

$$\begin{aligned} E_3 &= F + D_3 + k^2 E_2 \\ &= F + D_3 + k^2 (F + D_2) + k^2 (F + D_1) + k^4 E_0 \\ &= (F + D_3) + k^2 (F + D_2) + k^4 (F + D_1) + k^6 E_0 \end{aligned}$$

⋮

$$E_i = \sum_{m=0}^{i-1} (F + D_{i-m}) k^{2m} + k^{2i} E_0$$

We wish to examine the results as i approaches infinity for the function E_i . This limit will be examined term by term.

1. $\lim_{i \rightarrow \infty} k^{2i} E_0 = 0$, since $0 < |k| < 1$.

2. The summation can be rewritten:

$$\sum_{m=0}^{i-1} (F + D_{i-m})k^{2m} = \sum_{m=0}^{i-1} Fk^{2m} + \sum_{m=0}^{i-1} D_{i-m}k^{2m}.$$

a.) The first summation on the right generates the series,

$$\begin{aligned} & Fk^0 + Fk^2 + Fk^4 + \dots + Fk^{i-1} \\ &= F(k^0 + k^2 + k^4 + \dots + k^{i-1}). \end{aligned}$$

This series generates a sequence of partial sums,

$$S_i = 1; 1 + k^2; 1 + k^2 + k^4; \dots$$

The sequence for the base run is,

$$S_i = 1; 1.5050; 1.7600; \dots$$

The generator form is,

$$\begin{aligned} S_i &= \frac{k^2}{1-k^2} (1 - k^{2i-2}) + 1 \\ &= \frac{k^2}{1-k^2} - \frac{k^2 k^{2i-2}}{1-k^2} + 1 \\ &= \frac{k^2}{1-k^2} - \frac{k^{2i}}{1-k^2} + \frac{1-k^2}{1-k^2} \\ &= \frac{1}{1-k^2} (1 - k^{2i}) \text{ and hence} \end{aligned}$$

$$\lim_{i \rightarrow \infty} \left[\frac{1}{1-k^2} (1 - k^{2i}) \right] = \frac{1}{1-k^2}.$$



Since the sequence of partial sums limits to $1/(1-k^2)$, then the original series converges to it, that is, $F \sum_{m=0}^{i-1} k^{2m}$ converges to $F/(1-k^2)$.

b.) We examine the remaining summation term,

$$\sum_{m=0}^{i-1} D_{i-m} k^{2m}.$$

The generator form for this sequence of partial sums is,

$$\begin{aligned} S_i &= \left[\frac{1}{1-k^2} (1 - k^{2i}) \right] D_i \\ &= \left[\frac{1}{1-k^2} (1 - k^{2i}) \right] (Ak^{i-1} + Bk^{2i-2}) \end{aligned}$$

(see Appendix B.5) and thus

$$\text{Limit}_{i \rightarrow \infty} \left[\frac{1-k^{2i}}{1-k^2} (Ak^{i-1} + Bk^{2i-2}) \right] = 0.$$

The original series, therefore, converges to zero.

Thus all terms approach zero as i approaches infinity except for the summation involving F . The total limit for the entire error equation is

$$\text{Limit}_{i \rightarrow \infty} E_i = \frac{F}{1-k^2}.$$

B.7 Generator Form of the Error Equation

The error equation is

$$\begin{aligned} E_i &= \sum_{m=0}^{i-1} (F + D_{i-m}) k^{2m} + k^{2i} E_0 \\ &= \sum_{m=0}^{i-1} F k^{2m} + \sum_{m=0}^{i-1} D_{i-m} k^{2m} + k^{2i} E_0. \end{aligned}$$

Substituting the generator form for the summations yields

$$\begin{aligned} E_i &= \frac{F}{1-k^2} (1-k^{2i}) + \frac{D_i}{1-k^2} (1-k^{2i}) + k^{2i} E_0 \\ &= \frac{F}{1-k^2} - \frac{F k^{2i}}{1-k^2} + k^{2i} E_0 + \frac{D_i}{1-k^2} (1-k^{2i}) \\ &= \left(E_0 - \frac{F}{1-k^2} \right) k^{2i} + \frac{F}{1-k^2} + \frac{D_i}{1-k^2} (1-k^{2i}) \end{aligned}$$

From Appendix B.5

$$D_i = (A + Bk^{i-1}) k^{i-1} \quad \text{and hence}$$

$$E_i = \left(E_0 - \frac{F}{1-k^2} \right) k^{2i} + \frac{F}{1-k^2} + \frac{(A + Bk^{i-1}) (1-k^{2i})}{1-k^2}.$$

In this form the equation cannot be solved for i . However, if the D_i 's are relatively "small," that is, if

$$\begin{aligned} D_i &= (A + Bk^{i-1}) k^{i-1} \\ &\approx 0 \end{aligned}$$

then a good approximation to i is available:

$$E_i \approx \left(E_0 - \frac{F}{1-k^2}\right) k^{2i} + \frac{F}{1-k^2}$$

$$E_i - \frac{F}{1-k^2} \approx \left(E_0 - \frac{F}{1-k^2}\right) k^{2i}$$

$$\frac{E_i - \frac{F}{1-k^2}}{E_0 - \frac{F}{1-k^2}} \approx k^{2i}$$

$$\approx \left(\frac{1}{e^{v_s/z + 1/\tau}}\right)^{2i}$$

$$\ln \left(\frac{E_i - \frac{F}{1-k^2}}{E_0 - \frac{F}{1-k^2}} \right) \approx i \cdot \frac{-2(v_s/z + 1/\tau)}{1}$$

B.8 Behavior of the Coefficient of Variation

The coefficient of variation is the ratio of the standard deviation to the mean. Since both the standard deviation and the mean asymptotically limit to a final value, the coefficient of variation does also. Thus,

$$CV_{\infty} = \frac{\sqrt{E_{\infty}}}{P_{\infty}} .$$

However, the coefficient need not behave in the same manner as the prediction or its error. Due to the different rates of convergence and the different initial values, the coefficient can actually exceed the limit before asymptotically approaching its final value.

This becomes clear when graphing the prediction and the error together as normalized functions. That is,

$$\begin{aligned} P_i &= SS + (P_0 - SS)k^i \\ &= SS(1-k^i) + P_0 k^i \end{aligned}$$

$$\frac{P_i}{SS} = 1-k^i + \frac{P_0}{SS} k^i .$$

And, as the variance, is

$$\begin{aligned} E_i &\approx \frac{F}{1-k^2} + (E_0 - \frac{F}{1-k^2})k^{2i} \\ &\approx \frac{F}{1-k^2} (1-k^{2i}) + E_0 \cdot k^{2i} \text{ and} \end{aligned}$$

$$\frac{E_i}{F/(1-k^2)} \cong 1-k^{2i} + \frac{E_0}{F/(1-k^2)} k^{2i}.$$

There are six cases to consider in this approach. The first case occurs when two conditions are satisfied:

1. both the prediction and the error grow, that is,

$$\frac{P_0}{SS} < 1$$

$$\frac{E_0}{F/(1-k^2)} < 1 \text{ and}$$

2. the initial normalized error is greater than the initial normalized prediction,

$$\frac{E_0}{F/(1-k^2)} > \frac{P_0}{SS}.$$

When these conditions are met, and because of the faster rate of convergence for the error equation, the coefficient of variation, CV, will always have a normalized value greater than one and will asymptotically decay toward one.

The second case is similar to the first in that both the prediction and the error grow, that is, the first condition is met but the second condition is not satisfied. Instead, the initial normalized error is smaller than the initial normalized prediction,

$$\frac{E_0}{F/(1-k^2)} < \frac{P_0}{SS}.$$

In this case the error equation will eventually cross the prediction equation due to the fact that it converges more quickly. Before the

crossing,

$CV < 1$ and after it

$CV > 1$.

When the two functions equal each other exactly then

$$\frac{E_i}{F/(1-k^2)} = \frac{P_i}{SS} \quad \text{and}$$

$CV = 1$.

The point at which these functions are equal is the point at which the CV equals the limit. This can be easily calculated:

$$\begin{aligned} \frac{E_i}{F/(1-k^2)} &= \frac{P_i}{SS} \\ 1-k^{2i} + \frac{E_0}{F/(1-k^2)} k^{2i} &= 1-k^i + \frac{P_0}{SS} k^i \\ -k^{2i} + \frac{E_0}{F/(1-k^2)} k^{2i} &= -k^i + \frac{P_0}{SS} k^i \\ \left(\frac{E_0}{F/(1-k^2)} - 1 \right) k^i &= \frac{P_0}{SS} - 1 \\ k^i &= \frac{(P_0/SS - 1)}{(E_0/F/(1-k^2) - 1)} \\ i &= \ln \left(\frac{P_0/SS - 1}{\frac{E_0/F/(1-k^2) - 1}{-v_s/z + 1/\tau}} \right). \end{aligned}$$

The third case arises when the following two conditions are met:

1. both the prediction and the error decay, that is,

$$\frac{P_0}{SS} > 1$$

$$\frac{E_0}{F/(1-k^2)} > 1 \quad \text{and}$$

2. the initial normalized error is smaller than the initial normalized prediction,

$$\frac{E_0}{F/(1-k^2)} < \frac{P_0}{SS} .$$

When these conditions are satisfied the coefficient of variation, CV, will always have a normalized value less than one and will asymptotically grow toward one.

The fourth case is similar to the third case except that the second condition is not met. Instead, the initial normalized error is greater than the initial normalized prediction,

$$\frac{E_0}{F/(1-k^2)} > \frac{P_0}{SS} .$$

In this situation, as in the second instance, the error equation will cross the prediction equation. Before the crossing

$$CV > 1 \text{ and after it}$$

$$CV < 1 .$$

When the two functions are exactly equal,

$$CV = 1 .$$

After crossing the limit, the CV reaches a minimum and asymptotically approaches the limit from below. The calculation of i is the same as before.

The fifth case occurs when the prediction decays but the error grows,

$$\frac{P_i}{SS} > 1 \text{ and}$$

$$\frac{E_i}{F/(1-k^2)} < 1$$

then

$$CV < 1 .$$

The CV will asymptotically approach the limit from below.

The sixth case is the reverse of the above, namely, when the prediction grows but the error decays,

$$\frac{P_i}{SS} < 1 \text{ and}$$

$$\frac{E_i}{F/(1-k^2)} > 1$$

then

$$CV > 1.$$

The CV will asymptotically approach the limit from above.

Only in the second and fourth cases do the functions equal one another and hence, only in these two instances do the coefficients of variation equal and exceed the limit before they truly asymptotically approach their final values.

B.9 Relationships of the Time-Dependent Error to the Steady-State Error

A comparison of the steady-state error terms with those of the time-dependent form shows that the time dependent error is equal to the steady-state error scaled by the decay factor, i.e., it constitutes a fraction of the steady state error:

$$FSSE = (1 - k)^2 SSE$$

Since $0 < k < 1$, it follows that one relation is,

$$FSSE < SSE.$$

The final time-dependent error has a similar relationship,

$$\begin{aligned} E_{\infty} &= \frac{FSSE}{1-k^2} \\ &= \frac{(1 - k)^2 SSE}{1-k^2} \\ &= \frac{1 - k}{1 + k} SSE < SSE. \end{aligned}$$

Hence,

$$\frac{FSSE}{1-k^2} < SSE .$$

The final equilibrium value is always greater than the fractional steady-state value. This follows from a consideration of the decay factor:

$$0 < k < 1$$

$$0 < k^2 < 1$$

$$-1 < k^2 - 1 < 0$$

$$1 > 1 - k^2 > 0$$

$$\frac{1}{1 - k^2} > 1 \text{ hence}$$

$$\frac{\text{FSSE}}{1 - k^2} > \text{FSSE}.$$

The same argument holds for the model error ME, giving the relation

$$\frac{\text{ME}}{1 - k^2} > \text{ME}.$$

It should be noted that if the model error is sufficiently large, the time-dependent error can be larger than the steady-state error as shown below:

$$\frac{\text{FSSE}}{1 - k^2} + \frac{\text{ME}}{1 - k^2} > \text{SSE} + \text{ME}$$

$$\frac{\text{ME}}{1 - k^2} - \text{ME} > \text{SSE} - \frac{\text{FSSE}}{1 - k^2}$$

$$\left(\frac{1}{1 - k^2} - 1 \right) \text{ME} > \left(1 - \frac{(1 - k)^2}{1 - k^2} \right) \text{SSE},$$

recalling that $\text{FSSE} = (1 - k)^2 \text{SSE}$. And,

$$\left(\frac{k^2}{1+k^2}\right) \text{ME} > \left(1 - \frac{1-k}{1+k}\right) \text{SSE}$$

$$\frac{k^2}{1-k^2} \text{ME} > \frac{2k}{1+k} \text{SSE}$$

$$\frac{\text{ME}}{\text{SSE}} > 2(1/k - 1).$$

Thus, when the ratio of the model error to the steady-state error is greater than $2(1/k - 1)$, the equilibrium value of the time-dependent form will be greater than the steady-state value. In the base run, the ratios are

$$1.914 > .9802.$$

Relationships of the Dynamic Error
to the Steady-State Error

1. $\frac{\text{FSSE}}{1-k^2} > \text{FSSE}$

2. $\text{FSSE} < \text{SSE}$

3. $\frac{\text{FSSE}}{1-k^2} < \text{SSE}$

4. $\frac{\text{ME}}{1-k^2} > \text{ME}$

5. $\frac{\text{FSSE}}{1-k^2} + \frac{\text{ME}}{1-k^2} > \text{SSE} + \text{ME}$

when $\frac{\text{ME}}{\text{SSE}} > 2(1/k - 1).$

B.10 The Calculator Program

The results of this study were calculated on a Texas Instruments programmable calculator model 58C. The workspace was partitioned into 320 programmable steps (0 - 319) and 20 data registers (0 - 19). Listed below are the data register constants, including their base run values, and the 294 step program.

	<u>Data Registers</u>	<u>Base Run Values</u>
R ₀₀	P _i	
R ₀₁		
R ₀₂		
R ₀₃		
R ₀₄		
R ₀₅		
R ₀₆	ME	.0032
R ₀₇	r(q _s ,L)	.6822
R ₀₈	r(L,τ)	-.7078
R ₀₉	r(q _s ,τ)	-.9902
R ₁₀	exp (-v _s /z - 1/τ)	-19.1910/89 - 1/7.9402 = -.3416
R ₁₁	v _s + q _s	19.1910 + 10.6650 = 29.8560
R ₁₂	S ² of v _s	1.4311
R ₁₃	S ² of q _s	2.1339
R ₁₄	L	.6352

	<u>Data Registers</u>	<u>Base Run Values</u>
R ₁₅	SD ² of L	.0066
R ₁₆	τ	7.9402
R ₁₇	SD ² of τ	1.0860
R ₁₈	P _{i-1}	.0206*
R ₁₉	SD ² of P _{i-1}	.00000729*

The Program

Prediction

```
Rc1 00 Sto 18
Rc1 14 ÷ Rc1 11 x (1 - Rc1 10 Inv ln) + Rc1 18 x Rc1 10 Inv ln
= Sto 00 R/S
```

Parameter

$$v_s \left((Rc1\ 10\ Inv\ ln - 1) \times Rc1\ 14 \div Rc1\ 11 \times x^2 + (Rc1\ 18 - Rc1\ 14 \div Rc1\ 11) \times .0112 \div x \times Rc1\ 10 \right. \\ \left. Inv\ ln \right) \times x^2 \times Rc1\ 12 = R/S$$

Error

Input Variables

$$P_{i-1} + Rc1\ 10\ Inv\ ln \times x^2 \times Rc1\ 19 \\ q_s + (Rc1\ 14 \times (Rc1\ 10\ Inv\ ln - 1) \div Rc1\ 11 \times x^2) \times x^2 \\ L \times Rc1\ 13 + ((1 - Rc1\ 10\ Inv\ ln) \div Rc1\ 11) \times x^2 \times \\ Rc1\ 15 \\ \tau + ((Rc1\ 18 - Rc1\ 14 \div Rc1\ 11) \times Rc1\ 10\ Inv\ ln \\ \div Rc1\ 16 \times x^2) \times x^2 \times Rc1\ 17$$

* Initial Value

Error	Input Variables	$r(q_s, L) \quad + 2 \times \text{Rc1 } 14 \times (\text{Rc1 } 10 \text{ Inv ln} - 1) \div \text{Rc1 } 11 \times x^2$ $\times (1 - \text{Rc1 } 10 \text{ Inv ln}) \div \text{Rc1 } 11 \times \text{Rc1 } 15 \times$ $\times \text{Rc1 } 13 \times \text{Rc1 } 07$
		$r(q_s, \tau) \quad + 2 \times \text{Rc1 } 14 \times (\text{Rc1 } 10 \text{ Inv ln} - 1) \div \text{Rc1 } 11 \times x^2$ $\times (\text{Rc1 } 18 - \text{Rc1 } 14 \div 11) \times \text{Rc1 } 10 \text{ Inv ln} \div \text{Rc1 } 16$ $x^2 \times \text{Rc1 } 13 \times * \text{Rc1 } 17 \times * \text{Rc1 } 09$
		$r(L, \tau) \quad + 2x (1 - \text{Rc1 } 10 \text{ Inv ln}) \div \text{Rc1 } * (\text{Rc1 } 18 -$ $\text{Rc1 } 14 \div \text{Rc1 } 11) \times \text{Rc1 } 10 \text{ Inv ln} \div \text{Rc1 } 16 \times x^2$ $\times \text{Rc1 } 15 \times \text{Rc1 } 17 \times \text{Rc1 } 08$
	ME	$+ \text{Rc1 } 06 = \text{Sto } 19 \text{ R/S RST}$

(The program corresponds to the error equation in a straightforward manner using calculator keyboard notation.)

BIBLIOGRAPHY

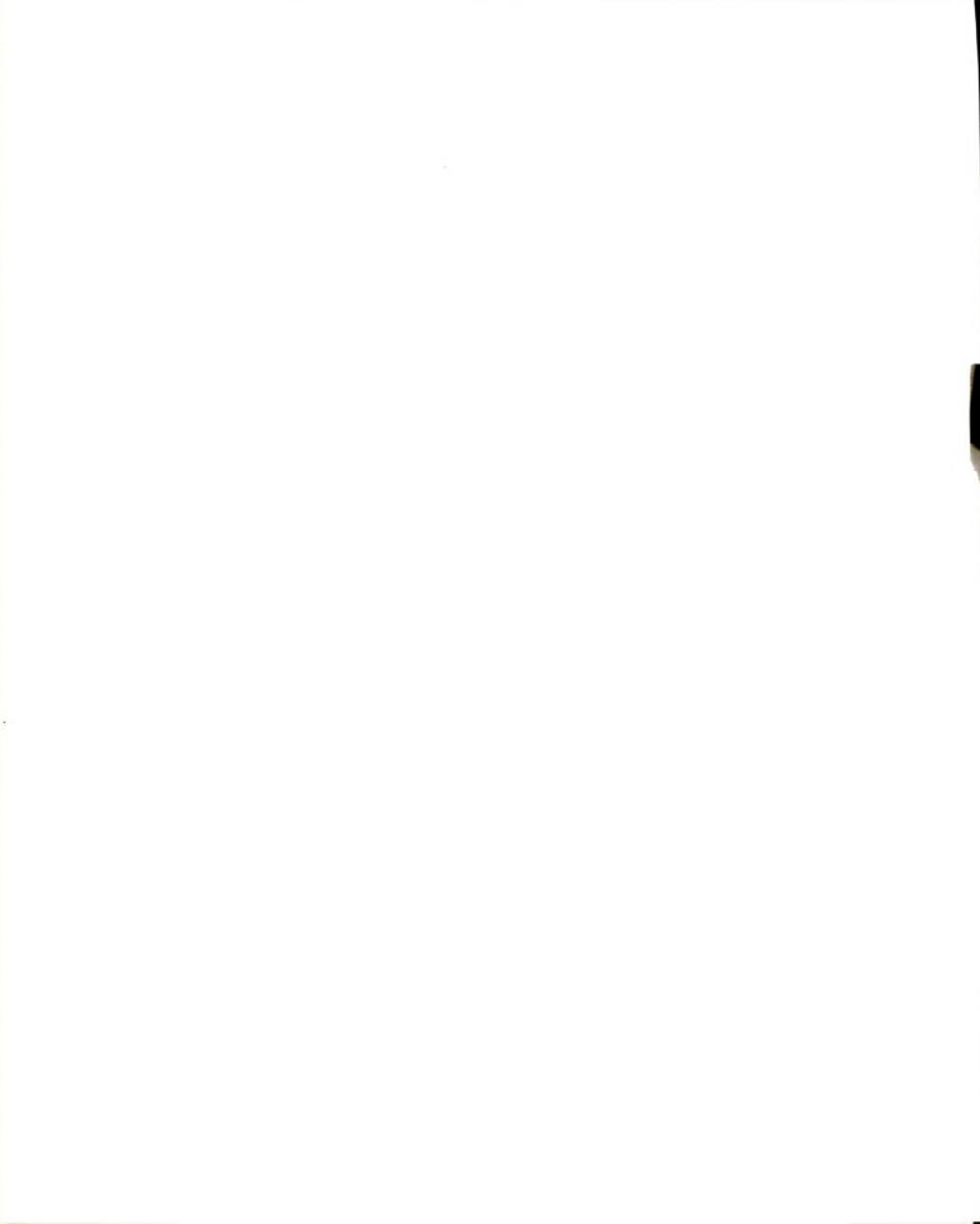
BIBLIOGRAPHY

- Beaulac, M. N.; 1980; Nutrient Export Coefficients: An Examination of Sampling Design and Natural Variability Within Differing Land Uses; Master's Thesis, Michigan State University, East Lansing, Michigan.
- Benjamin, J. R. and C. A. Cornell; 1970; Probability, Statistics, and Decision for Civil Engineers; McGraw-Hill Book Company. New York. pp. 180-186.
- Berthouex, P. M.; 1975; Modeling Concepts Considering Process Performance, Variability and Uncertainty. In Mathematical Modeling for Water Pollution Control, pp. 405-440. T. M. Keinath and M. P. Wanielista, editors. Ann Arbor Science Publishers, Inc. Ann Arbor, Michigan.
- Burges, S. J. and D. P. Lettenmaier; 1975; Probabilistic Methods in Stream Quality Management; Water Resources Bulletin, 11, 115-130.
- Chapra, S. C.; 1977; Total Phosphorus Model for the Great Lakes: ASCE Journ. Econ. Eng'g. Div. 103; 147-161.
- Chapra, S. C. and K. H. Reckhow; 1979; Expressing the Phosphorus Loading Concept in Probabilistic Terms; Journ. Fish. Res. Board Can. 36: 225-229.
- Chapra, S. C.; 1981; National Oceanic and Atmospheric Administration, Ann Arbor, Michigan. Personal communication.
- Coleman, G. and D. G. DeCoursey; 1976; Sensitivity and Model Variance Analysis Applied to Some Evaporation and Evapo Transportation Models. Water Resources Research 12 (5): 873-879.
- Cornell, C. A.; 1969; A Normative Second-Moment Reliability Theory for Structural Design; Seminar No. 6, December 15, 1969; Soil Mechanics Division, University of Waterloo, Waterloo, Ontario, Canada.

- Cornell, C. A.; 1970; A First-order Reliability Theory for Structural Design. In Structural Reliability and Codified Design, M. C. Lind, editor. Waterloo University, Waterloo, Ontario.
- Cornell, C. A.; 1972; First-order Analysis of Model and Parameter Uncertainty. In Proceeding of International Symposium on Uncertainties in Hydrologic and Water Resource Systems, pp. 1-31, University of Arizona, Tucson, 1972.
- DePinto, J. V., J. K. Edzwalk, M. S. Switzenbaum and T. C. Young; 1980; Phosphorus Removal in Lower Great Lakes Municipal Treatment Plants. United States Environmental Protection Agency, in press.
- Fiering, M. G.; 1967; Streamflow Synthesis. Harvard University Press, Cambridge, Massachusetts.
- Gardner, R. H., R. V. O'Neill, J. B. Mankin and J. H. Carney; 1981; A Comparison of Sensitivity Analysis and Error Analysis Based on a Stream Ecosystem; Ecological Modelling 12 (1981): 173-190.
- Garen, D. C.; 1979; An Analysis of the Effects of Parameter Uncertainty on Runoff Predictions Using the Stanford Watershed Model. Master's Thesis. University of Washington.
- Gumbel, E. J.; 1958; Statistics of Extremes; Columbia University Press, New York.
- Howard, R. A.; 1971; Proximal Decision Analysis; Management Science; Vol. 17, No. 9; May, 1971. pp. 507-541.
- IJC; 1977; International Joint Commission; Land Use and Land Use Practices in the Great Lakes Basin, Summary Report--Task B; Windsor, Ontario, Canada.
- Kendall, M. G. and A. S. Stuart; 1958; The Advanced Theory of Statistics; Vol. 1, Distribution Theory; Charles Griffen and Company.
- Lee, V. D.; 1980; Lake Model Uncertainty Analysis: A Case Study Using Monte Carlo Simulation; Master's Thesis; Michigan State University; East Lansing, Michigan.
- McGrath, E. J., and D. C. Irving; 1973; Techniques for Efficient Monte Carlo Simulation; Office of Naval Research Rep. HD762 721-723 (3 volumes), 408 pp. Nat. Tech. Inform. Serv., Springfield, VA.
- Mein, R. G., and B. M. Brown; 1978; Sensitivity of Optimized Parameters in Watershed Models; Water Resources Research 14(1) (4): 299-303.

- Michigan State University, Computer Laboratory; 1978; SPSS-6000 Supplement.
- Montgomery, R. H., V. D. Lee, and K. H. Reckhow; 1980; A Comparison of Uncertainty Analysis Techniques: First-order Analysis vs. Monte Carlo Simulation. Presented at 1980 Conference on Great Lakes Research International Assoc. for Great Lakes Research, Queens College, Kingston, Ontario.
- Norman, H. N., C. H. Hull, J. G. Jenkins, K. Steinbrenner, D. H. Brent; 1975; Statistical Package for the Social Sciences (SPSS); second edition; McGraw-Hill Book Company, New York.
- Reckhow, K. H.; 1978; Sampling Designs for Lake Phosphorus Budgets. Paper presented at the 1978 American Water Resources Association Symposium on the Establishment of Water Quality Monitoring Programs.
- Reckhow, K. H.; 1979; Quantitative Techniques for the Assessment of Lake Quality; U.S. Environmental Protection Agency, Office of Water Planning and Standards; Washington, D.C. 20460; pp. 85-88.
- Reckhow, K. H., M. N. Beaulac, and J. T. Simpson; 1980; Modeling Phosphorus Loading and Lake Response Under Uncertainty: A Manual and Compilation of Export Coefficients. United States Environmental Protection Agency, EPA-440/5-80-011.
- Reckhow, K. H., R. E. Ancil, M. N. Beaulac, V. D. Lee, and R. H. Montgomery; 1981; Methods for the Analysis of Errors and Variability in Lake Trophic State Determination; National Oceanic and Atmospheric Administration Completion Report, Contract Number 03-78-bol-109; Appendix F.
- Scavia, D., W. F. Powers, R. P. Canale, J. L. Moody; 1981; Comparison of First-order Error Analysis and Monte Carlo Simulation in Time-dependent Lake Eutrophication Models; Water Resources Research; Vol. 17 (4): pp. 1051-1059 and 1115-1124.
- Snodgrass, W. J.; 1974; A Predictive Phosphorus Model for Lakes-- Development and Testing; Ph.D. dissertation, University of North Carolina; Chapel Hill, North Carolina.
- Sonzogni, W. G., T. J. Monteith, W. N. Bach, U. G. Hughes; 1978; United States Great Lakes Tributary Loadings; Great Lakes Basin Commission Staff; Ann Arbor, Michigan. Prepared for International Joint Commission's Pollution from Land Use Activities Reference Group (PLUARG); pp. 134 and 137.

- Thomann, R. V., R. P. Winfield, D. S. Symuski; 1977; Estimated Response of Lake Ontario, Phytoplankton Biomass to Varying Nutrient Levels; J. Great Lakes Research, 3, 123-131.
- Uttormark, P. D. and M. L. Hutchins; 1980; Input/Output Models as Decision Aid for Lake Restoration; Water Resources Bulletin Vol. 16, No. 3, pp. 494-500.
- Vollenweider, R. A.; 1975; Input-Output Models with Special Reference to the Phosphorus Loading Concept in Limnology; Sch. Zeit. Hydrologic, 37, pp. 53-84.
- Vollenweider, R. A.; 1969; Possibilities and Limits of Elementary Models Concerning the Budget of Substances in Lakes (in German), Arch. Hydrobiol., 66, 1-36.
- Weibull, W; 1951; A statistical Distribution Function of Wide Applicability. J. App. Mech. 18(3): 293-297.
- World Weather Records, 1941-1970. Volume 1, North America.



MICHIGAN STATE UNIV. LIBRARIES



31293008373973