This is to certify that the

thesis entitled

LAKE MODEL UNCERTAINTY ANALYSIS
- a lake Ontario case study using Monte Carlo Simulation

presented by

V. David Lee

has been accepted towards fulfillment
of the requirements for

Master of Science degree in Resource Development

Kenneth H. Reckhow
Major professor

Date January 5, 1981

LAKE MODEL UNCERTAINTY ANALYSIS

-a Lake Ontario case study using Monte Carlo simulation-

By

V. David Lee

A THESIS

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

MASTER OF SCIENCE

Department of Resource Development

1980

ABSTRACT

LAKE MODEL UNCERTAINTY ANALYSIS

- A LAKE ONTARIO CASE STUDY USING MONTE CARLO SIMULATION -

By

V. David Lee

A growing concern over the reliability of modeling and model predictions has highlighted the need for a measure of the value of the information contained in the model result. Uncertainty analysis meets this need by allowing quantification of model prediction uncertainty. Monte Carlo simulation accomplishes this by incorporating all known sources of error as stochastic inputs which reflect the natural variability and estimation uncertainty for each parameter.

A discussion of general modeling concepts, an introduction to Monte Carlo simulation, and an overview of distribution selection considerations serve as an introduction to the Lake Ontario case study. A detailed development of the model expression and the simulation structure is presented, and model sensitivities as well as several planning scenarios are experimentally tested. Following a discussion, conclusions and recommendations are made and analyzed in terms of the meaningfulness of the results and the utility of the model as a planning tool.

ACKNOWLEDGEMENTS

This research was supported through funding granted by the National Oceanic and Atmospheric Administration (grant #03-78-b01-109) and by the Michigan State University Agricultural Experiment Station.

I would like to acknowledge the invaluable contributions made by Janine Niemer and Sue Watt for preparing this document in its final form, and by Paul Schneider for his superb graphics work. Special thanks are also extended to my friends and colleagues Michael Beaulac, Ralph Ancil, and Robert Montgomery for their constructive criticism, editorial advice, and the intellectually stimulating environment they helped to provide.

I expressly thank my committee members Dr. Kenneth Reckhow, Dr. Daniel Chappelle, and Dr. Stanley Zarnoch for their comments and criticisms. These gentlemen have not lost sight of the most noble goals of higher education, and as a result, I have not only gained facts and skills but a valuable learning experience in the process.

Dr. Reckhow is especially deserving of very special thanks. As my major professor and friend, he has provided me with an example of professionalism and high academic standards that will remain with me always. Through his guidance, personal concern, and seemingly endless patience, I have matured both as a professional and as a person. I owe him a great deal.

Deeply felt appreciation is also due to my close friend, Cynthia Hauger. It has been her understanding, support, and love that have sustained me for these many months, and I thank her with all my heart.

Finally, I wish to thank my mother and father. Their support and encouragement have provided the opportunity for me to realize and overcome my own shortcomings without being penalized for my own inability to recognize good advice when I first hear it. I love them dearly, and it is to them that I dedicate this thesis.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
CHAPTER ONE: MODELS	4
Model Types	4
Mathematical Models	5
Considerations Central to Model Use	6
Model Costs	7
Complexity, Accuracy, and Uncertainty	9
Summary Comments	11
CHAPTER TWO: MONTE CARLO SIMULATION	13
The Monte Carlo Procedure	13
Distribution Selection	14
Bases for Distribution Selection	16
Qualitative Basis	16
Quantitative Basis	17
Quantitative Techniques	18
Sensitivity Analysis	19
Graphical Analysis	19
Parameter Estimation	20
Simple Parametric Distributions	21
Complex Parametric Distributions	22
Goodness-of-Fit Testing	23
CHAPTER THREE: THE SIMULATION MODEL APPLICATION	24
The Model Expression	24
System Representation	28
Distributions and Random Number Generation	28
Phosphorus Loading	29
Diffuse Source Flux	30
Atmospheric Flux	32
Point Source Flux	34
Lake Erie Flux	35
Settling Velocity	36
Areal Water Loading	37
Additional Model Terms	38
Simulation Flow and Structure	39

Page

CHAPTER FOUR: RESULTS 40

Model Results 40
Flux 40
Waterload 41
Model Sensitivities 42
Number of Runs 42
Distribution Selection 43
Random Seed 43
Planning Scenarios 44
Point Source Flux 44
Erie Concentration 46
Land Use Fractions 47
Variability 48

CHAPTER FIVE: ANALYSIS AND DISCUSSION 51

FINAL COMMENTS 60

BIBLIOGRAPHY 62

APPENDICES 66

A. Model Parameter Data Distributions 66
B. Model Results - Graphical Summary 79
C. Model Results - Tabular Summary 99
D. 1000 Iteration Periodicity 111
E. Flow Diagram and Computer Listing 114

LIST OF TABLES

Table		Page
3.1	Land Use Projections for the Great Lakes Basin	33
C.1	100 Iteration Base Run - Model Terms	100
C.2	100 Iteration Base Run - Flux Terms	101
C.3	100 Iteration Base Run - Water Load Terms	102
C.4	Model Sensitivities - Number of Runs, Distribution Selection	104
C.4a	Model Sensitivities - Random Seed	105
C.5	Planning Scenarios - Point Source, Lake Erie Concentration	106
C.5a	Planning Scenarios - Land Use Fractions	107
C.6	Single Year Prediction Distributions - Base Run, All Normal	109
C.7	Prediction Variability Over Time	110

LIST OF FIGURES

Figure		Page
A.1	Runoff Concentrations From Urban Watersheds - Data	67
A.2	Runoff Concentrations From Agricultural Watersheds	68
A.3	Runoff Concentrations From Forested Watersheds	69
A.4	Atmospheric Flux to Urbanized Land Uses	70
A.5	Atmospheric Flux to Agricultural Land Uses	71
A.6	Atmospheric Flux to Forested Land Uses	72
A.7	Atmospheric Water Loading to Lake Surface	73
A.8	Water Load From Lake Ontario Basin Tributaries	74
A.9	Water Load From Lake Erie at Niagara-Welland	75
A.10	Phosphorus Concentration From Lake Erie at Niagara-Welland	76
A.11	Apparent Settling Velocity	77
A.12	Hydraulic Detention Time	78
B.1	100 Iteration Base Run - Flux Sources	80
B.2	100 Iteration Base Run - Water Load Sources	81
B.3	100 Iteration Base Run - Concentration Predictions	82
B.4	500 Iteration Run - Concentration Predictions	83
B.4a	500 Iteration Run, Increased Random Seed - Concentration Predictions	84
B.5	1000 Iteration Run - Concentration Predictions	85
B.5a	1000 Iteration Run, Random Seed = 135432.DO - Concentration Predictions	86
B.5b	1000 Iteration Run, Increased Markov Constant - Concentration Predictions	87

Figure		Page
B.5c	1000 Iteration Run, Decreased Settling Velocity - Concentration Predictions	88
B.5d	1000 Iteration Run, Random Seed = 987225.DO - Concentration Predictions	89
B.6	Base Run, 1/2 Point Source Flux Variability - Concentration Predictions	90
B.7	Base Run, Upgrade All Point Sources - Concentration Predictions	91
B.8	Comparison of Point Source Predictions - Flux Predictions	92
B.9	Base Run, Eastern Erie Concentrations - Concentration Predictions	93
B.10	Comparison of Shifts to Single Land Use Types - Concentration Predictions	94
B.11	Four Single-Year Prediction Distributions, Base Run	95
B.12	Four Single-Year Prediction Distributions, All Normal Distributions	96
B.13	Base Run, All Normal Distributions - Concentration Predictions	97
B.13a	All Normal Distributions, Random Seed = 987225.DO - Concentration Predictions	98
E.1	Flow Diagram of Simulation	115
E.2	Diagramatic Representation of Prediction Altorithm	116

INTRODUCTION

Concern for the environment follows a pendulum-like path; sometimes foremost in our thoughts, often superseded by seemingly more immediate concerns. Since the time of the industrial revolution, degradation of the environment has been a prominent side effect of the rapid increases in technology and leisure time. The associated demands placed on natural resource systems, combined with the rapid expansion of technological expertise, has resulted in an acute awareness of our ability to alter the environment. Not until most recently, however, has this awareness begun to alter the way in which the world and its resources are perceived. The past 20 years have highlighted the dangers of pollution, and the hazards of exploiting the earth's resources. Demands for effective action to protect the environment have been voiced in increasing numbers. This change in collective consciousness has been remarkably swift and far-reaching, and represents perhaps the most important of the recent social, political, and economic changes which bears directly and critically on the future of our planet.

Our planet is both fragile and finite. Decisions made today will dictate the differences between a polluted, unproductive world, and a world that can sustain future generations. However, making those decisions is never the clear cut, straight forward procedure it could be ideally. This multifaceted process must give a high priority to

assessing potential impacts of different resource use strategies. The extent to which negative impacts are minimized in favor of positive impacts, however, relies heavily on the judgment of the decision-maker. Traditionally, this responsibility has been met with very little quantitative information. While qualitative information is essential to any resource utilization decision, decisions solely based on this type of information are risky at best. The trend toward increasing accountability and justification requirements imposed on decision-makers has led toward a growing reliance on explicit quantitative analysis of the essential cause and effect relationships. This type of analysis requires considerable amounts of specific information; information that is often not readily available. Major decisions regarding resource management are made daily as part of numerous planning programs, and modeling has become a popular means for meeting the increasing expectations which planners encounter in their continual efforts to justify expenditures and fully achieve all expected results.

The major concern that has prompted the following work is the potential use and, most importantly, misuse of models in the resource planning context. As one of the most powerful planning tools available, modeling needs to be critically evaluated to effect their greatest positive impact. In particular, the emphasis of this investigation is focused on the topic of uncertainty analysis as an aid to decision-making. From a cost-effectiveness standpoint, incorporating uncertainty analysis into the modeling process presents one of the few excellent opportunities to gain a great deal of meaningful information for very little additional cost.

In order to establish a common ground for all readers, as well as a point of departure for the example application, some basic information on models and modeling is included. This is followed by a discussion of the Monte Carlo procedure with special attention given to distribution selection, the heart of Monte Carlo simulation. The remainder of this analysis consists of the application of this stochastic process to the Lake Ontario system, and a discussion of simulation development and model outputs.

CHAPTER 1

MODELS

Models are simplifications of reality. For this reason, any particular system can be modeled in a multitude of ways because there are many ways to simplify the same reality. Different model types can be classified by their structural characteristics.

Model Types

Physical models are intended to closely resemble the subject in appearance. Characteristically, scaling techniques are used to yield products such as a globe or photograph. Analog models are characterized by the use of graphical or schematic representation. Often transformed equivalents are employed for the development of typical products such as block diagrams and flowcharts. Finally, there are symbolic models which are characterized by the use of symbols, mathematical and logical, to represent the system components and interactions. Of these general types, only the symbolic has been extensively applied in the resource management contest. Of course, physical and analog models can be useful in specific instances.

The physical model is useful in studying the physical characteristics of certain systems. This modeling process, however, is often expensive, highly system specific, and thus quite limited in usefulness. Analog models are often very useful for general planning. There is

however, no facility for describing other than qualitative relationships. For these reasons, the mathematical model now bears the lion's share of the modeling performed for resource planning and management, and has revolutionized information systems.

Mathematical Models

This model type is actually a set of model types, as diverse in make up as the set of all models as a whole. Basically, these models can be classified as either empirical or theoretical. Empirical models are developed from statistical analysis of the available data, while the theoretical models attempt to describe the system components and their interactions through the use of equations derived to match the mechanistic operation of the system. In addition, mathematical models may be specialized to perform optimization or simulation duties. Optimization involves attempting to discover the best conditions, as defined by an objective function and constraints for a given situation. Simulation seeks only to create circumstances which emulate the system processes. Three additional classifications must be taken into account to characterize the entire spectrum of mathematical models. First, the time dependency question must be addressed. Static models do not allow for conditions in which the values of the variables change with time. These models tend to be simpler, however, and require less computational effort than their dynamic counterparts, which are capable of accounting for the effects of transient phenomena. Second, cross-sectional and longitudinal dimensionality can vary according to need. Real world resource systems are three dimensional in nature. However, two dimensions, or often one, may be appropriate when the processes

desired can be adequately described. Temporal dimensionality is dictated primarily by a model's time dependency; dynamic models being capable of continuous longitudinal forecasting. Third, and last, to be considered is the degree to which real world variability is built into the model. Deterministic models are based on physical laws and empirical formulas, and are frequently regarded as expected-mean-value models. Stochastic (or probabilistic) models take into account the randomness or variability inherent in system phenomena as well as the errors associated with quantitatively characterizing the system. Although more realistic in their representation of physical processes, excessive data requirements often limit their opportunity for use.

Considerations Central to Model Use

Mathematical models can be used in a variety of ways. Whether or not they should be employed in any given situation is a separate and important question. A direct answer which applies to all situations across the board does not exist. The decision to model must be made on a case by case basis according to the specific requirements of each case. The availability of an existing model which is suitable or the need for developing a new model to meet specifications must be considered. Time constraints, data availability, and data collection requirements additionally all exert a strong influence on decisions to model, and may differ substantially from case to case.

A basic understanding of the advantages and disadvantages of model building provides a basis on which to make this decision. Chappelle (1972) elaborates the major strengths of model building. He notes that:

1) the most successful predicting systems to date employ model building, 2) since the model represents the designer's view of the real world, all of the inherent simplifying assumptions are explicitly recognized, 3) models force recognition of shortcomings at an early stage, 4) once represented symbolically, the system can be related more easily in the literature, and can be more easily manipulated, 5) mathematical models are often the least expensive means to accomplish prediction, and 6) modern theoretical statistics can be used as a powerful tool in the analysis and manipulation of the model. Many dangers also exist however. Inappropriate simplification, unnecessary complexity, and model building for model building's sake all reduce the probability that the model can be used to provide meaningful results.

Foremost in the minds of model users is the ability of modeling to give quantitative answers to complex planning problems through the elucidation of fundamental cause-effect relationships. Depending on the solution technique used, these quantitative answers can be derived quickly and inexpensively. Computer assisted computation is one example.

Model Costs

As mentioned above, however, computation is only a portion of the time and cost necessary to complete a successful modeling exercise. Model selection and development (or modification), verification, degree of complexity and uncertainty are all important cost considerations. In addition, the availability of all necessary data, in the appropriate

format to meet all these requirements must be taken into account.

A wide variety of models are currently available, and as new models are developed, the probability of existence of a relevant model, which needs little modification, increases. Despite the fact that there are many models in existence and more being developed, this does not necessarily represent a duplication of efforts (Reckhow et. al., 1980b). Models are needed for a wide range of problem types, addressing issues at different levels of complexity and precision. Additionally, in very few instances, at this point in time, can a model be expected to efficiently and effectively meet the need for specific information without some modification, and therefore some cost. A major consideration in the use of any model is the type, amount and accuracy of the data needed to carry out experiments with specified reliability. Data are needed for calibration and verification of the model, as well as for experimental applications. Calibration and verification are distinctly different procedures. Calibration is performed using one or more data sets for model inputs and outputs to adjust and tune the model. Verification involves testing the calibrated model using independent data. Preferably, conditions for verification should differ from those used for calibration. Favorable comparisons between model predictions and corresponding field data serve as verification. Collection of field data, or assembly of historical data, is a costly and time consuming procedure. Once verification is accomplished, however, repeated experiments may be carried out quickly and with a small additional cost. Therefore, numerous alternative scenarios can be avaluated quickly and at low

cost. The cost of experimental data, however, remains a major expense in terms of time as well as money.

Complexity, Accuracy, and Uncertainty

The cost of information is most acutely felt in the consideration of model complexity. The more complex and detailed a model is, the more data it requires to describe the system and the state of the system. The model developer must carefully evaluate the need for complexity in relation to the cost of the data necessary for meaningful applications. In addition to this data cost, computational costs are directly related to model complexity, and thus increase model operation costs. The introduction of additional complexity does not necessarily increase the accuracy of the simulation. In fact accuracy as well as precision may be adversely affected. Although an increase in accuracy often implies an increase in complexity, the reverse is not true, and a common fallacy is to mistake complexity for accuracy. The accuracy and precision (variability) of modeling results depend directly on these same attributes of the model, parameters, variables, and input data. One of the common pitfalls in using mathematical models is to attribute much greater weight to the model than is warranted by the accuracy and precision of the results. Such over-reliance on the simulation results may be counter productive and lead to misinterpretation of the data. Consideration of accuracy and precision, then, is a very real and vitally important issue. Model users are often placed in situations of great personal and professional risk by models and model results. It is no surprise,

therefore that concern over reliability has begun to be voiced. A measure of the value of the information contained in the model result is needed, and uncertainty analysis meets this need.

Quantification of uncertainty is relatively new to lake modeling research. From early work by Cornell (1972) and Berthouex (1975), many other studies have followed, which address uncertainty analysis in lake phosphorus modeling (Reckhow, 1978b; Lehman, 1978; Lettenmaier and Richey, 1979; Chapra and Reckhow, 1979; Reckhow, 1979b; Reckhow and Chapra, 1979; and Simpson and Reckhow, 1979). More recently, comparative applications of different uncertainty analysis techniques, such as first-order, Extended Kalman Filter, and Monte Carlo simulation have been presented (Scavia, 1980; Scavia et al., 1980).

Traditionally, the term uncertainty has been used to describe variability in situations where too little information is available to quantify that variability. Risk, on the other hand, is used to characterize variability in situations where enough information exists to quantify variability, often in the form of probabilities. In the present context, some liberty is taken with these definitions. Uncertainty will be estimated quantitatively to enable some qualitative comparative risk reduction.

A model's inherent simplification of reality results from the inability to census all information completely through time and space. As a result, all model predictions reflect a certain degree of imprecision or uncertainty. Through the use of procedures such as Monte Carlo simulation, quantification of model prediction uncertainty is possible. This in turn provides yet one more piece of information

which can be used to weigh the value of model outputs. A final consideration is the tradeoff between reliability and time. A decision based on modeling which yields results with relatively broad error bounds but that is ready in time to be used in appropriate planning activities may be preferable to a decision based on a more accurate yet complex model which is not available for planning applications until a later time.

Summary Comments

Mathematical models are powerful tools. By following a few simple guidelines, models can complement the qualitative information used in decision making and thereby provide valuable input to directing the planning process. The following guidelines represent a partial listing of general statements meant to serve as reminders but not as strict rules:

1. Critically define the problem and determine both what questions are to be answered and what information is needed.
2. Use the simplest method that can provide the answers to the questions.
3. Fit the model to the problem, not the problem to the model.
4. Weigh carefully increased time and cost involved in increasing accuracy, and use the simplest model that will yield the desired results.
5. Do not confuse complexity with accuracy.

6. Explicitly note the underlying assumptions of the model and avoid attributing more significance to the results than is actually there.

A model may be thought of as really nothing more than an hypothesis or set of hypotheses which are logically combined into an integrated whole (Chappelle, 1972). The goodness of a model prediction depends on how well these guidelines are used in formulating the appropriate hypothesis and connecting elements. Most important of all, it must be recognized that mathematical models, when used properly, can expand the range of alternatives for consideration, and assist in providing information in an organized form. Models are not the panacea for resource information systems, however. When used within their limitations, models are at best tools to assist in the difficult task of evaluating alternatives. They are not a substitute for experience and good judgment, rather they are a means for permitting these qualities to be used more effectively.

CHAPTER II

MONTE CARLO SIMULATION

Monte Carlo simulation has broad application potential as a research tool. Any system which can be characterized by parameters that exhibit some degree of variability is a candidate for Monte Carlo analysis. By allowing explicit treatment of parameter variability and error, this technique is an ideal tool for combining estimates of error from all sources in an examination of model prediction error.

The Monte Carlo Procedure

Monte Carlo simulation presumes construction of a mathematical model, which describes the stochastic behavior of the variables in the process under study. By characterizing each of these variables as a probability density function (an explicit representation of the variability and error) and not as a single value, all known components of the uncertainty associated with the model prediction are internalized. Repeatedly selecting a random value from the distribution representing each term, and using each of these randomly selected values in the model to calculate a predicted value of the dependent variable, results in a distribution of dependent variable predictions which reflects the combined uncertainties. This resulting dependent variable distribution allows evaluation of the potential impacts on the dependent variable as the result of perturbations to the system, just

as a single value would. However, the frequency distribution also quantifies the value of the prediction by indicating the degree of prediction precision (reliability of the information), represented by the distributional shape and spread. In addition, this information can be used to evaluate alternative models by comparing the precision of the resulting distributions.

Distribution Selection

Before beginning this procedure, selection from among a myriad of possible probability density functions for each parameter is necessary. McGrath and Irving (1973) provide an excellent review of this subject. It is vitally important to know how the particular process, which a variable represents, relates to the entire model in order to select an appropriate probability distribution for any given random variable. Each of the following points must be carefully considered:

1. the underlying theory of each process or event
2. data representing the variability of the process
3. sensitivity of the process to probable values of the variable
4. programming considerations

When the variable under consideration is just one of many which affects the overall problem or system, the simulation is often not very sensitive to the choice of the distribution. For example, in summing a series of random variables, none of which dominate the sum, the total will tend to have a normal distribution, irrespective of the nature of the individual distributions. On the other hand, when only

a few variables dominate the process, or the process is greatly influenced by rare occurances, the selection of probability distributions becomes critical to effective simulation.

A balance must be struck between theoretical justification and empirical evidence in the selection of the appropriate form of probability distribution. Typically, some form of parametric distribution can be justified. Available data can then be used to estimate its parameters. If no empirical data are available, theory and intuition must suffice for selection. Carrying out sensitivity analyses using several different distributions is another means of selecting the appropriate distribution in the absence of empirical data. If abundant empirical data are available, the histogram or more elaborate parametric models can be used. The final choice of a particular distribution must also depend on the relative ease of implementation. Computer storage requirements, length of computation, and programming difficulty are also key considerations.

Generating a random variable from a simple parametric distribution is a relatively simple procedure. Histograms, too, are fairly easily incorporated into Monte Carlo simulations. For more complex distributions, however, simple procedures for generating random numbers are not available, and other more lengthy computational algorithms must be used. In these cases, a compromise must be made between the time and cost of complex computation and simulation rigor. Consideration of the simulation's sensitivity to individual probability distribution assumptions should be foremost in this compromise.

Bases for Distribution Selection

Underlying the distribution selection process are two concepts: the extent to which the qualitative nature of the process is known, and the amount of quantitative data that are available. It is possible in some cases, for instance, to characterize a certain process as normal based solely on a firm understanding of particular characteristics and behavior. Of course, little or nothing may be known of the process. In these cases, data must be relied on for supporting the selection of one distribution over another. The amount of data, however, may range from extensive (readily characterizing the variable), to none. Each case is unique in its particular combination of qualitative and quantitative information, and therefore should be judged individually.

Qualitative Basis

Efforts to establish a qualitative basis for distribution selection are generally based on the following:

1. similarity to a known process
2. underlying theory
3. certain other qualitative aspects

A process may be similar enough to one whose behavior is well known to justify use of the known characterization for the process under study. McGrath and Irving (1973) add that even though the specific situations may not be particularly similar, an assumption of similar process may be reasonable. That is to say, that even though the particular events seem to bear no resemblance, they may share a

behavioral similarity which permits the assumption that the two events may be characterized by representatives from the same family of distributions.

Many types of processes which are modeled stochastically have been characterized by examining the underlying theory of the process. The failure of electrical components have been widely assumed to follow exponential or Weibull distributions (Weibull, 1951). The deviation of shots from a bulls-eye is supported as having a Maxwell distribution in three dimensions and a Rayleigh distribution to two dimensions (Kendall and Stuart, 1958). The exponential distribution reflects reliability and queueing phenomena, as well as characterizing random arrival times (Goodman, 1979).

There are cases, however, for which little is known of the theory of the process, and the process bears no discernable relation to any process whose behavior can be described. There are certain other qualitative aspects which may serve as clues for the identification of an applicable distribution in these cases. This is particularly true if something is known of the behavior of the process or if some data is available. Although probably not sufficient for positive identification, consideration of whether the variable is discrete or continuous, symmetric, bounded, or can be otherwise characterized, can be useful in making a reasonable selection of a distribution.

Quantitative Basis

The amount of data available is perhaps the most important consideration in the selection of probability distributions. Very

often, not having or not being able to collect the data necessary to describe a particular variable is the primary constraint. The collection process may be too lengthy or costly, or in some cases impossible. When sufficient data are available, however, an empirical approach can be used. In essence, this involves using the data to derive a characteristic distribution which represents the variable's behavior. When insufficient data are available, and acquisition is difficult or impossible, justification of the additional resources expended in further data gathering may also be difficult. This is especially true if a reasonable distribution can be selected using the limited data. However, if a workable distribution cannot be identified, additional data may be essential for selecting a valid distribution. This is of particularly great concern if the results of the simulation are highly sensitive to that variable.

Quantitative Techniques

The basis for selecting a specific stochastic model, then, depends on the amount of qualitative and quantitative information available. Of course, this information may vary from nothing to overabundance. In this latter case, almost certain characterization of the process is possible based on sound theory and empirical observation. Development of the underlying theory traditionally involves chains of inductive and deductive reasoning. Although generation of these logical pathways often involves substantial quantitative information, the secondary use of this information is primarily qualitative. Developing empirical evidence, on the other hand, may require

a number of quantitative methods. Among these, sensitivity analysis, graphical analysis, parameter estimation, and goodness-of-fit testing are most common.

Sensitivity Analysis

Sensitivity analysis is performed to determine the extent to which one particular variable or assumption impacts the outcome of the analysis. This technique can be useful in determining if the behavior of the random component must be accurately described, especially when very little characterizing information is available. By varying the values or assumptions of the variables in question, significant differences may be revealed using standard statistical tests. This should never involve lengthy or tedious labor. Most of the variables' behavior must be characterized in the first place, if the simulation is to be relied on. If many of the variables are not accurately described, the simulation lacks validity.

Graphical Analysis

Frequency histograms are one means for identifying appropriate distribution models under the proper circumstances. Foremost, the modeler may benefit from the relative simplicity of histograms. They plot the frequency with which each value or class of values occurs in the sample data. In addition to being used directly in the simulation as the stochastic model of a particular process, these frequency plots provide a visual model of the distribution shape and thus can be useful in selecting an appropriate distribution. For many

applications, the use of a histogram is acceptable if abundant data are incorporated. Care must be exercised, however, in removing obvious errors while retaining all elements of the data set which reflect the existence of low probability events. This graphical technique is most applicable in cases of limited theoretical information and at least some quantitative support. Constructed using all the data collected for a given variable, the histogram presents only the empirical information available. No assumptions of probable behavior are made. Also, by using all the data, errors in the observation set are perpetuated, and may contribute to the inaccuracy of the simulation results. Of special significance is the case of very limited data. Under this circumstance, accuracy is especially heavily influenced by sampling irregularities and from a lack of low probability events being represented. When this is the case, other qualitative considerations (ranging from underlying theory to intuition) may be used to develop better stochastic representations of the variable.

Parameter Estimation

The concept of parameter estimation for the development of a parametric distribution is another common approach to variable characterization. A parametric distribution is defined as either a functional or analytical representation of a probability distribution which depends on one or more parameters (Hastings and Peacock, 1975).

Several conveniences and advantages support the use of a parametric distribution function rather than an histogram, despite the additional effort required for parameter estimation (McGrath and Irving, 1973). Among other advantages, use of a parametric distribution:

1. allows reasonable models to be developed in cases of no data or very limited data,
2. allows incorporation of additional information inherent in the shape of the distribution (e.g., continuity, bounds, and symmetry) if there is theoretical justification,
3. allows meaningful extrapolation into the tail(s) of the distribution, and into other regions where no data are available,
4. provides a reproducible means of representing the data since manual fits to the same data will vary, and
5. provides important summary information about the variable in the form of estimated parameters of the fitted distribution.

In addition, if the random variable may be characterized by a "simple" parametric distribution, random number generation is facilitated.

Simple Parametric Distributions

Simple parametric distributions are conceived as being those which are commonly encountered, relatively easy to recognize, and have some theoretical basis for their functional form and application.

Also,

one c

distr

empir

by ei

cal i

tions

this

eleme

tend

distr

are c

are a

distr

appea

distr

Funct

Param

appli

the r

avail

flexi

Also, somewhat arbitrarily, only those which are characterized by one or two parameters are considered simple, In general, these distributions may often be derived from theoretical assumptions or empirically-based graphical evidence, and are justified for selection by either knowledge of the underlying theory or by preliminary graphical information.

Perhaps the most used and useful of the simple probability functions is the normal distribution. Based on the central limit theorem, this function of course assumes that, if none of the independent elements of a sum dominates, the variable representing the sum will tend toward normality.

Complex Parametric Distributions

In contrast to the simple parametric distributions, the complex distribution families lack well-defined physical interpretations and are difficult to express in simple functional form. Rather, they are abstractions which provide greater flexibility than the simple distributions, and better allow projection of events which would appear in the distribution tails. The Weibull, Johnson, and Pearson distribution families commonly represent the complex distributions. Functionally complex, all these distributions require three to five parameters for specific characterization. They are especially applicable when a simple distribution cannot be justified, and when the results depend on rare events, for which insufficient data are available to accurately define the tail regions. These families are flexible enough to include a number of shapes. This flexibility

allows a reasonable fit to any set of observations. Most simple parametric distributions are in fact special cases of these complex parametric distributions families.

Goodness-of-Fit Testing

Following the preliminary distribution selection process, goodness-of-fit testing should be conducted to substantiate the choice. It is especially important to validate these choices if it is determined that the Monte Carlo result will be sensitive to the distribution selection. Sokal and Rohlf (1969) present an excellent, detailed review of goodness-of-fit testing. The Chi-Square test and the Kolmogorov-Smirnov test are most popular due to their wide range of application. These tests evaluate whether or not a group of data supports the assumption that the hypothesized random variable has come from the assumed probability distribution. The underlying assumptions here are particularly important. Since the statistical inferences based on these tests rely on asymptotic properties, substantial data are required to obtain valid interpretations. The more data there are, the better the chances are of rejecting an inadequate distribution. Perhaps more importantly, the converse is also true; the less data there are, the better the chances are of failing to reject an inadequate distribution. In general, the Kolmogorov-Smirnov test is more sensitive than the Chi-Square test, and does not require arbitrary grouping. In addition to these widely applicable tests, several distribution specific tests are available such as the W-test for a normal distribution and the WE-test for an exponential distribution.

CHAPTER III

THE SIMULATION MODEL APPLICATION

The Model Expression

The model expression used in this simulation is an empirical, black box, lake phosphorus model. Such models are especially applicable in situations which allow a high degree of aggregation, both temporal and spatial (Reckhow, 1979a). The black box label results from the basic assumption of all these models that the system (the lake in this instance) may be treated as a black box. Simulation of internal lake mechanics is not attempted. Only processes which occur at the system boundaries are modeled; material input and output, and interface interaction.

Being less complex mathematically than mechanistic simulations, empirical black box models are appealing to planners because they are often more compatible with the level of mathematical training and the availability of financial resources. Additionally, the relatively straightforward application of uncertainty analysis techniques to these models has encouraged analysis of prediction precision by model users who would otherwise work with deterministic methods only (Reckhow et al., 1980b).

Reckhow (1979a) reviews those basic model forms which empirically address lake phosphorus concentration. The first is a steady state form derived from the definition of R_p, the fraction of influent

phosphorus retained in the lake. This term is defined as follows:

$$R_p = \frac{M - QP_o}{M} \qquad (3\text{-}1)$$

where, on an annual basis,

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

Q = annual volume rate of water outflow from lake (10^6m^3/yr)

P_o = average outflow phosphorus concentration (mg/l).

With M = LA and Q/A = z/τ, equation 3-1 becomes:

$$R_p = \frac{L - (z/\tau)\, P}{L}$$

$$= \frac{\frac{L\tau}{z} - P_o}{\frac{L\tau}{z}}$$

where: L = annual areal phosphorus loading (g/m^2-yr)

z = lake mean depth (m.)

τ = hydraulic detention time (yr)

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

$\frac{L\tau}{z}$ = average influent phosphorus concentration (mg/l)

P = annual in-lake phosphorus concentration (mg/l)

If it is assumed that in-lake and lake outflow concentrations are equal,

$$P_o = P = \frac{L\tau}{z}(1 - R_p) \tag{3-2}$$

The remaining model forms are based on the following phosphorus mass balance (Vollenweider, 1969):

$$V\frac{dP}{dt} = M - \sigma PV - QP \tag{3-3}$$

where: V = lake volume ($10^6 m^3$)

σ = sedimentation coefficient (yr^{-1}).

In essence, this states that the change in phosphorus mass in the lake (VdP) per unit time (dt) is equal to the mass input of phosphorus (M) minus the mass output of phosphorus via the outflow (QP) minus the net mass of phosphorus deposited to the sediments. This sediment "sink" term (σPV) assumes the rate of phosphorus deposition to the sediments to be proportional to the total mass of phosphorus in the lake (PV). Also, as in the first instance, the lake and outflow concentrations are assumed to be equal.

When $\frac{dP}{dt} = 0$, the steady state solution is

$$P = \frac{L}{z(1/\tau + \sigma)} \tag{3-4}$$

The time-dependent solution is as follows:

$$P_t = \frac{L}{\sigma z + z/\tau}\left[1 - e^{-(1/\tau + \sigma)\,\Delta t}\right] + P_{t-1}e^{-(1/\tau + \sigma)\,\Delta t} \tag{3-5}$$

where: Δt = time step (yrs.)

Alternatively, the phosphorus mass balance may be expressed as:

$$V\frac{dP}{dt} = M - v_s PA - QP \qquad (3\text{-}6)$$

where: v_s = apparent settling velocity (m/yr)

This differs from equation 3-3 in the sediment sink term by employing an areal sink, which expresses the rate of deposition to the sediments as a function of the bottom (surface) area. The resulting steady-state expression is as follows:

$$P = \frac{L}{v_s + z/\tau} \qquad (3\text{-}7)$$

where: $z/\tau = q_s$ = areal water loading (m/yr)

Therefore,

$$P = \frac{L}{v_s + q_s} \qquad (3\text{-}8)$$

with the time dependent solution

$$P_i = \frac{L}{v_s + q_s}\left[1 - e^{-(1/\tau + v_s/z)\Delta t}\right] + P_{i-1}e^{-(1/\tau + v_s/a)\ \Delta t} \qquad (3\text{-}9)$$

The major difference between these last two model forms is the manner in which the settling velocity is expressed. The first (3-4), is based on an assumption of depth-dependent settling velocity (σz), while the other model (3-7, 3-8) assumes a constant settling velocity (v_s). It is the time-dependent solution of this last model form (3-9) which is used as the experimental base for this application of Monte

Carlo simulation to lake model uncertainty analysis.

System Representation

As mentioned above, empirical black box lake models do not attempt to represent internal lake mechanics. Rather, they focus on material movement into and out from the system, and activity at the system interfaces. It is not surprising therefore, to find that this model addresses in-lake phosphorus concentration primarily through modeling the interrelationship among phosphorus flux via the lake interfaces (L), settling velocity (v_s) and water loading (q_s).

Distributions and Random Number Generation

Fifteen parameters are characterized using random variables. Six are represented as log normal variates:

1. runoff concentration-urban
2. runoff concentration-forest
3. runoff concentration-agriculture
4. atmospheric flux-urban
5. atmospheric flux-forest
6. atmospheric flux-agriculture.

The remaining nine are represented as normal variates:

7. point flux-primary treatment
8. point flux-secondary treatment
9. point flux-phosphorus removal
10. water load-atmospheric
11. water load-Ontario basin tributaries

12. water load-Niagra River/Welland Canal
13. concentration-Niagra River/Welland Canal
14. settling velocity
15. hydraulic detention time.

All random number generation is accomplished through the use of International Mathematical and Statistical Library (IMSL) subroutines; subroutine GGNML for normal random variates, and GGNLG for log-normal random variates (IMSL, 1979).

Goodness-of-fit testing, as described earlier, is not performed to verify the selection of these parametric probability density functions. Severely limiting data restrictions, push the selection criteria balance toward the qualitative. Consequently, the generated data does not always fit the empirical histogram exactly. In this instance of limited empirical information, knowledge of the underlying theory, as well as intuition have been employed to supplement the existing empirical observations. That is to say, the parameter distributions generated by the simulation are meant to approximate, as reasonably as possible, the actual behavior of the parameter and not to exactly match the date.

Phosphorus Loading

This simulation (see Appendix E) places greatest emphasis on characterizing the loading term (L). This loading term is represented as the sum of the following individual terms:

1a) agricultural runoff concentration * tributary flow rate
1b) urban runoff concentration * tributary flow rate

1c) forest runoff concentration * tributary flow rate

1a + 1b + 1c = diffuse source flux (DIFFLX)

2a) atmospheric flux to agricultural land use

2b) atmospheric flux to urban land use

2c) atmospheric flux to forested land use

2a + 2b + 2c = atmospheric flux (ATMFLX)

3a) point source flux attributable to primary wastewater treatment

3b) point source flux attributable to secondary wastewater treatment

3c) point source flux attributable to phosphorus removal wastewater treatment

3a + 3b + 3c = point source flux (PNTFLX)

4) concentration from Lake Erie * flow from Lake Erie

= flux from Lake Erie (EREFLX)

Therefore,

$$L = (DIFFLX + ATMFLX + PNTFLX + EREFLX)/LSA$$

where LSA = lake surface area.

Diffuse Source Flux

Diffuse source flux is obtained be selecting random values from the runoff concentration distribution for urban, forested and agricultural land uses. Each of these random concentration values are then multiplied by a randomly selected value from the distribution of Lake Ontario tributary flows. The runoff concentrations are represented by a log normal variate for each of the three land use types. The

distributions of these random variates (Figures A.1, A.2, A.3) are characterized by means and standard deviations derived from relevant data aggregated from the literature (Beaulac, 1980). For further details regarding the selection of these data, see Reckhow et al. (1981f).

Tributary flows, defined as all those tributaries in the Lake Ontario watershed other than the Niagara-Welland complex, are characterized by a normal random variate defining a distribution (Figure A.8) with mean and standard deviation calculated from flow data of several tributaries (Chapra, 1979). The cross-correlations among tributary flows (Reckhow et al., 1981g) are sufficiently high to allow considerations of all tributary flow (other than Niagara-Welland Canal flow) by a single distribution. It is not unreasonable to believe that these tributary flows are correlated with the corresponding atmospheric water loading values (i.e., the more rain that falls, the higher the tributary flows are expected). Future consideration should be given to incorporating this relationship into the model. One procedure in particular (Fiering, 1967) seems appropriate for use in this application. Preliminary analysis with limited data, however, does not indicate a strong correlation ($r = .60$).

The size of each land use fraction over the 40 year experimental period (1980-2020) is expected to change relative to the others, and therefore alters the total flux contribution from each land use fraction during any given year. The simulation mechanics account for this change by increasing or decreasing the relative size of each land use fraction according to projected land use information

(see Table 3-1). The projected difference for each land use fraction (annual loss or gain) 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 is done for the second 20-year period (2000-2020) using the projected loss or gain for each land use during this period.

Atmospheric Flux

Atmospheric flux is obtained by summing random values from three log normal atmospheric flux distributions. These distributions are defined by the mean and standard deviation of data characterizing this flux source for each of the three land use categories (Figures A.4, A.5, A.6). The data used (Reckhow et al., 1980a) reflect total bulk loads. This includes both solution phase and dry fallout components. It should be noted at this point that the dry fallout component may be a substantial and significant portion of total atmospheric flux (Reckhow et al., 1980a). All values in Figure A.4, A.5, and A.6 are measurements from terrestrial stations, however, and as such, probably incorporate much higher levels of dry fallout than is representative of mid-lake stations. Some appropriate reduction in these measurements, therefore, may be in order. Until further information is available, however, this lack of detailed information plus the broad inherent variability make any reduction at this point speculation.

The relative weighting of atmospheric flux for each land use

Table 3.1: Land Use Projections for the Great Lakes Basin (10^3ha.) (IJC, 1977a)

	URBAN	AGRICULTURE	FOREST
1980	3416.4	16446.9	34459.4
% Total	6.29%	30.28%	63.43%
Δ%	+10.8%	-11.5%	+1.3%
2000	3933.6	15818.4	34325.1
% Total	7.274%	29.252%	63.474%
Δ%	+15.14%	-3.82%	-.004%
2020	4212.8	15497.8	33821.5
% Total	7.87%	28.95%	63.18%
Δ%	+7.1%	-2.03%	-1.47%
Net Δ% 1980-2020	+23.31%	-5.77%	-1.85%

category is determined using an average land use fraction, obtained by determining the average relative size of each land use fraction over the entire 40-year experimental period. This average fraction size is used in each year's calculation of relative land use contributions, unlike the weighting for atmospheric water loading values for each land use which changes on an annual basis.

Point Source Flux

Point source flux is a major contributor to the total phosphorus load to Lake Ontario. In order to facilitate examination of several different policy scenarios involving changes in waste water treatment strategies, and therefore point source loads, this flux source is subdivided into primary treatment, secondary treatment, and phosphorus removal. Each of these subcategories is represented by a normal distribution defined by a mean determined from survey data (De Pinto et al., 1980). The variability of these point flux contributions is estimated by standard deviations about the mean concentrations for each treatment type (Reckhow, 1978a). Two points which are deserving of future consideration are: 1) the effect of population shifts on point source loads, and 2) the effect of the Finger Lakes and other watershed phosphorus traps (e.g., wetlands) on flux rates from point tributary sources. Population shifts will determine to a great extent the relative fractions serviced by primary, secondary, and phosphorus removal treatments. Additionally, point tributary loads upstream from the Finger Lakes will be affected to a certain degree by these lakes. Phosphorus flux, especially in particulate form, will be attenuated to

some extent by these water bodies, but exactly to what extent is not known.

Lake Erie Flux

Flux from Lake Erie is the final component of the phosphorus loading term (L). It is calculated by multiplying a selected random value from a normal concentration distribution by a lag-one auto-correlated flow value. The concentration distribution (Figure A.10) is defined by a mean and standard deviation from annual concentration data (Chapra, 1979). The flow data is represented by an autocorrelation flow model (Reckhow et al., 1981h). To incorporate this feature, a lag-one Markov model is employed to longitudinally correlate the flows. With flow initialized at the historical mean of 1.85×10^{11} m^3/yr (Chapra, 1979), this model combines 80% of the previous year's flow, 20% of the historical mean, plus a random component based on the standard deviation and correlation coefficient of the historical data (Figure A.9) to calculate the lag-one flow value.

Some controversy exists over which concentration data most accurately reflect the flux values when multiplied by the lag-one flows. Since flows are measured in the Niagara-Welland Complex, Niagara-Welland concentrations can easily be justified. Chapra (1980), however, notes that severe shoreline erosion contributes an inordinate amount of particle-bound phosphorus to the inflow, and suggests use of eastern Erie concentration data as more representative of loadings that determine Lake Ontario concentrations. The particle-bound phosphorus is included in the total-P concentration measurements

taken in the Niagara-Welland area, and yet is thought not to contribute greatly to the in-lake concentration due to rapid settling near the inflow. Estimates of mid-lake availability vary, and indicate a need for examining this question further (see Reckhow et al., 1981i for a discussion of phosphorus fractions and availability). In an effort to resolve this controversy, runs using both Niagara-Welland values ($\bar{x}$ = .022, s = .002) and eastern Erie values ($\bar{x}$ = .017, s = .004) are analyzed.

Settling Velocity

The second major term in the model expression is v_s, the apparent settling velocity. This term is obtained by selecting random values from a normal distribution (Figure A.11), defined by a mean and standard deviation calculated from Lake Ontario data (Chapra, 1979). This term is additionally important in that it serves as the repository for model standard error. This error term is incorporated in the standard deviation of the settling velocity term, and is approximately equal to 2.7 g/m^3. It is calculated by propagating model error for P through the steady state model, fit from Lake Ontario data (see Equation 3-8). When re-expressed as error in v_s, this estimate for model error in the Monte Carlo simulation tests the degree to which a constant v_s is appropriate for modeling Lake Ontario. In essence then, it may be thought of as measuring the lack of model fit for a constant settling velocity model.

Areal Water Loading

The last major term in the model expression is the areal water loading term (q_s), and $q_s = Q/A$, where Q is the sum of the following terms:

1. Ontario tributary flows
2. Flow from Lake Erie
3. Atmospheric water loading

Annual water load from all Ontario basin tributaries other than the Niagara River and the Welland Canal is calculated as described above (see Phosphorus Loading: Diffuse Source Flux). In summary, all tributary flows are characterized by a single normal distribution since cross-correlations are high (Reckhow et al., 1981g).

In addition, total water load is considered to be similar per unit area of land surface, regardless of the land use (Reckhow et al., 1981e). The water load from Lake Erie is calculated as described above (see Phosphorus Loading: Lake Erie Flux) from Lake Ontario data (Chapra, 1979). Summarizing, the water load is represented as a lag-one auto-correlated parameter value. The Markov model combines 80% of the previous year's flow, 20% of the historical mean, plus a random component based on the standard deviation coefficient to derive the new flow (Reckhow et al., 1981h).

Atmospheric water loading from precipitation is calculated from historical rainfall data (World Weather Records), obtained from the meteorologic records for the Ontario basin. Random values are generated from a normal distribution (Figure A.7) defined by the mean and standard deviation which characterize these data.

Additional Model Terms

To fill out the model expression, lake mean depth (z), hydraulic detention time (τ), lag-one phosphorus concentration (P_{i-1}), and time step (Δt) are needed.

Hydraulic detention time is represented by a normal random variable selected from a distribution (Figure A.12) defined by a mean and standard deviation calculated from lake volume and outflow data provided by Chapra (1979). Twenty years of outflow data were each divided by the respective year's lake volume to arrive at a distribution of hydraulic detention times.

For the entire exercise, Δt remains one year. This of course implies annual data only, and points to a shortcoming of the model. By modeling annual loading and flow processes only, many planning objectives (especially those which hold seasonal occurrences as key events) cannot be addressed. However, taking into account the primary purpose of this study, to examine the application of uncertainty analysis on a lake model using Monte Carlo simulation, and the severe data restrictions which would be encountered in an attempt to simulate seasonal processes as well as the fact that much useful planning information can be obtained from a study of average annual conditions, this approach seems valid.

The lag-one phosphorus concentration is selected from a distribution of phosphorus concentration prediction values from the previous year (P_{i-1}). This concentration value is randomly generated from a normal distribution defined by the mean and standard deviation of P_{i-1} predictions and then used to calculate the present year's

prediction (P_i). It is initialized (P_o) at .020 mg/l. Lake mean depth is set deterministically at 89 meters (Snodgrass, 1974), and used for all runs.

Simulation Flow and Structure

All the values are substituted into the model expression (Equation 3-9) for calculation of a Lake Ontario phosphorus concentration prediction. The first prediction represents the concentration estimate for the year 1981. Following a number of 1981 runs to achieve a prediction distribution for that year's phosphorus concentration, the land use fractions are incremented as described above and multiple iterations are again performed to achieve a prediction distribution for the year 1982. This is repeated 40 times to include a prediction distribution for the year 2020. The flow of the computer simulation is somewhat different from that described above in order to streamline certain calculations (see Appendix E).

CHAPTER IV

RESULTS

Model Results

The initial test of the model is a 40-year, 100 iteration - per-year run. The phosphorus concentration prediction mean for the 40-year experimental period is 18.2 micrograms per litre (μg/l), with a coefficient of variation for those concentration predictions of 2.75 percent. This measure of variability actually represents the year-to-year variation in the annual phosphorus concentration predictions. The 14.3 percent mean annual coefficient of variation and the 2.6 μg/l mean annual standard deviation serve to estimate within-year variability, which reflects combined input errors and variabilities. Figure B.3 demonstrates the slight decreasing trend with the slope of the regression line equal to -0.0194. The annual predictions for the apparent settling velocity ($\bar{x}$ = 16.0367 m/yr), areal water load ($\bar{x}$ = 12.0348 m/yr), and the phosphorus loading ($\bar{x}$ = .5108 g/m^2-yr) terms are summarized in Table C.1.

Flux

Figure B.1 graphically presents the breakdown of phosphorus flux sources. Diffuse source flux comprises 32.5 percent of the total phosphorus flux to the lake, with a 40-year mean of 3.161 x 10^9 g/yr. The coefficient of variation for the concentration mean over this same

period is 8.8 percent, making it the most variable of all the delineated flux sources, and the major influence on overall variation for total flux. Point source flux is 17.5 percent of the total flux, with a 40-year mean of 1.703×10^9 g/yr. The variability of this source is 70 percent lower than that for the diffuse sources with a coefficient of variation of 2.5 percent. Phosphorus flux from the atmosphere is the smallest component of total flux at only 8.0 percent with a mean of 7.762×10^8 g/yr. Variability over the experimental period is slightly higher at 3.7 percent than that for point sources. The major source of phosphorus flux to Lake Ontario is from Lake Erie. 42 percent of all the phosphorus input to Lake Ontario comes by way of the Niagara River and Welland Canal (calculated using Niagara-Welland concentrations). The model estimates a 40-year mean to be 4.064×10^9 g/yr with very little variability (coefficient of variation = 1.1 percent). See Table C.2 for a more complete statistical summary of the model's characterization of phosphorus flux to the lake.

Water Load

Water loading is broken down into atmospheric, Ontario basin tributary, and Lake Erie source components (Figure B.2). Atmospheric water loading is the smallest component. The 1.659×10^{10} m^3/yr comprises only 7.0 percent of the total water load of 2.287×10^{11} m^3/yr. The coefficient of variation is 2.0 percent. The variability of the load from Ontario basin tributaries is approximately the same at 2.1 percent, with a 40-year mean of 2.721×10^{10} m^3/yr, or about 12

perce

Lake

water

water

to be

0.3 p

summa

Model

sensi

to th

Numbe

result

shade

deviat

run.)

artifa

aggreg

magnit

either

Variat

percent of the total. The remaining 81 percent of the water load to Lake Ontario comes from Lake Erie. At 1.849×10^{11} m^3/yr, this water loading source is the most significant influence on the overall water load to the system. Additionally, one would expect this load to be consistently a significant influence as indicated by the small 0.3 percent coefficient of variation. Table C.3 presents a tabular summary of the water loading terms.

Model Sensitivities

Tables C.4 and C.4a summarize the results of the three different sensitivity experiments. The first involves the model's sensitivity to the number of annual iterations performed.

Number of Runs

Figures B.3, 4, and 5 are the graphical representations of the results for 100-, 500-, and 1000-iteration runs respectively. The shaded portion reflects variability of plus or minus one standard deviation. (See Appendix D for a discussion of the 1000-iteration run.) There appears to be little sensitivity to this factor other than artifacts of the increasing number of iterations per year. On an aggregate basis, all experiments reflect predictions of comparable magnitude. In addition, no conclusive differences are evidenced by either measure of mean annual prediction variability (coefficient of variation, $\overline{CV}$ and standard deviation, $\bar{s}$).

Distribution Selection

Comparing results for the distribution selection experiments (Table C.4, Figure B.13) reveals a higher prediction for the run using all normal distributions to characterize the model parameters. The average annual variability, however, is much lower than that for the 100 iteration base run ($\overline{CV}$ = .084 as opposed to .144).

Random Seed

Sensitivity to a change in the random seed appears to a limited extent in two of the three cases tested. Table C.4a presents a comparison of these cases. In the case of 500 iterations per year, the general appearance of the time series plot is noticeably altered by changing the random seed from 123457.DO (double precision) to 987543.DO (Figures B.4 and B.4a respectively). The summary statistics of each of these runs, however, do not differ appreciably. The increased random seed results in a mean annual concentration prediction that is approximately 2 percent higher (19.1 μg/l vs. 18.7 μg/l), and a mean annual variability that is just slightly lower (14.3 percent vs. 14.4 percent). In both instances, the concentration appears to change very little over the 40-year experimental period.

In the case of all normal distributions characterizing the parameters, the increase in the magnitude of the random seed from 123457. DO (Figure B.13) to 987225.DO (Figure B.13a) again results in a distinct change in the overall appearance of the time series plot. The larger random seed results in an average annual concentration prediction just less than 9 percent greater than that for the small random

seed (21.1 µg/l vs. 19.4 µg/l). The average annual variability is comparable for the two runs with the larger random seed resulting in a somewhat smaller value. In addition, the tendency of the concentration prediction to increase or decrease over time is altered by the change in random seed; in this instance changing from a negative slope when the smaller random seed is used, to a positive slope, when the larger random seed is used.

Unlike the first two cases, changing the random seed for the 1000 iteration-per-year run appears to have little effect on either the general appearance of the time series plot or the magnitude of the predicted concentrations (Figures B.5, B.5d). Table C.4a does, however, point out an increase in the average annual variability of approximately 3.5 percent when the larger random seed is used. (See Appendix D for a discussion of the periodicity present in the 1000 iteration-per-year runs.)

Planning Scenarios

Testing the effects of planning scenarios is one of the primary advantages of modeling. Tables C.5 and C.5a summarize the results of eight experimental scenarios.

Point Source Flux

The first two involve changing the nature of the point source flux term by: a) reducing point source flux variability, and b) upgrading all wastewater treatment facilities to phosphorus removal status. The first case attempts to experimentally test whether or not regulations

designed to tighten up operational efficiency, rather than to set upper bounds on discharge, will have any effect on in-lake phosphorus concentration. Figure B.6 indicates no great deviations from the 100 iteration-per-year base run (Figure B.3). In fact, the two runs are identical in many respects. Most importantly, they share a common average annual concentration prediction (18.2 µg/l) as well as a common average annual coefficient of variation (14.3 percent). It should be re-emphasized at this point that this analysis of point source flux variability may be questionable since the characterization of input variabilities is in such doubt. The second case is designed to test the effect of upgrading all wastewater treatment facilities by hypothesizing phosphorus removal treatment for those plants which do not currently employ this treatment (approximately 47% of the total point source contribution). It should be noted at this point that three treatment plants are not included in this treatment change over. Influent/effluent data are not available for these plants, which in fact contribute less than 4% of the total point source load. Figure B.7 displays a remarkable resemblance to both the 100 iteration-per-year base run (Figure B.3) as well as the run using one half of the point source variability (Figure B.6). The major differences are a slightly lower mean annual concentration prediction (17.4 µg/l vs. 18.2 µg/l) and a mean annual coefficient of variation about 4 percent higher than the 14.3 percent variability of the base run.

The effects of these hypothesized changes on the total point source flux is graphically represented in Figure B.8. Total point source flux is defined here to be phosphorus flux from all wastewater

treatment plants with an average daily flow in excess of 1 million gallons per day (mgd). There are 91 such facilities in the Lake Ontario basin which account for well above 90 percent of the actual total point source phosphorus load (DePinto, 1980). Industrial sources are not included, as their contribution is relatively minor (Chapra and Sonzogni, 1979). The reduced variability has very little effect on the total point source flux. A reduction of less than two percent in the average annual flux from 1.70×10^9 g/yr to 1.67 x 10a g/yr results. The upgrading to all phosphorus removal treatment, however, has a marked effect; reducing the mean annual point flux value by 27 percent.

Erie Concentration

The third experimental scenario measures the effect of alternate concentration values used to characterize the flux from Lake Erie. The concentration values used for all simulation runs are those taken from Niagara River-Welland Canal data. Chapra (1979) contends that excessive shoreline erosion in the area of the Niagara-Welland complex may contribute a large portion of particle-bound phosphorus to the lake. How much of this particulate fraction is actually available at mid-lake sites is questionable, however. Chapra therefore suggests that a more accurate estimate of this flux source may be derived using eastern Lake Erie concentration values. The time series plot of this run (Figure B.9) is again very similar in general appearance to the 100 iteration-per-year base run. Since flux from Lake Erie is the largest single source, it is not surprising

to note that changes in this term result in distinct changes in the final concentration prediction distribution (see Table C.5). The mean summarizing the eastern Lake Erie concentration data is 23 percent lower than that used for the Niagara-Welland data (17 µg/l vs. 22 µg/l), and results in an average annual concentration prediction 1.7 µg/l less than the 18.2 µg/l mean value for the Niagara-Welland run. In addition, the eastern Lake Erie data is 15 percent more variable than the Niagara-Welland data. As a consequence, the mean annual coefficient of variation for the entire experimental period rises by 2 percent.

Land Use Fractions

The five remaining experimental planning scenarios test the effects of a) altering the balance of land use fractions, and b) altering the rate of that change (Table C.5a). Doubling, or even tripling the rate of current projected land use pattern shifts has virtually no effect whatsoever on the concentration predictions over the 40-year test period. Graphical representations of these runs are not included in Appendix B. They are identical to the 100 iteration-per-year base run in almost every respect. One slight difference is evident, however. As the rate of change increases, the average annual coefficient of variation decreases slightly from 14.3 percent to 14.1 percent (doubling), and again to 13.8 percent (tripling).

The last three experiments hypothesize land use shifts to all urban, or all agriculture, or all forest by the end of the 40-year

period (Table C.5a). Beginning with the current land use fractions, one land use type is incrementally enlarged until the entire lake basin is covered with this land use at the end of the test period. The other two land use types are incrementally decreased until they occupy none of the lake basin at the end of the test period. The results of these experiments (Figure B.10) dramatically point out the distinct dichotomy between disturbed and undisturbed ecosystems as far as their contribution to lake phosphorus concentration is concerned. The two disturbed systems, urban and agriculture, contribute to increasing in-lake concentrations as their respective fractions increase. Agricultural expansion impacts the concentration more quickly than urban expansion. That is to say, the rate of concentration increase is higher than that for the urban experiment. The projected increase is 1.5 μg/l every 5 years for agriculture, as opposed to 0.4 μg/l every 5 years for urban. Within-year variability for the agriculture projections is more than 1-1/2 times greater than the 13.6 percent average annual coefficient of variation for the urban projections. Increasing the undisturbed system fraction, forest, results in a decrease of the concentration predictions of 0.7 μg/l every 5 years. The average annual variability of these predictions is a very low 10.4 percent.

Variability

Table C.6 summarizes eight representative single year concentration distributions; four from the 100 iteration-per-year base run, and four from the run using all normal distributions to characterize the

model parameters. It is these single year distributions which distinguish stochastic methods such as Monte Carlo simulation from strictly deterministic techniques. While deterministic time series models are able to generate estimates of variability over years, only stochastic processes allow specification of variability for any one year. These distributions (Figure B.11 and B.12) provide a great deal of information concerning the nature of the model prediction. In addition to the customary measure of location, this information allows some feeling for spread as well as third and fourth moments. In the case of the 100 iteration-per-year base run, annual variability is quite large. Actually, for the entire 40-year run, the average annual coefficient of variation ranges from 9.6 percent to 28.4 percent. Skewness may be easily estimated by calculating the mode to mean ratio, R, as follows:

$$R = (1 + CV^2)^{-1.5}$$

The base run distributions consistently demonstrate R-values very close to 1.0. Some positive skewness can be seen in the sample distributions. The run using all normal distributions manifests much lower annual variability. For the entire 40-year run, the average annual coefficient of variation ranges from 6.8 percent to 11.0 percent. In this second case, the R-values approach 1.0 even more closely. The histograms bear this out, being remarkably normal in appearance.

One of the major questions to be answered is whether or not this within-year variation displays any tendency to increase or decrease

over time. Table C.7 summarizes the simulation results of this experiment. Testing four model sensitivity runs results in very small differences in annual variability. Over the 40-year test period, none of these runs demonstrate much of a tendency to increase or decrease prediction variability over time. Regression statistics for prediction standard deviation and time reveal slight decreasing tendencies on the order of 10^{-6} mg/l every year; -13×10^{-6} being the single largest rate of change. Even the most radical planning scenario, the shift to all agricultural land use, produces only a mild 13×10^{-5} mg/l annual increase in the average annual standard deviation.

CHAPTER V

ANALYSIS AND DISCUSSION

A primary concern of this investigation, as with all modeling exercises, is how well this tool characterizes the system it is intended to characterize. The degree to which the accuracy of such a tool may be measured is limited, especially in the predictive work it is most often employed. However, three criteria which may be used for such an evaluation are:

1) Does the simulation generate valid parameter values?
2) Using these parameter values, does the model predict reasonable values for the dependent variable?
3) Does the model react to perturbations with consistency?

The first of these may be easily evaluated. The fifteen parameters that are randomly generated from distributions characterized by literature data (see Chapter III) are of course accurately represented by the simulation. The ten remaining major model parameters, which are calculated parameters, also agree closely with the findings of other Lake Ontario studies (IJC, 1978; Sonzogni et al., 1978; Chapra and Sonzogni, 1979; Chapra, 1980; Simons and Lam, 1980; and Fraser, 1980).

The second point is somewhat more difficult to address. The dependent variable, in-lake phosphorus concentration, is certainly characterized within reason by the model expression. The model

prediction agrees with all the major Lake Ontario studies listed above. However, prediction behavior over the entire 40-year test period must be assessed. In the few years immediately preceding 1980, there seems to be a "significant decreasing trend in total phosphorus concentrations" (IJC, 1977b). The 100 iteration-per-year base run of this simulation agrees very closely with the magnitude of these concentrations, and continues the decreasing trend in the 40-year prediction period. Major responsibility for this decrease is suggested to result from changes in the sedimentation rate of total phosphorus (Fraser, 1980). This of course demands a much clearer picture of the nature of the system's sensitivity to the rate of phosphorus sedimentation than that which the current data provide. If this is the case, however, more sophisticated modeling treatment of the settling velocity term (v_s) is necessary (if meaningful projections are to be made).

The final point is the most difficult to assess in the present context. Experimental perturbations, detailed in Chapter IV, affect the model predictions in different ways, and to different degrees. That is to say, the simulation does not react to these perturbations with any recognizable consistency. The results, however, may indeed be consistent with the behavior of the natural system. This dilemma is due in part to the relatively recent origins of studies of this type that stress prediction uncertainty. There is no cohesive body of experimental information on which to base a set of expectations concerning responses to certain given perturbations.

Model Sensitivities

The first set of sensitivity experiments indicates no substantial effects caused by changing the number of annual iterations performed. Some change in the overall appearance of the time series plots is evident, however, due primarily to the increased probability of generating extreme values when greater numbers of iterations are performed. Aggregate statistics, however, belie this difference.

Revising the simulation to employ only normal distributions to characterize model parameters results in a slightly elevated prediction range and a significantly reduced average annual coefficient of variation (.084 vs. .144). This lower variability and higher prediction average may be explained by the fundamental differences between the manner in which normal and log-normal distributions are treated. Random log-normal parameter values can only be positive by definition. Generation of random normal parameter values, however, involves generating a random standard normal variate which will be negative 50 percent of the time. Since negative concentrations and fluxes are not easily interpreted, all but the positive random numbers are screened out before this standard normal random variate is adjusted using the sample mean and standard deviation. As a result, random values generated from the same population, characterized by each of these two probability functions will result in two different sets of summary statistics due to the inherent differences in the density functions and the simulation's treatment of them. Whether or not this revision to all normal distributions results in a more realistic representation of system behavior is not clear, however. Adequate data for comparison are

not available. This revision, however, does seem to increase the simulation's sensitivity to changes of the random seed.

Although there are certain minor changes in the 500 and 1000 iteration-per-year runs due to changes in the random seed, none appear as significant as those changes in the run using all normal distributions. There is an 8.8 percent increase in the average annual concentration and a complete reversal of the decreasing trend accompany the experimental change of the random seed. This suggests that perhaps the use of normal distributions is less characteristic of the six parameters, previously described by log-normal distributions. The interrelationships among all the parameters, however, are so involved that such a suggestion cannot possibly be validated without considerable future investigation. Since distribution selection is at the heart of Monte Carlo simulation, model sensitivity to various distributions is critical. Future Monte Carlo investigations of Lake Ontario should focus on better characterization of each parameter by a well defined parametric distribution.

Planning Scenarios

Experimental testing of regulations that would require greater consistency in the operation of wastewater treatment plants seems to alter the magnitude of the predictions very little. In fact, a substantial reduction in treatment plant effluent concentration results in only a relatively small, although distinct, 1 μg/l reduction in average annual in-lake phosphorus concentration. In this hypothesized planning scenario, plants accounting for about half of the total point

source load are experimentally upgraded from 47 percent phosphorus concentration reduction (average secondary treatment) to 79 percent concentration reduction (average phosphorus removal treatment). It is difficult, once again, to know with any certainty if these experimental predictions are truly indicative of the system's reactions to these point source flux perturbations. Point source flux is a minor contributor at 17.5 percent of total flux, and as such would not be expected to impact the overall prediction to a major degree. The sparse information currently available on the projected impacts of point source controls on future in-lake phosphorus concentrations for Lake Ontario (see Chapra, 1980), however, suggests more substantial impacts than are evidenced by this study. It is painfully obvious that much more effort needs to be spent in gathering data to better characterize the relationships between treatment plant operation, effluent concentration, and in-lake concentration. In addition, it is quite surprising that the multi-year data that have been gathered are neither centralized nor aggregated in any useful form. As an example, no estimates of treatment plant operational variability are available. In order to simulate the natural variability in the point source flux term on an annual basis, these estimates of how great the variations in concentration reduction are from year to year are vital. It seems a rather simple matter to calculate a standard deviation for the concentration reduction performance over a period of years. Extended studies of this kind which include measures of variability, however, are non-existent (Elridge, 1980; Eastman, 1980; Palancic, 1980; King, 1980; Berthouex, 1980, Heidtke, 1980; and

Hore, 1980).

As introduced in Chapter IV, the use of eastern Lake Erie concentration values to characterize phosphorus flux to Lake Ontario via the Niagara/Welland Complex inlet results in a substantial 2 μg/l decrease in the average magnitude of the annual concentration predictions. This inflow is the single greatest source of phosphorus, representing 42 percent of the total flux to the lake. It is not surprising, then, to find the model prediction directly affected by a reduction in the concentration value used to generate estimates of flux from Lake Erie. Since the use of either concentration value constitutes a major assumption (and a difference of 1.0 to 1.7 billion g/yr), fundamental understanding of why these two terms are different through critical assessment of shoreline erosion at the inlet and bio-availability at mid-lake locations is essential to meaningful planning applications.

The final planning scenarios involve altering the distribution of land use types and the rate at which that distribution changes. The current projections for land use over the next 40 years, however, manifest only minute annual changes of 1/100 to 5/100 on one percent. Therefore, even though diffuse source flux represents 32.5 percent of total flux, doubling, or even tripling, the rate of change has no impact on the predicted concentrations over the test period. Only the drastic and unrealistic shifts to single land use types stimulate noticeable changes. These last changes, however, serve to highlight the distinct dichotomy between the behavior of disturbed versus undisturbed systems, and the difficulty in distinguishing any further

specificity in land use classification. This is to say that, although more detailed distinctions between row crop and non-row crop, pasture, fallow, and feedlot, or between deciduous and coniferous forest, or between industrial and residential land uses may be desirable from the planning perspective, no data are currently available to support any distinction beyond disturbed versus undisturbed.

Variability

Prediction variability within years and over time is the heart of this study. Results of variability experiments are presented in Chapter IV. Table C.7 clearly points out the difference between modeling with uncertainty analysis and modeling without uncertainty analysis. Traditionally, the prediction is characterized by a single value; usually the mean. Through the use of Monte Carlo simulation techniques, a fully descriptive prediction distribution is obtained. This distribution not only allows some estimation of location, but also allows estimation of prediction precision. Qualitatively, this may be accomplished through visual examination of the prediction histogram. Quantitatively, statistics such as standard deviation (s), coefficient of variation (CV), and the mode to mean ratio (R) may be used to estimate higher order moments.

The tendency for variability to increase or decrease over time is unclear from the information generated by this study. Linear regression statistics calculated for the annual prediction standard deviations indicate slightly decreasing slopes; the largest of the

sensitivity runs being 13/1000 of one microgram per year. Even the planning scenario which hypothesizes the shift to all agricultural land use, the most radical change, results in an annual change in standard deviation of only 13/100 of one microgram per year.

Summary

In summary, there are three concrete findings that can be made directly from the results. They are:

1) Upgrading all wastewater treatment plants to phosphorus removal status substantially reduces the magnitude of the overall concentration predictions.
2) The flux from Lake Erie is the single greatest flux source. The decision to use either the Niagara-Welland or the eastern Lake Erie concentration data, therefore, is critical to the final predictions.
3) The only distinction which can be meaningfully made with regard to land use type as it affects phosphorus flux is that between disturbed (urban and agriculture) and undisturbed (forest) systems.

From these findings, several recommendations may be made to aid in the development of priorities for future investigations. First, if uncertainty analysis is to play a major role in improving the predictive capabilities of models, then measures of variability must become a routine element in data collection and presentation. Shortcomings in this area are particularly evident in the context of wastewater treatment plant performance. Second, the controversy over

which Lake Erie concentrations best reflect the actual impact of flux from this source on Lake Ontario must be resolved. Until this question is satisfactorily addressed, Lake Ontario water quality predictions will be subject to this additional uncertainty; a substantial difference of 1.0 - 1.7 x 10^9 grams per year. Third, existing data relating land use type and phosphorus runoff concentration must be augmented if meaningful land use planning is to be accomplished. If no such effort is undertaken, all modeling distinction of land use types beyond disturbed and undisturbed will continue to be artificial.

FINAL COMMENTS

In many ways, this research has uncovered many more questions than it has answered; highlighted more difficulties than it has overcome. It is an interim work, meant to be used as a stepping stone, and to continue the construction of a firm foundation from which this relatively new field may continue to grow.

The ultimate objective of these exercises is to provide the environmental planner with a valuable tool which will enhance the quality of the planning decisions by presenting one more piece of information to be factored into the decision-making process. Those who concern themselves with the model development area exclusively often lose sight of this objective, and begin to view model development as an end in itself rather than as a means to an end. It is essential to keep in mind the eventual real world application. Models are subject to data constraints, computation inadequacies, and training insufficiencies. Often, logical and efficient models are taken too soon from their pristine birthplaces and forced to fend for themselves in the somewhat imperfect real world. After the dust settles these well-intended tools are rejected as impractical or impossible to apply. Those that survive have been developed within the limitations of existing data restrictions. Of course, there is an important and valid function for those models that are developed beyond existing constraints. This simulation, for example, is not yet ready to be

used for sophisticated planning projections. It has, however, performed the necessary function of pointing out priority research areas, and thereby helped focus the scope of future investigations. Continued refinement is essential for this model to realize its ultimate objective.

Above all else, this work is intended to showcase uncertainty analysis, in the form of Monte Carlo simulation, through the use of a practical example. It is abundantly clear that the Lake Ontario data base is not yet comprehensive enough to support a model of even this modest detail with uncertainty analysis. It is equally clear, however, that Monte Carlo techniques in specific, and uncertainty analysis in general, possess enormous potential as modeling aids to decision-making.

BIBLIOGRAPHY

BIBLIOGRAPHY

Beaulac, M.N. 1980. Nutrient Export Coefficients: An Examination of Sampling Design and Natural Variability Within Differing Land Uses. Masters Thesis, Michigan State University. 243 pp.

Berthouex, A.M. 1975. Modeling Concepts Considering Process Performance, Variability, and Uncertainty. In Mathematical Modeling for Water Pollution Control Processes, p. 405-440. Keinath, T.M. and M.P. Wanielista (eds.), Ann Arbor Science Publications, Inc., Ann Arbor, Michigan.

Berthouex, P. 1980. University of Wisconsin-Madison, Department of Civil and Environmental Engineering. Personal Communication.

Chappelle, D.W. 1972. Lecture Notes: Model Building, Hypotheses, Development of Theory. Department of Resource Development, Michigan State University, East Lansing, Michigan.

Chapra, S.C. 1979. National Oceanic and Atmospheric Administration, Ann Arbor, Michigan. Personal Communication.

Chapra, S.C. 1980. Simulation of Recent and Projected Total Phosphorus Trends in Lake Ontario. J. Great Lakes Res. 6(2): 101-112.

Chapra, S.C. and K.H. Reckhow. 1979. Expressing the Phosphorus Loading Concept in Probabilistic Terms. J. Fish. Res. Board Can. 36: 225-229.

Chapra, S.C. and W.C. Sonzogni. 1979. Great Lakes Total Phosphorus Budget for the Mid 1970's. J. Water Pollution Control Fed. 51 (10): 2524-2533.

Cornell, C.A. 1972. First-Order Analysis of Model and Parameter Uncertainty. International Symposium on Uncertainties in Hydrologic and Water Resource Systems. Vol. 3: 1245-1274. Tucson, Arizona.

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.

Eastman, J. 1980. Michigan State University, Department of Civil and Sanitary Engineering. East Lansing, Michigan. Personal Communication.

Eldridge, S. 1980. Michigan Department of Natural Resources, Water Quality Division, Lansing, Michigan. Personal Communication.

Fiering, M.B. 1967. Streamflow Synthesis. Harvard University Press, Cambridge, Massachusetts.

Fraser, A.S. 1980. Changes in Lake Ontario Total Phosphorus Concentration 1976-1978. J. Great Lakes Res. 6(1): 83-87.

Goodman, E. 1979. Department of Electrical Engineering and Systems Science, Michigan State University, East Lansing, Michigan. Personal Communication.

Hastings, N.A.J. and J.B. Peacock. 1975. Statistical Distributions. Halsted Press, New York.

Heidtke, T. 1980. Great Lakes Basin Commission, Ann Arbor, Michigan. Personal Communication.

Hore, R. 1980. Ontario Ministry of the Environment, Toronto, Ontario, Canada. Personal Communication.

IJC. 1977a. International Joint Commission, Land Use and Land Use Practices in the Great Lakes Basin, Summary Report--Task B. Windsor, Ontario, Canada.

IJC. 1977b. International Joint Commission, Great Lakes Water Quality, Appendix B. Windsor, Ontario, Canada.

IJC. 1978. International Joint Commission Technical Group to Review Phosphorus Loadings, Fifth Year Review of Canada - United States Great Lakes Water Quality Agreement - Phosphorus Loadings. Report of Task Group 3, Great Lakes Regional Office, Windsor, Ontario, Canada.

IMSL. 1979. International Mathematical and Statistical Library Reference Manual. Edition 7, Vol. 2, Chapter G.

Kendall, M.G. and A.S. Stuart. 1958. The Advanced Theory of Statistics. Vol. 1, Distribution Theory. Charles Griffen & Co.

King, D. 1980. Michigan State University, Water Resources Research Center, East Lansing, Michigan. Personal Communication.

Lehman, J.T. 1978. Aspects of Nutrient Dynamics in Freshwater Communities. Ph.D. dissertation, University of Washington.

Lettenmaier, D.P. and J.E. Richey. 1979. Use of First-Order Analysis in Estimating Mass Balance Errors and Planning Sampling Activities. In Theory and Systems Analysis in Ecology, p. 80-106. Halfon, F. (ed.), Academic Press, New York.

McGrath, E.J. and D.C. Irving. 1973. Techniques for Efficient Monte Carlo Simulation. Office of Naval Research, National Technical Information Service, AD 762: 721-723 (3 volumes).

Palancic, S. 1980. United States Environmental Protection Agency, Facility Planning Branch, Chicago, Illinois. Personal Communication.

Reckhow, K.H. 1978a. 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. 1978b. Empirical Lake Models for Phosphorus: Development, Applications, Limitations and Uncertainty. In Perspectives on Aquatic Ecosystem Modeling. Scavia, D. and A. Robertson (eds.), Ann Arbor Science Publishers, Inc., Ann Arbor, Michigan.

Reckhow, K.H. 1979a. Quantitative Techniques for the Assessment of Lake Quality. United States Environmental Protection Agency, EPA-440/4-79-015.

Reckhow, K.H. 1979b. Uncertainty Analysis Applied to Vollenweider's Phosphorus Loading Criterion. J. Water Pollution Control Federation. 51(9): 2123-2128.

Reckhow, K.H. and S.C. Chapra. 1979. A Note on Error Analysis for a Phosphorus Retention Model. Water Resources Research. 15(6): 1643-1646.

Reckhow, K.H., M.N. Beaulac, and J.T. Simpson. 1980a. 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., V.D. Lee, and S.C. Chapra. 1980b. An Examination of Lake Model Prediction Uncertainty Using First Order Analysis and Monte Carlo Simulation. Paper presented at the winter meeting of the American Society of Limnology and Oceanography, January, 1980, Los Angeles, California.

Reckhow, K.H., R.E. Ancil, M.N. Beaulac, V.D. Lee, and R.H. Montgomery. 1981a. Methods for the Analysis of Error and Variability in Lake Trophic State Determination. National Oceanic and Atmospheric Administration Completion Report, Contract Number 03-78-b01-109, Appendix F.

Ibid. 1981b. Appendix G.

Ibid. 1981c. Appendix H.

Ibid. 1981d. Appendix I.

Ibid. 1981e. Appendix J.

Scavia, D. 1980. Uncertainty Analysis of a Lake Eutrophication Model. Ph.D. dissertation, University of Michigan.

Scavia, D., W.F. Powers, R.P. Canale, J.L. Moody. 1980. Use of First-Order Error Analysis and Monte Carlo Simulation in Time-Dependent Lake Eutrophication Models. Great Lakes Environmental Research Lab. (NOAA) Contribution #234. Ann Arbor, Michigan.

Simons, T.J. and D.C.L. Lam. 1980. Some Limitations of Water Quality Models for Large Lakes: A Case Study of Lake Ontario. Water Resources Research. 16(1): 105-116.

Simpson, J.T. and K.H. Reckhow. 1979. Assessing Phosphorus Loading on Lakes. Paper presented at the North American Lake Management Conference, Michigan State University. East Lansing, Michigan.

Snodgrass, W.J. 1974. A Predictive Phosphorus Model for Lakes - Development and Testing. Ph.D. dissertation, University of North Carolina, Chapel Hill, North Carolina.

Sokal, R.R. and F.H. Rohlf. 1969. Biometry. W.H. Freeman & Co., San Francisco.

Sonzogni, W.C., J. Montieth, W.N. Bach, and V.G. Hughes. 1978. United States Great Lakes Tributary Loadings. PLUARG Technical Reports, IJC, Great Lakes Regional Office, Windsor, Ontario, Canada.

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.

APPENDIX A

APPENDIX A

GRAPHICAL DATA INPUTS

Each of the twelve figures contained in this appendix displays the data used to formulate the parametric distributions which characterize the model parameters. The block histograms represent the actual data used in the distribution selection process. The continuous curves superimposed on these histograms are the graphical results of randomly generating values from parametric distributions defined by the mean and standard deviation of each of these data sets.

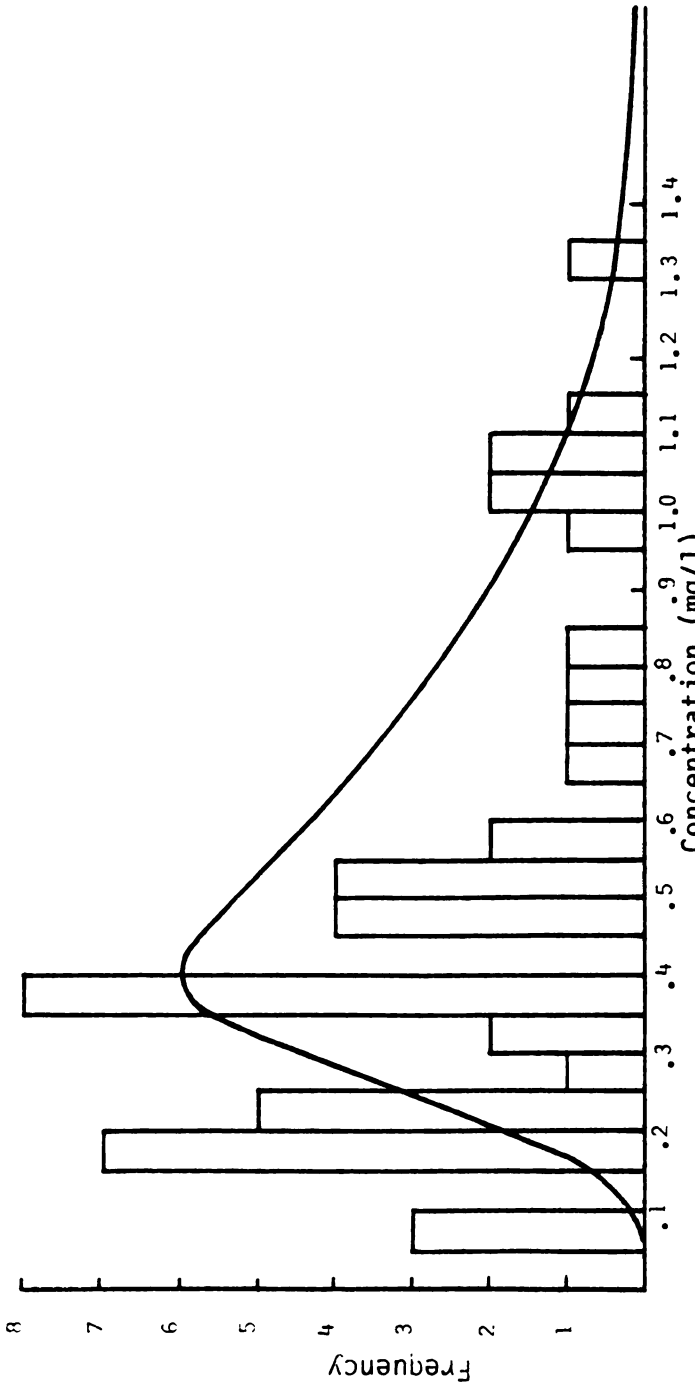

Figure A.1. Total Phosphorus Runoff Concentrations from Urban Watersheds.

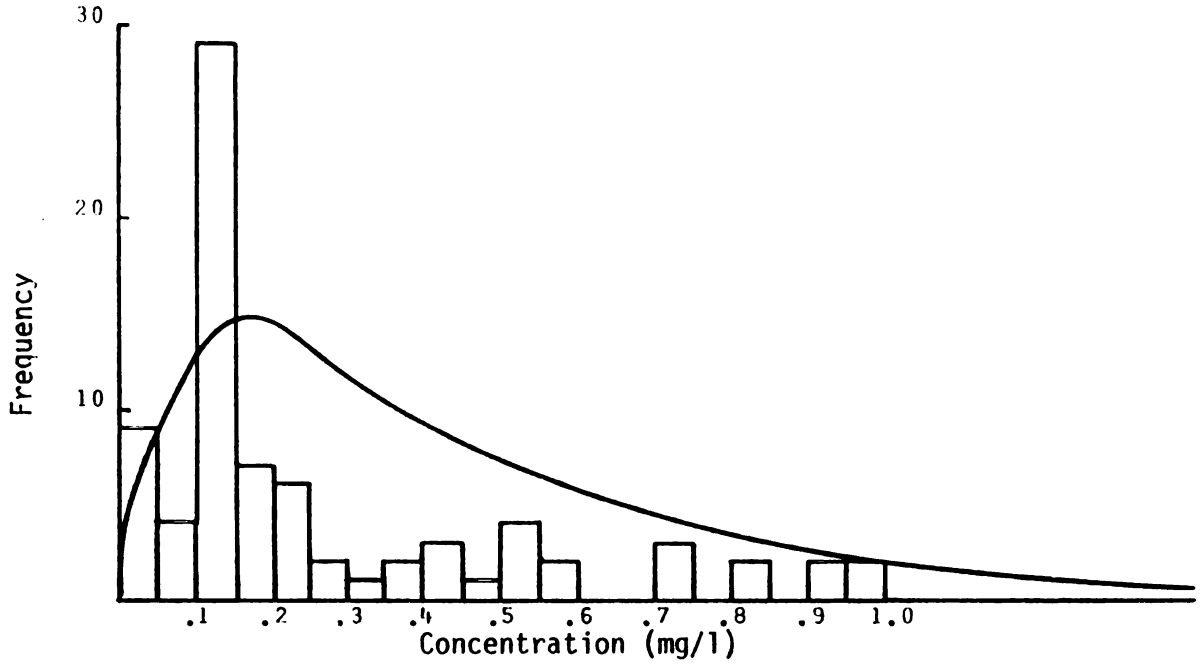

Figure A.2. Total Phosphorus Runoff Concentrations from Agricultural Watersheds.

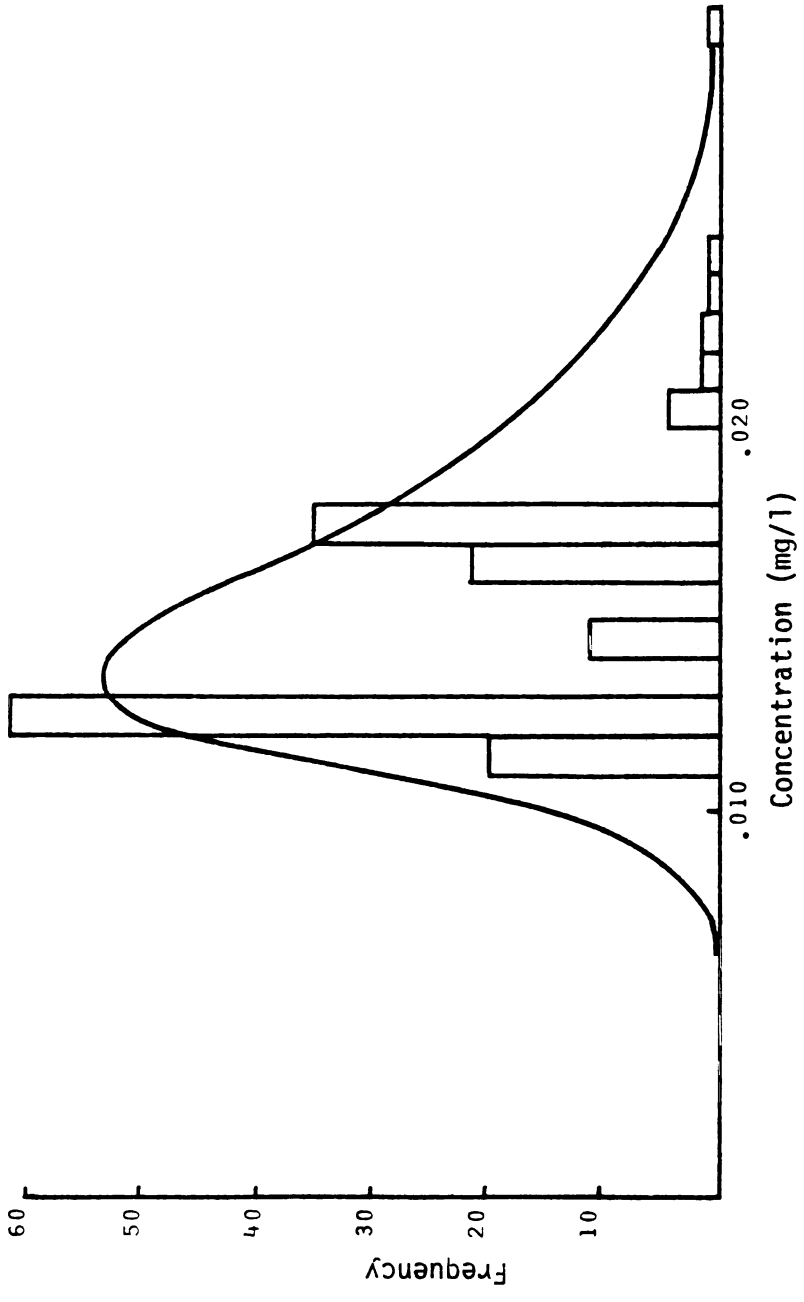

Figure A.3. Total Phosphorus Runoff Concentrations from Forested Watersheds.

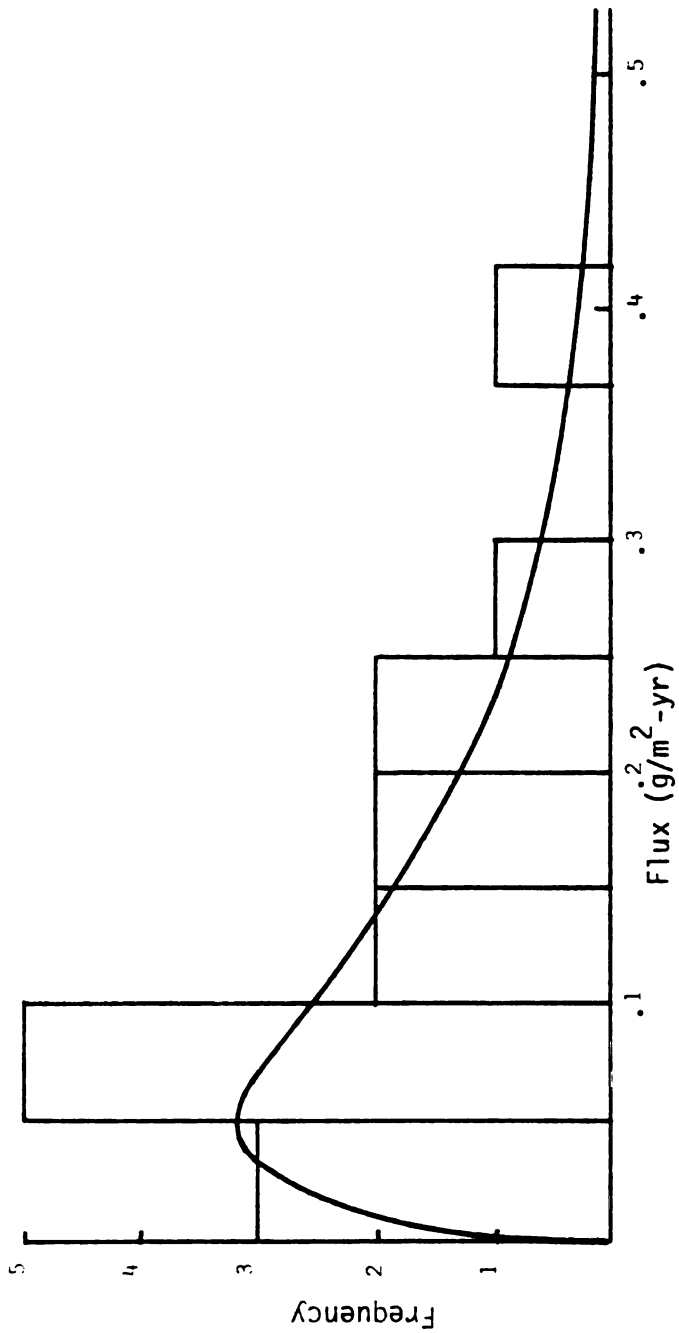

Figure A.4. Total Phosphorus Flux from the Atmosphere to Urbanized Land Uses.

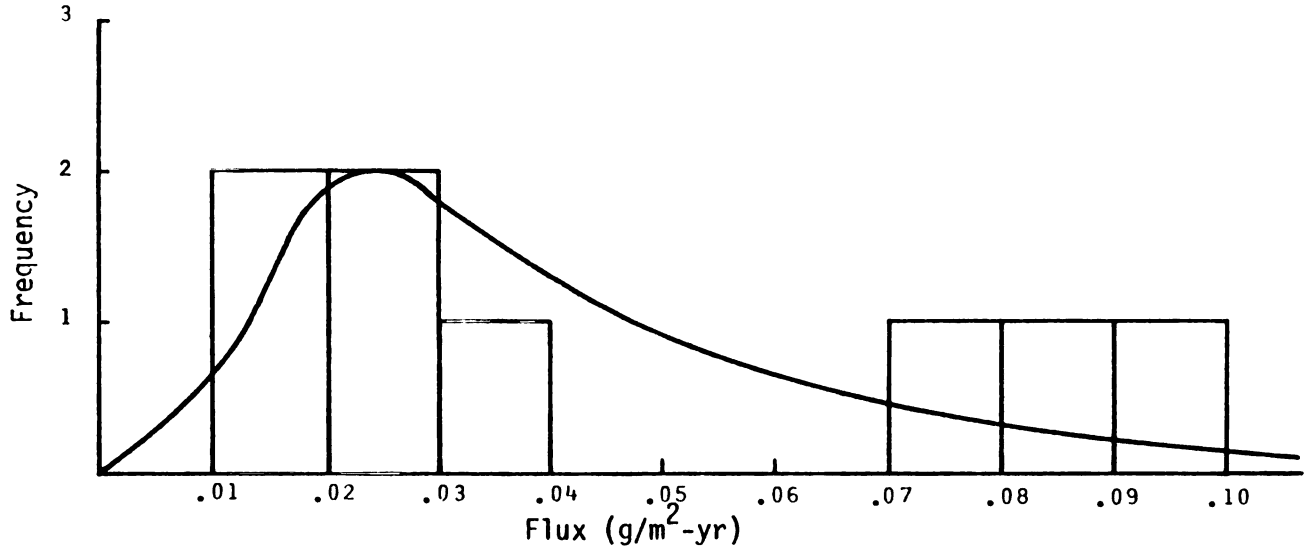

Figure A.5. Total Phosphorus Flux from the Atmosphere to Agricultural Land Uses.

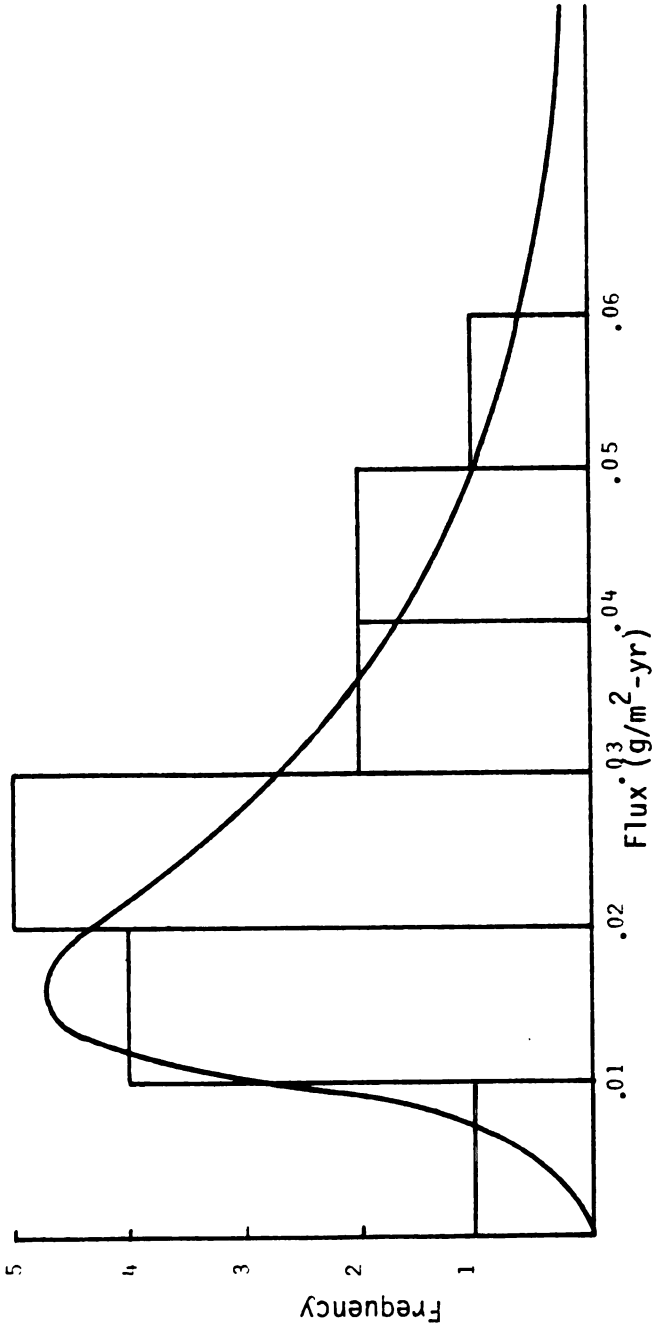

Figure A.6. Total Phosphorus Flux from the Atmosphere to Forested Land Uses.

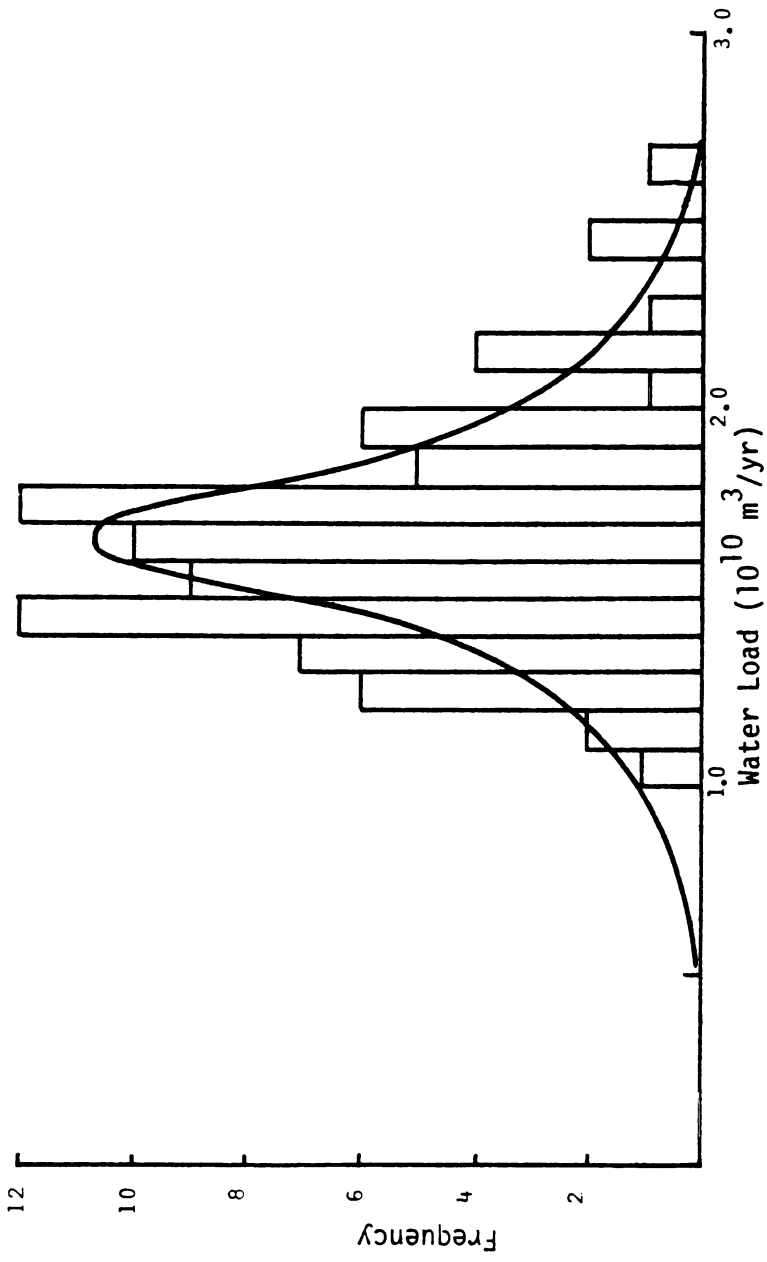

Figure A.7. Total Atmospheric Water Loading to Lake Surface.

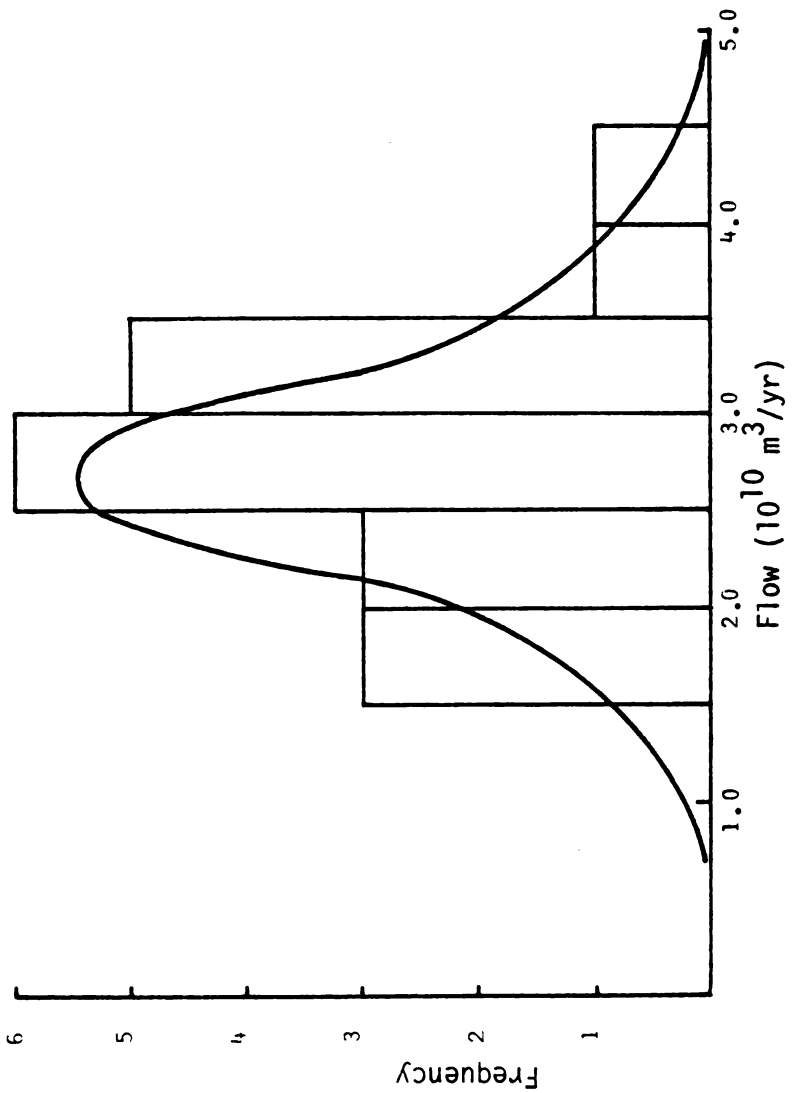

Figure A.8. Water Load from Lake Ontario Basin Tributaries.

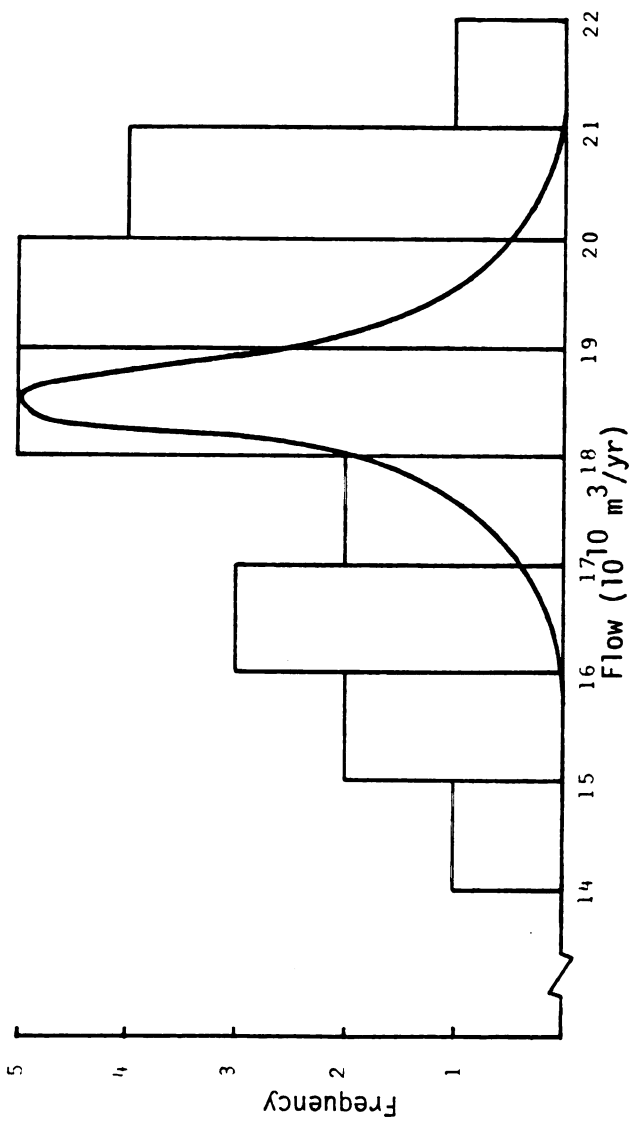

Figure A.9. Annual Flow from Lake Erie at Niagara-Welland.

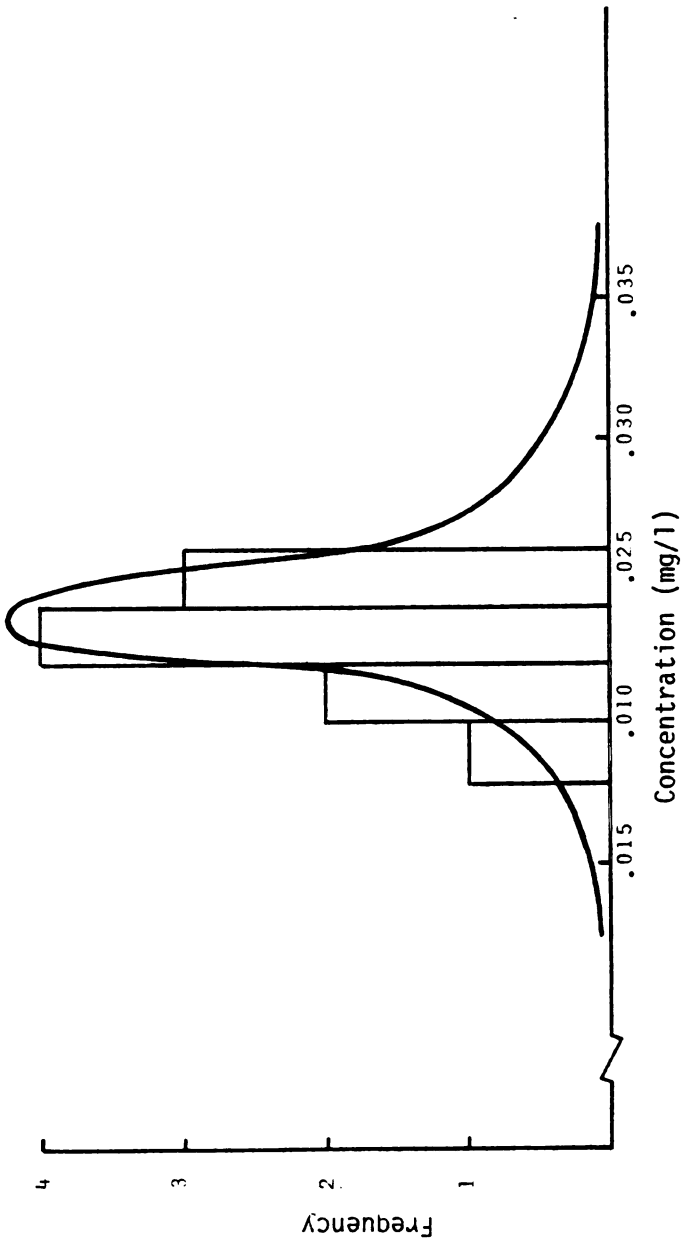

Figure A.10. Total Phosphorus Concentration from Lake Erie at Niagara-Welland.

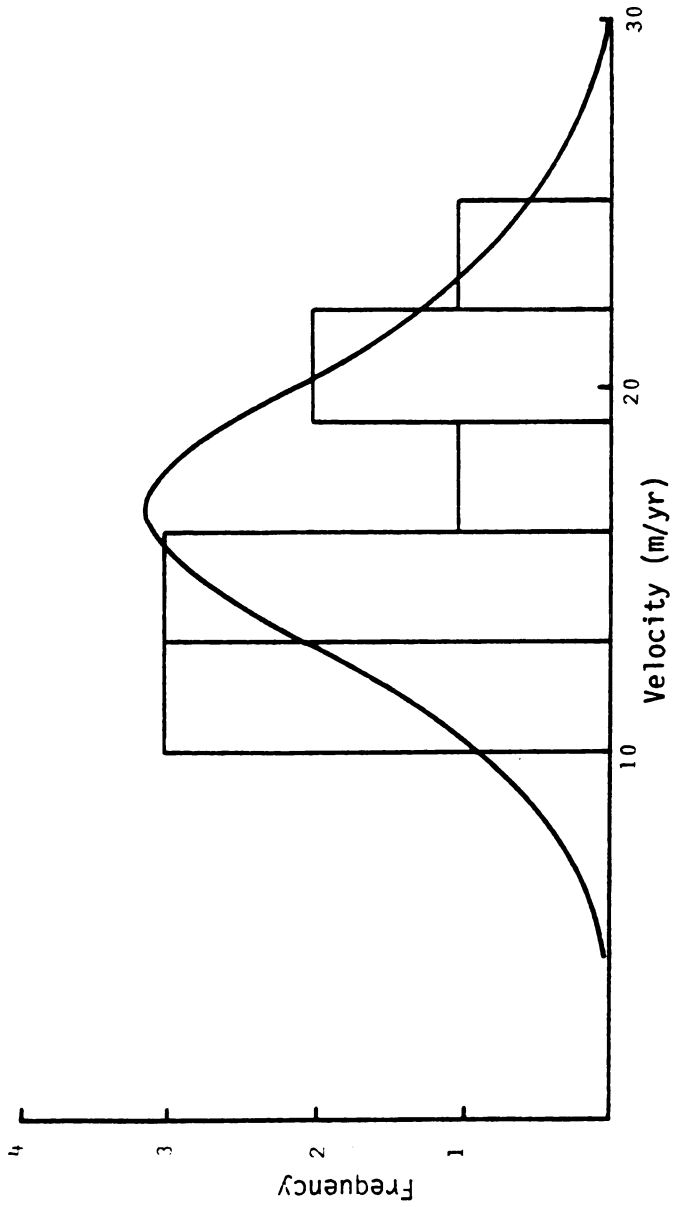

Figure A.11. Apparent Settling Velocity.

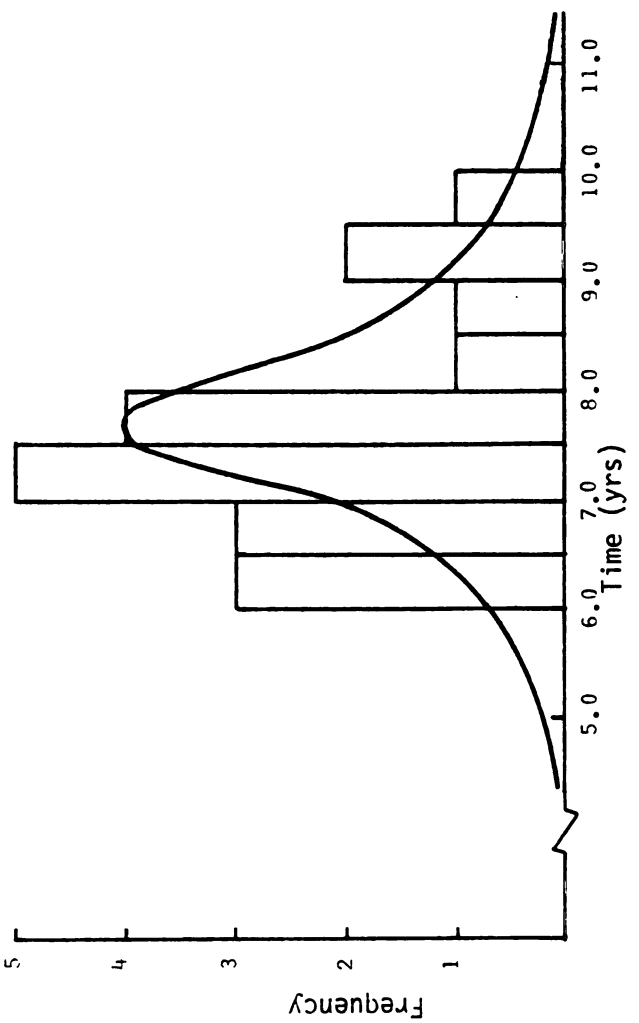

Figure A.12. Hydraulic Detention Time.

APPENDIX B

APPENDIX B

GRAPHICAL OUTPUTS

Graphical interpretations of the simulation results comprise this appendix. Figures B.3, 4, 5a, 6, 7, 9, 13, and 13a have shaded areas which represent the estimated uncertainty (#1 standard deviation) about the predictions. Figures B.11 and 12 display this same uncertainty through the use of histograms of representative single-year prediction distributions.

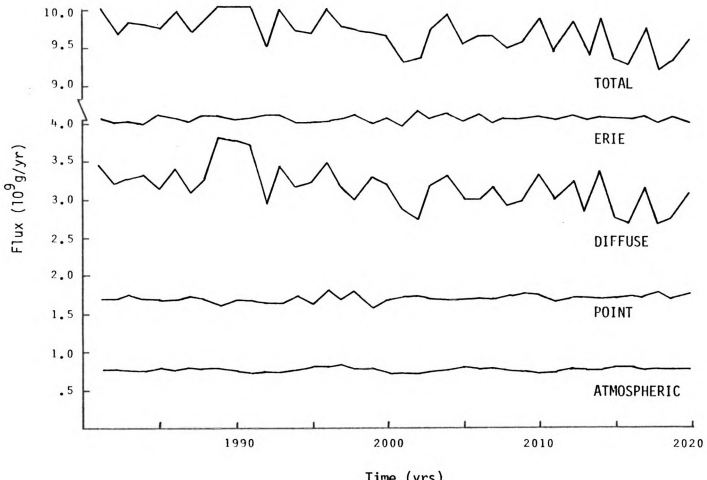

Figure B.1. Fluxes For 40 Years: Diffuse, Point, Atmospheric, Lake Erie, and Total (100 Iterations).

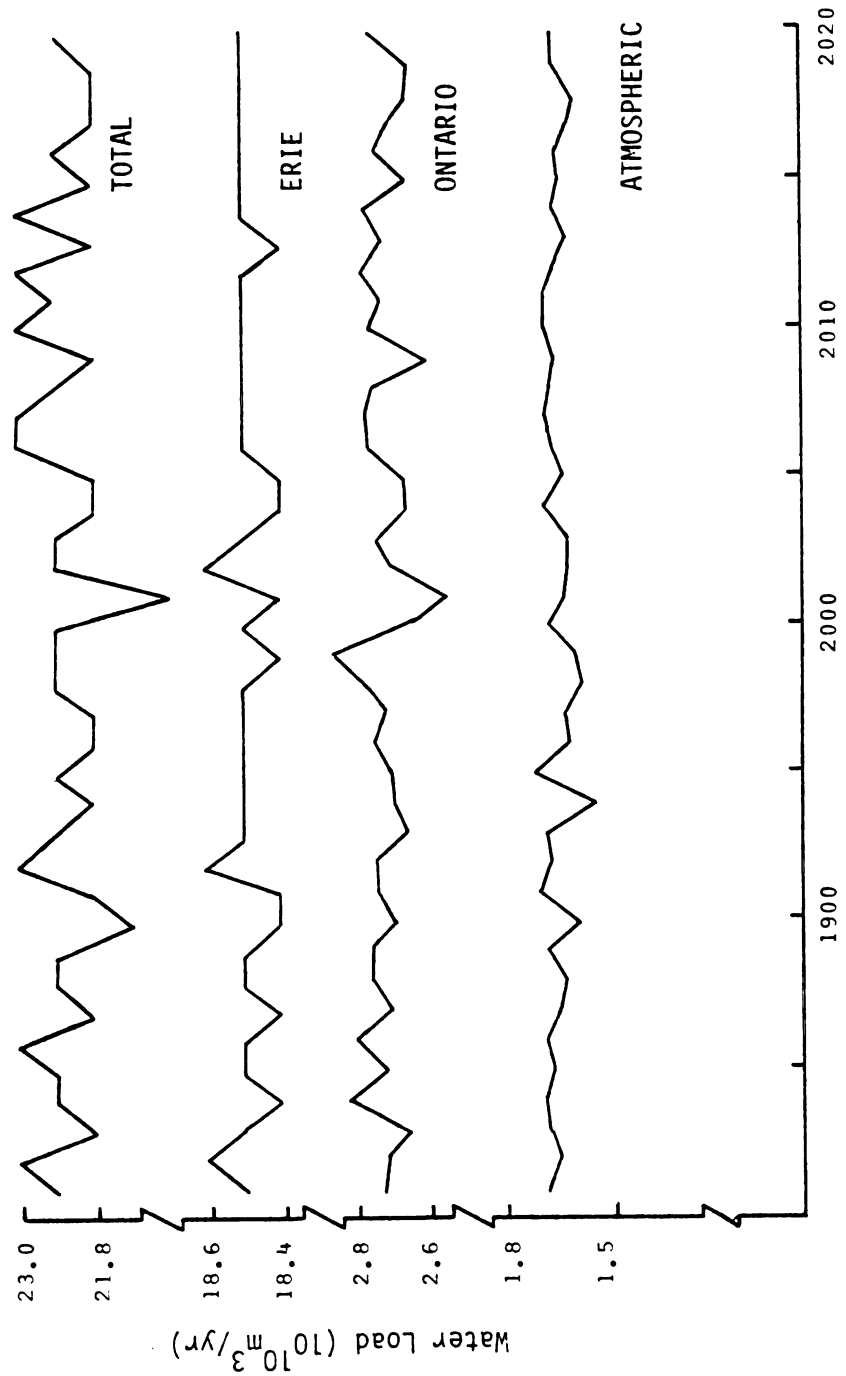

Figure B.2. Water Load for 40 yrs.: Atmospheric, Erie, Ontario, Total.

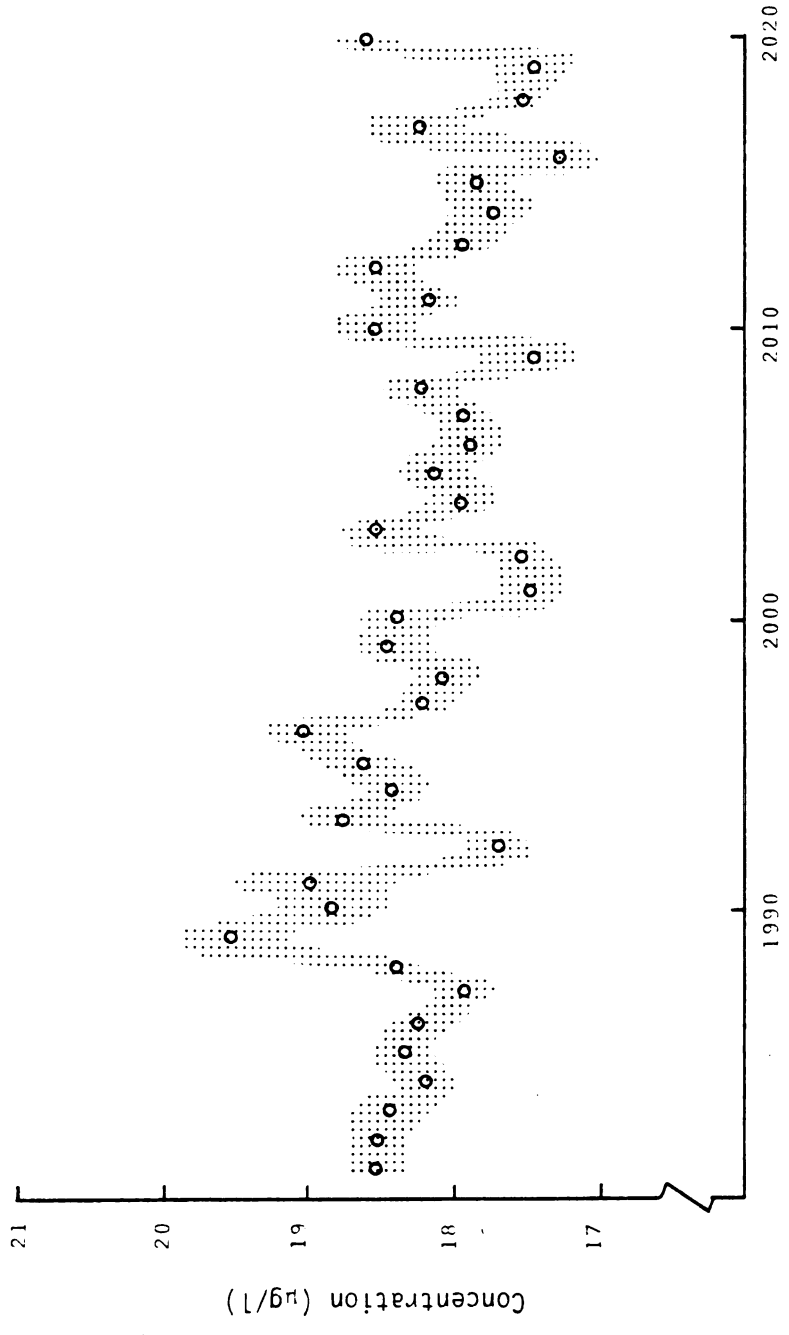

Figure B.3. 40 Years of Phosphorus Concentration Predictions at 100 Iterations Per Year.

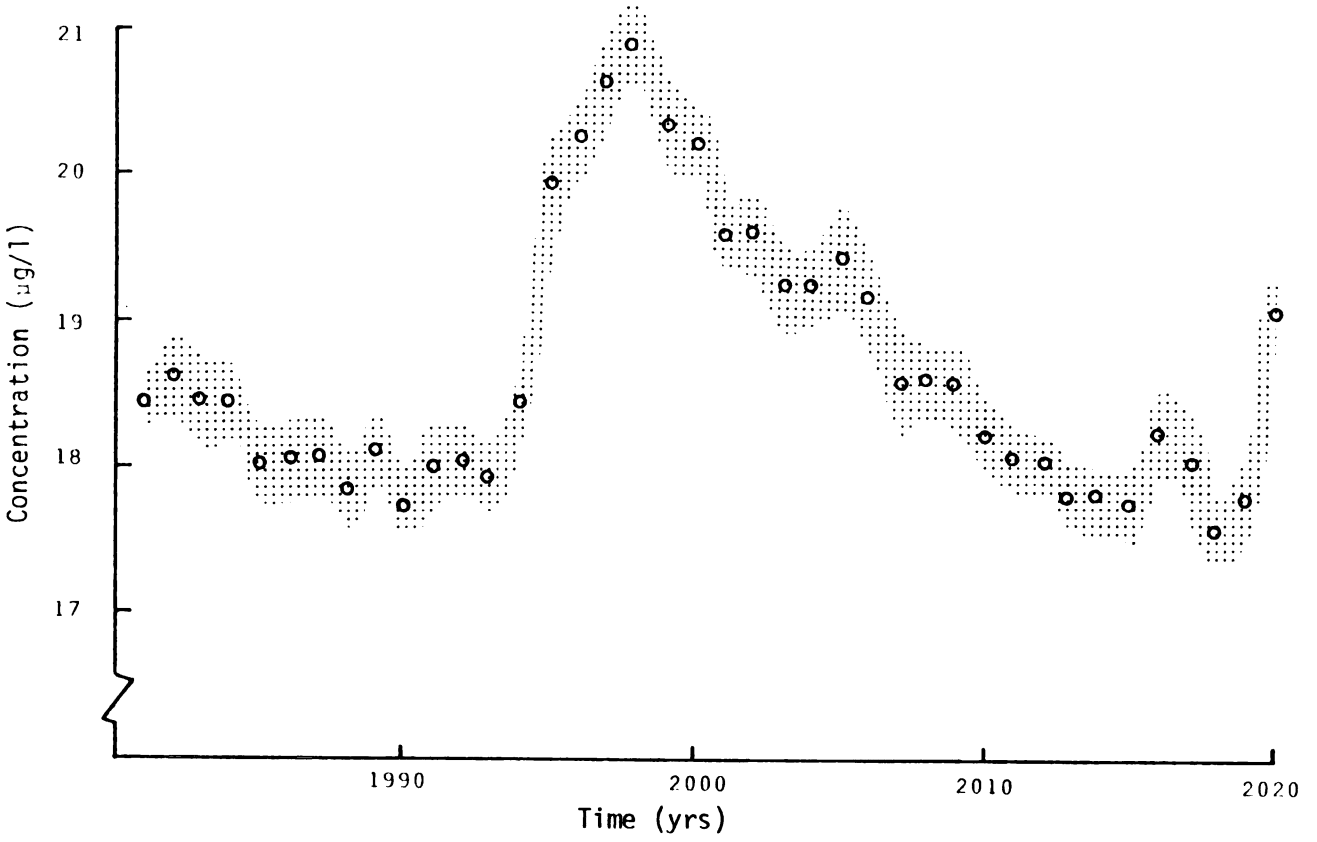

Figure B.4. 40 Years of Phosphorus Concentration Predictions at 500 Iterations Per Year.

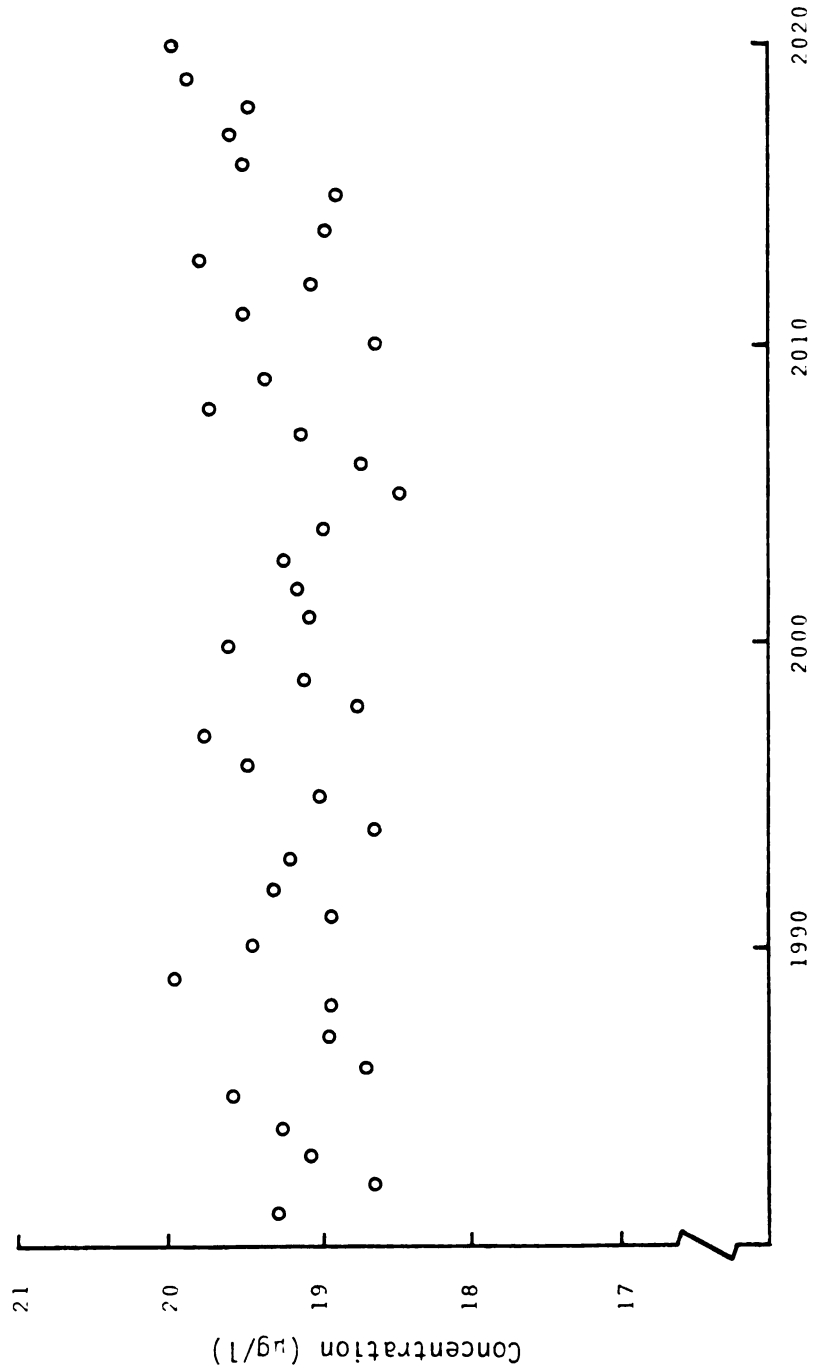

Figure B.4a. Concentration Predictions at 500 Iterations Per Year.

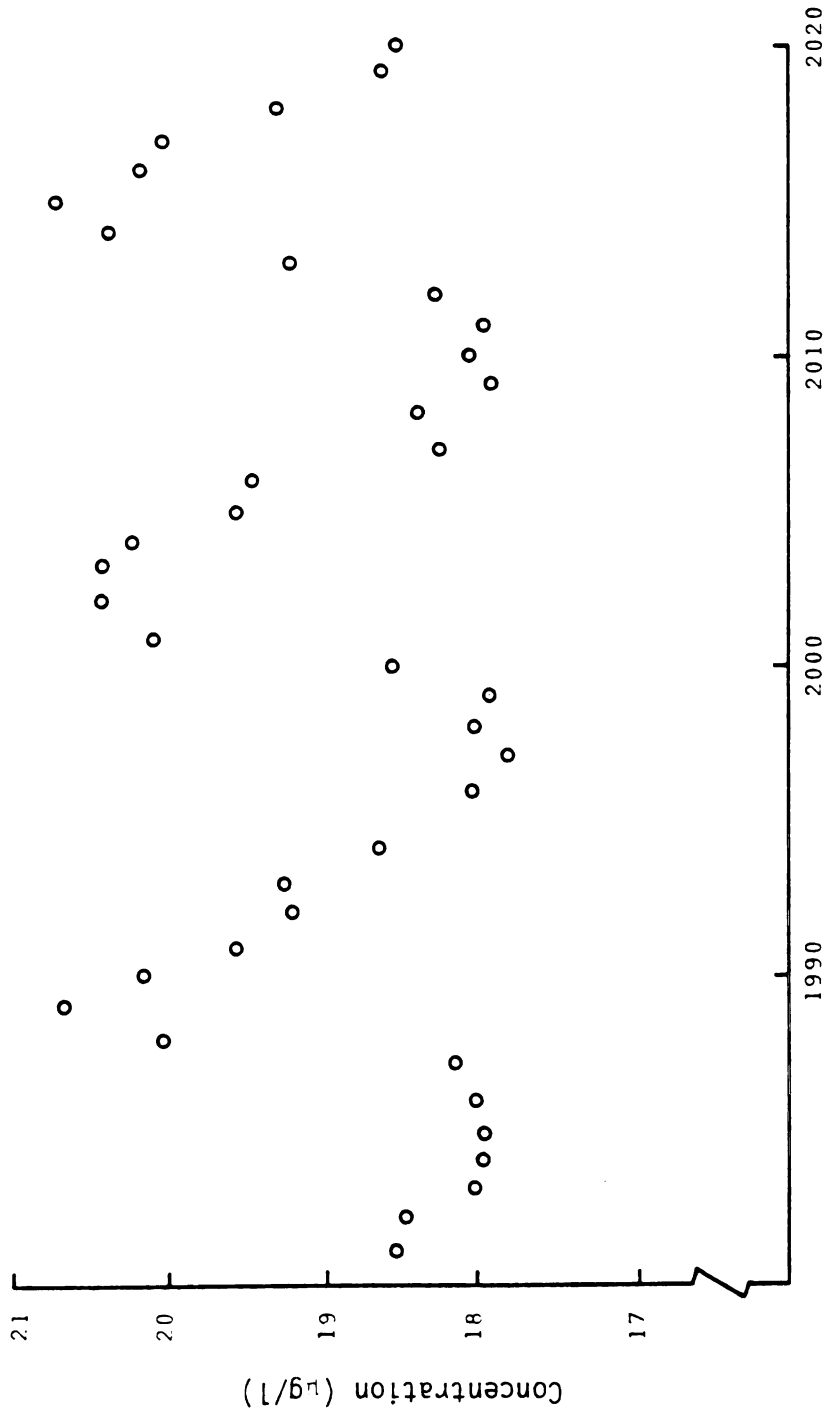

Figure B.5. 40 Years of Phosphorus Concentration Predictions at 1000 Interations Per Year.

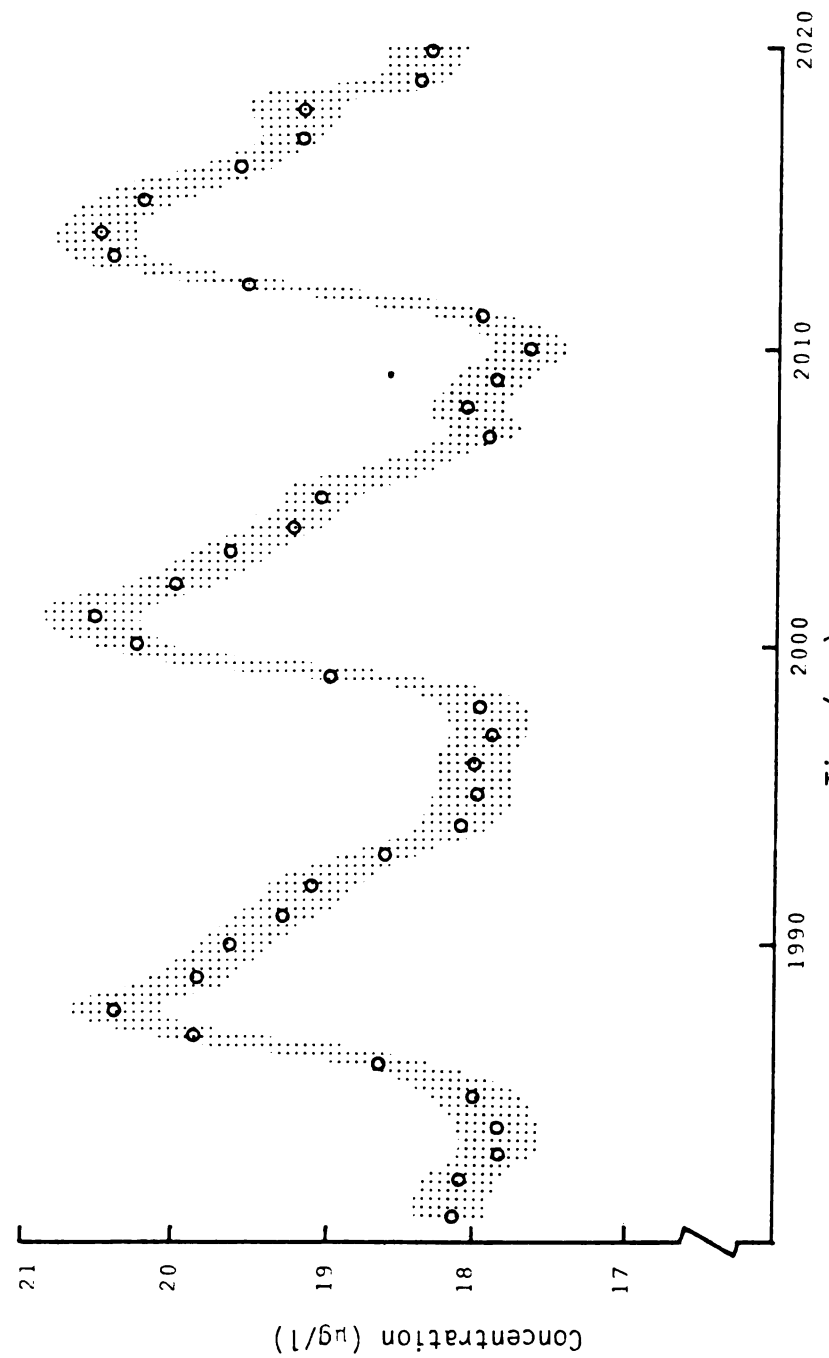

Figure B.5a. Concentration Predictions at 1000 Iterations Per Year.

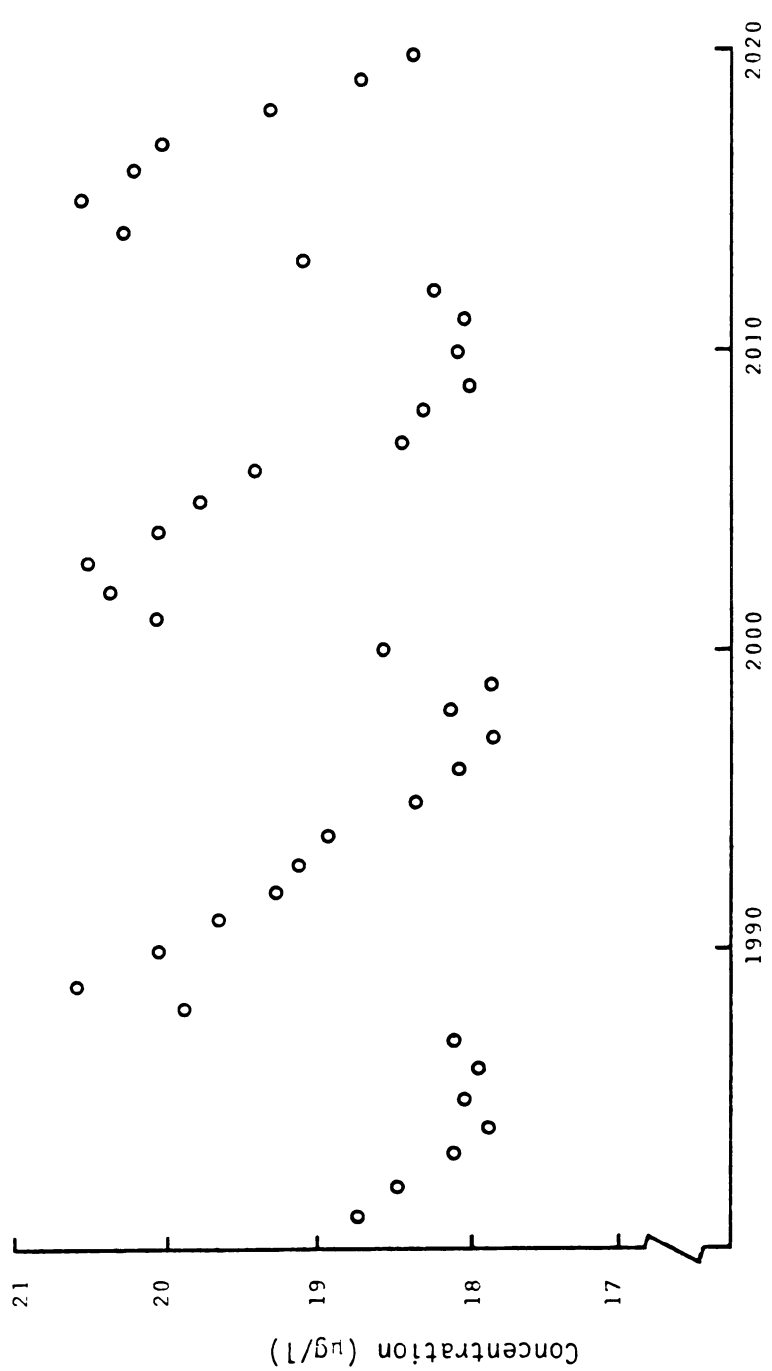

Figure B.5b. Concentration Predictions for MKC2 = 6,610,000,000. at 1000 Iterations Per Year.

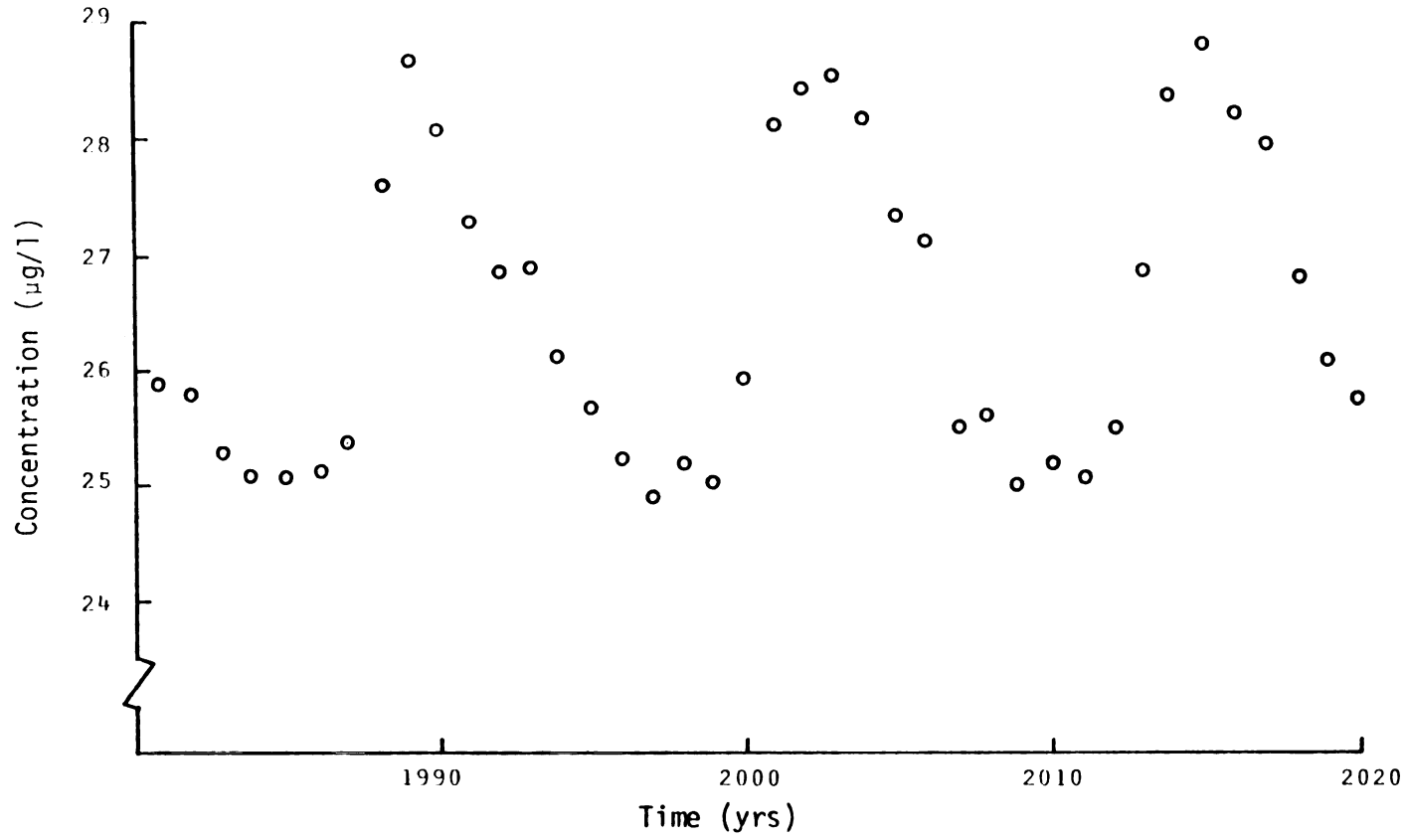

Figure B.5c. Concentration Predictions at 1000 Iterations Per Year Using 1/2 v_s.

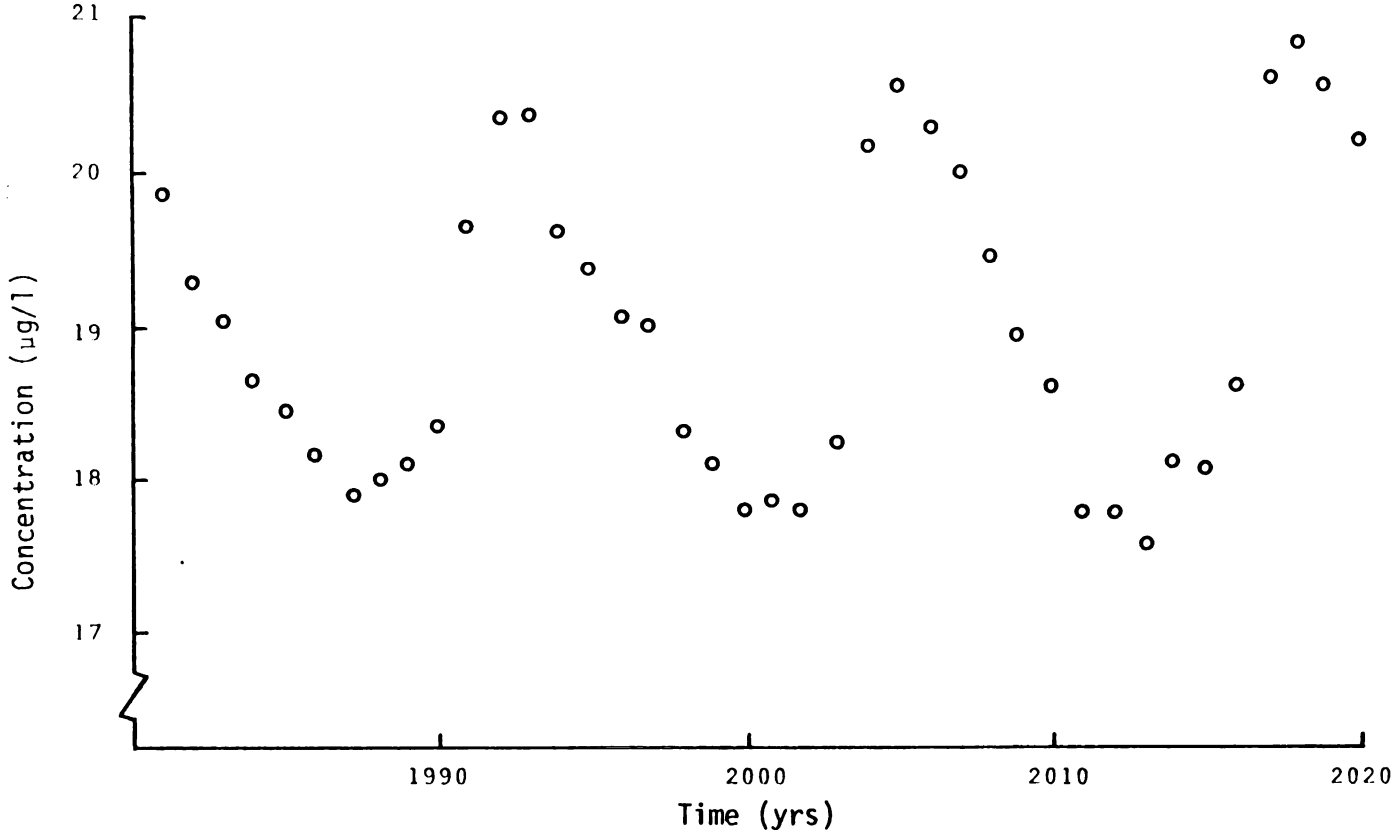

Figure B.5d. Concentration Predictions Using Larger Random Seed For 1000 Iterations Per Year.

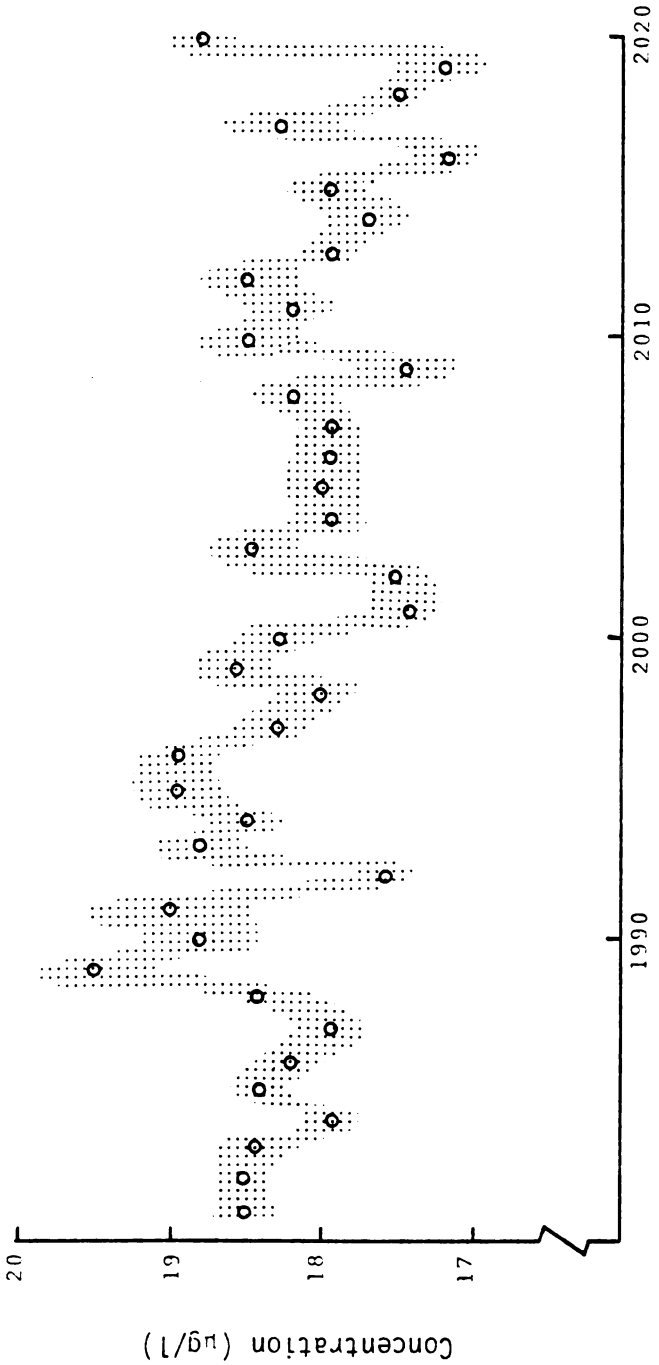

Figure B.6. 40 Years of Phosphorus Concentration Predictions Using 1/2 Variability for Point Source Flux.

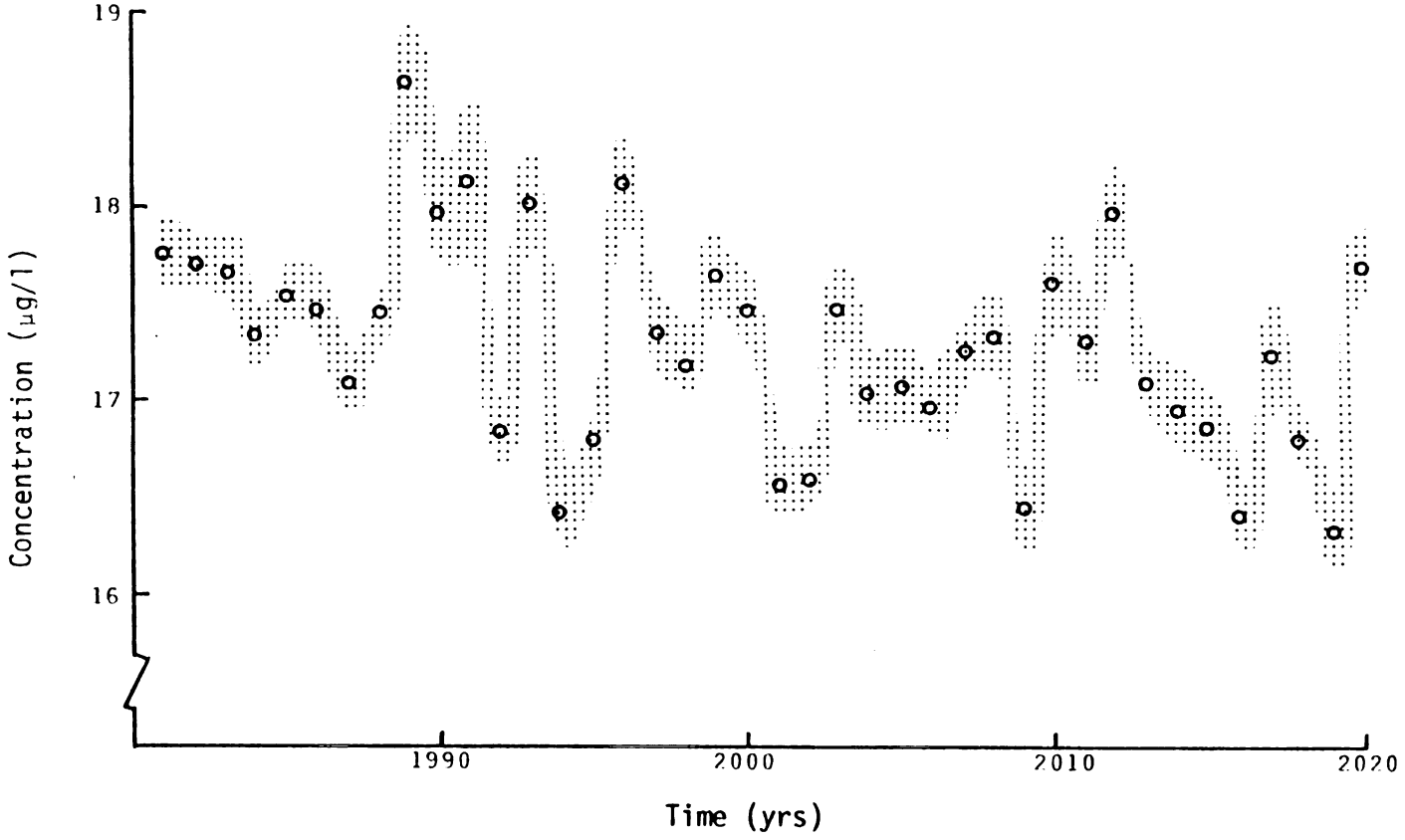

Figure B.7. 40 Years of Phosphorus Concentration Predictions with all Phosphorus Removal for Point Flux.

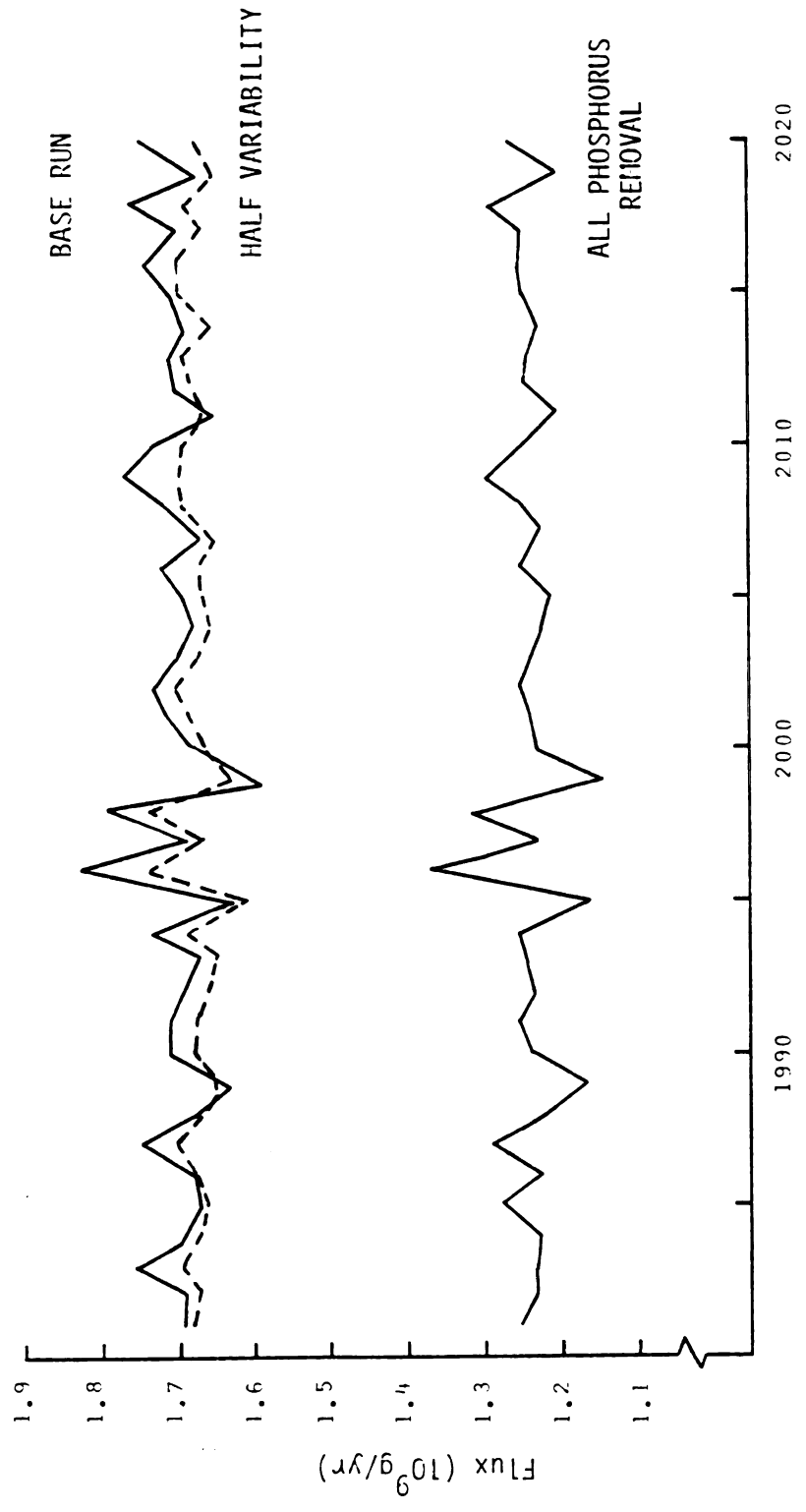

Figure B.8. 40 yrs of Point Source Flux Predictions -- Comparisons.

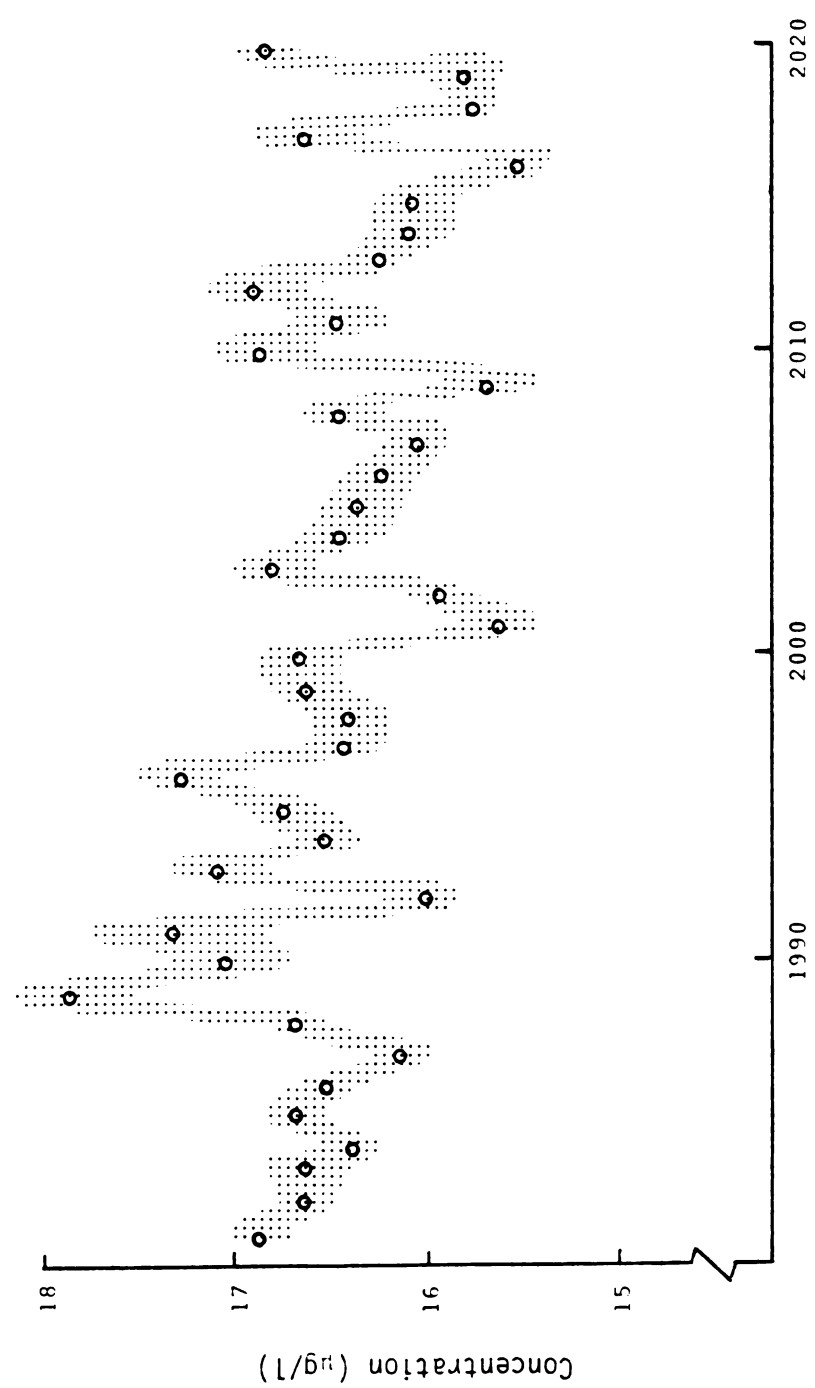

Figure B.9. 40 Years of Phosphorus Concentration Predictions Using Eastern Erie Concentration Data.

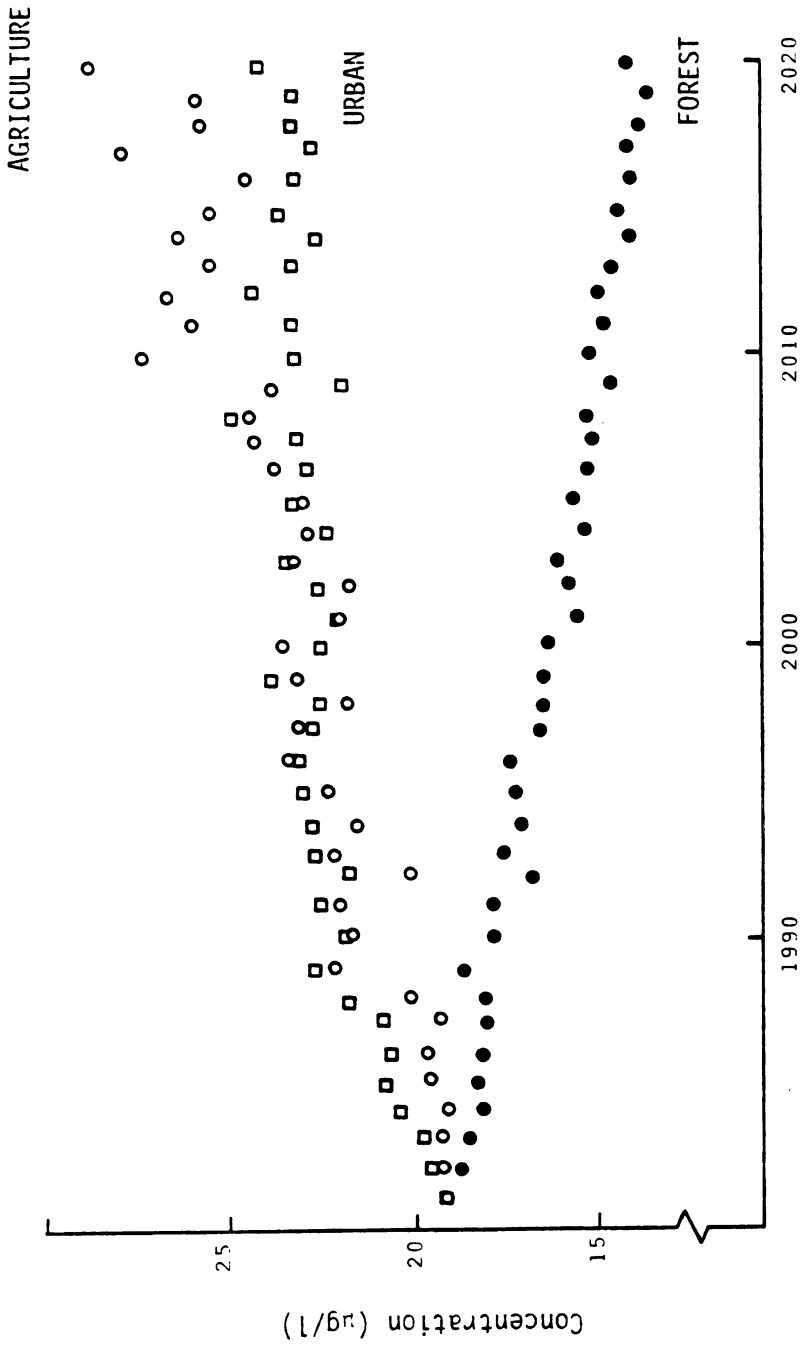

Figure B.10. 40 yrs of Phosphorus Concentration Predictions: Land Use Shifts to All One Type.

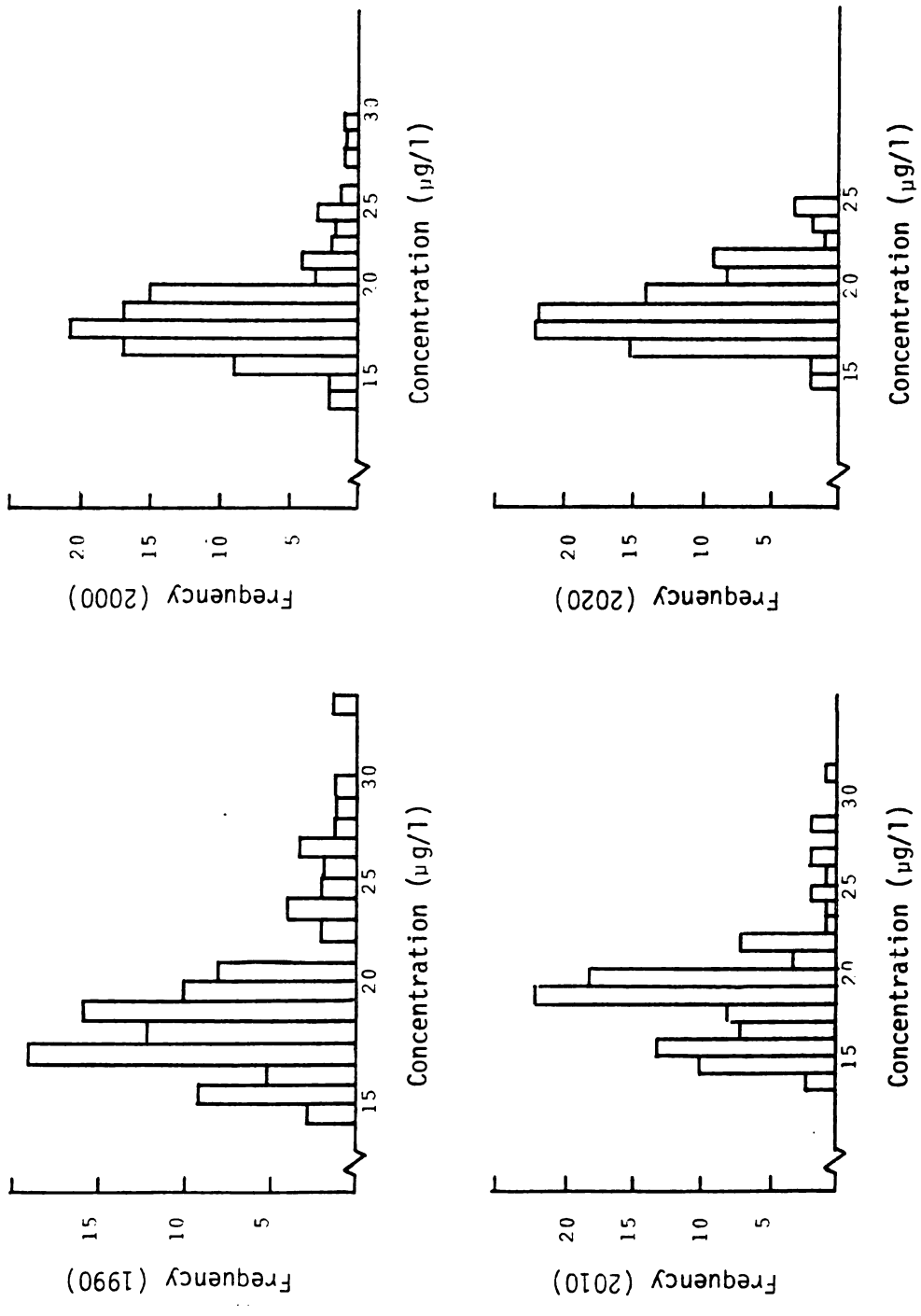

Figure B.11 100 Single-Year Concentration Predictions for Years 1990, 2000, 2010, 2020.

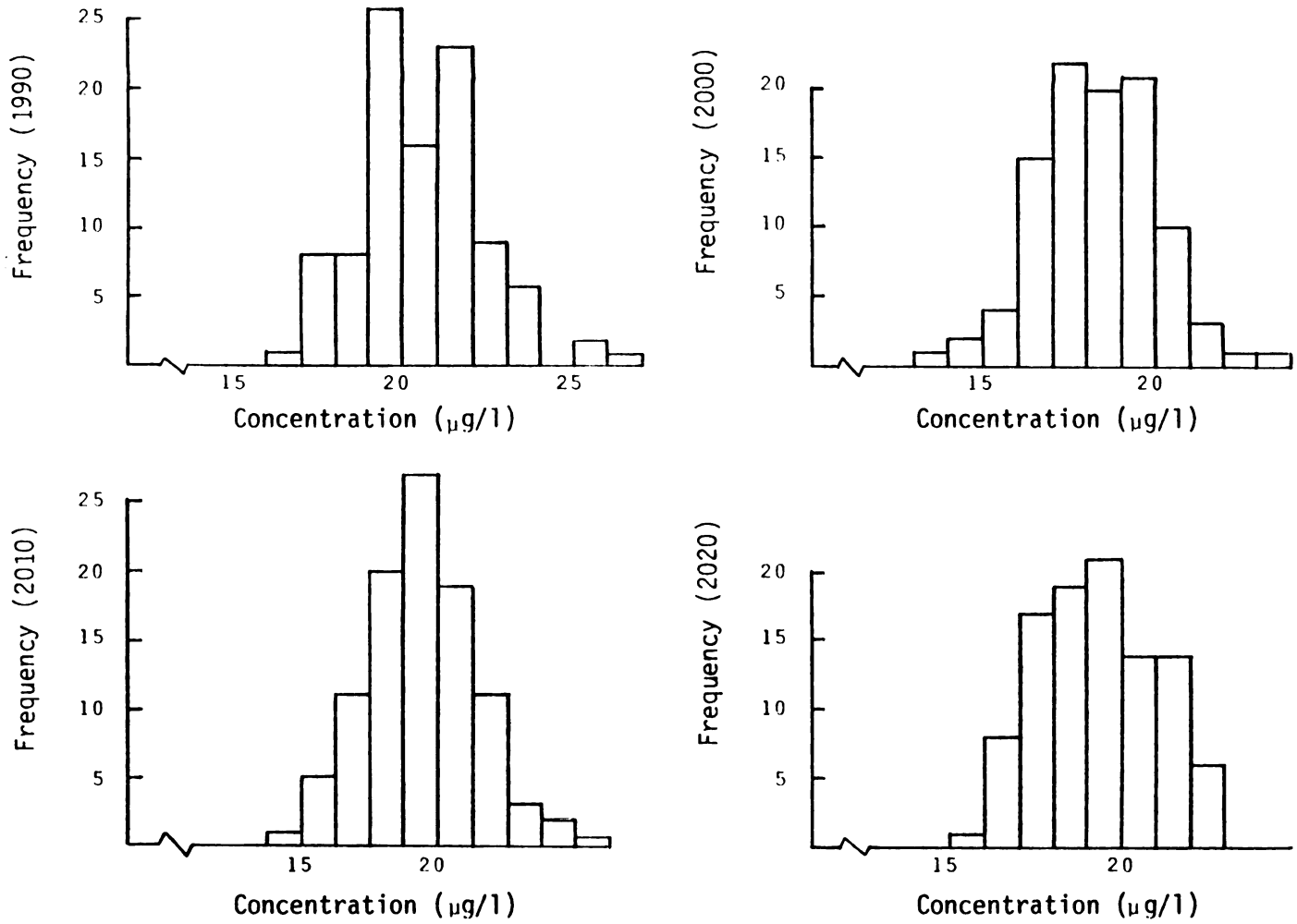

Figure B.12. 100 Concentration Predictions Using All Normal Distributions for Years 1990, 2000, 2010, 2020.

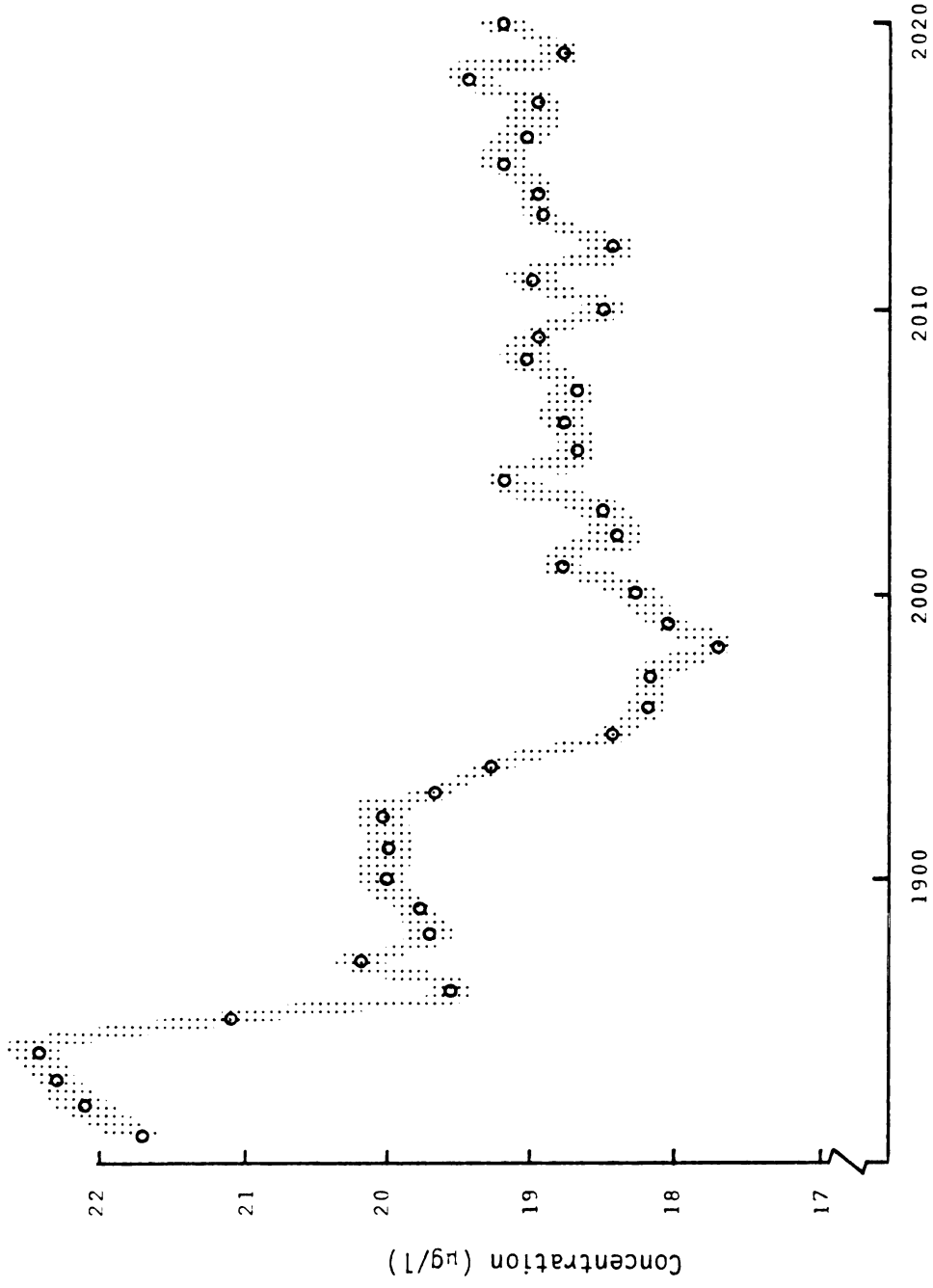

Figure B.13. 40 Years of Phosphorus Concentration Predictions Using All Normal Distributions.

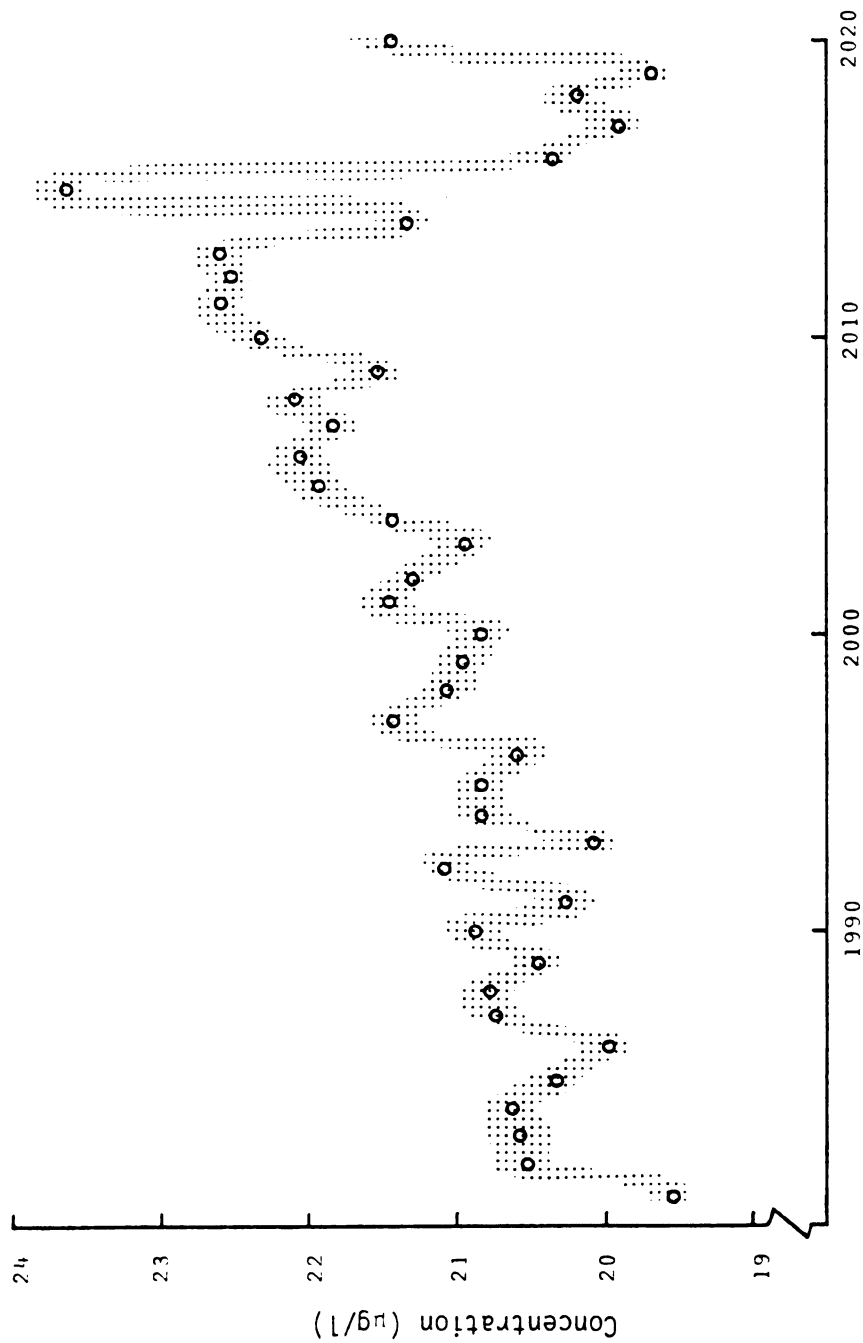

Figure B.13a. Concentration Predictions Using All Normal Distributions.

APPENDIX C

APPENDIX C

TABULAR OUTPUTS

Tabular presentations of summary statistics for simulation results are included in this appendix. Many of these statistics are not very meaningful as far as interpreting annual prediction variability. Statistics such as the 40-year mean of annual means, and variability about that mean are included for the sole purpose of relative comparison. It is helpful to use these values to compare, on a relative basis, different sensitivity or scenario experiments. Tables C.6 and 7, however, directly address the focus of the study by presenting specific annual variability information.

Table C.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		

Table C.2: Statistical Summary of 100 Iteration Base Run - Flux Terms (g/yr)

Diffuse Source Flux (32.5%)	$\bar{x} = 3.161 \times 10^9$	$CV_{\bar{x}} = .0882$	
	$s_{\bar{x}} = .276 \times 10^9$	$R_{\bar{x}} = .9885$	
	y-intercept = 3.44	slope = -0.01	r = -.57
Point Source Flux (17.5%)	$\bar{x} = 1.703 \times 10^9$	$CV_{\bar{x}} = .0252$	
	$s_{\bar{x}} = .043 \times 10^9$	$R_{\bar{x}} = .9990$	
	y-intercept = 1.69	slope = .0006	r = .17
Atmospheric Flux (8.0%)	$\bar{x} = 7.762 \times 10^9$	$CV_{\bar{x}} = .0370$	
	$s_{\bar{x}} = .287 \times 10^8$	$R_{\bar{x}} = .9980$	
	y-intercept = 7.77	slope = -.0003	r = -.012
Flux From Lake Erie (42.0%)	$\bar{x} = 4.064 \times 10^9$	$CV_{\bar{x}} = .0108$	
	$s_{\bar{x}} = .044 \times 10^9$	$R_{\bar{x}} = .9998$	
	y-intercept = 4.07	slope = -.0001	r = -.022
Total Flux (100%)	$\bar{x} = 9.702 \times 10^9$	$CV_{\bar{x}} = .0278$	
	$s_{\bar{x}} = .268 \times 10^9$	$R_{\bar{x}} = .9989$	
	y-intercept = 9.97	slope = -.013	r = -.564

Table C.3: Statistical Summary of 100 Iteration Base Run - Water Load Terms (m^3/yr)

Atmospheric Water Loading (7%)	$\bar{x} = 1.659 \times 10^{10}$	y-intercept = 1.64
	$s_{\bar{x}} = .033 \times 10^{10}$	slope = -.0002
	$CV_{\bar{x}} = .0199$	r = -.085
	$R_{\bar{x}} = .9994$	
Water Load From Lake Erie (81%)	$\bar{x} = 1.849 \times 10^{11}$	y-intercept = 1.83
	$s_{\bar{x}} = .005 \times 10^{11}$	slope = -.0011
	$CV_{\bar{x}} = .0027$	r = -.048
	$R_{\bar{x}} = 1.0000$	
Water Load From Lake Ontario Basin Tributaries (12%)	$\bar{x} = 2.721 \times 10^{10}$	y-intercept = 2.73
	$s_{\bar{x}} = .056 \times 10^{10}$	slope = -.0005
	$CV_{\bar{x}} = .0206$	r = -.1140
	$R_{\bar{x}} = .9994$	
Total Water Load (100%)	$\bar{x} = 2.287 \times 10^{11}$	y-intercept = 2.29
	$s_{\bar{x}} = .009 \times 10^{11}$	slope = -.00003
	$CV_{\bar{x}} = .0039$	r = -.040
	$R_{\bar{x}} = 1.0000$	

Table C.4: Statistical Summary of Sensitivity Experiments - Phosphorus Concentration Prediction (g/m^3)

Number of Iterations		
100	$\bar{x}$ = .0182	$\overline{CV}$ = .144
	$s_{\bar{x}}$ = .0005	$\bar{s}$ = .0027
	$CV_{\bar{x}}$ = .0275	y-intercept = .0186
	$R_{\bar{x}}$ = .9989	slope = -.00002
		r = -.4586
500	$\bar{x}$ = .01872	$\overline{CV}$ = .144
	$s_{\bar{x}}$ = .00090	$\bar{s}$ = .0027
	$CV_{\bar{x}}$ = .0479	y-intercept = .01878
	$R_{\bar{x}}$ = .9966	slope = -.000003
		r = -.0414
1000	$\bar{x}$ = .01899	$\overline{CV}$ = .143
	$s_{\bar{x}}$ = .00095	$\bar{s}$ = .0027
	$CV_{\bar{x}}$ = .0499	y-intercept = .0186
	$R_{\bar{x}}$ = .9963	slope = .00002
		r = .2282

Table C.4: Continued

Distribution Selection

All Normal Distributions

$\bar{x}$	= .0194	$\overline{CV}$ = .084	
$s_{\bar{x}}$	= .0021	$\bar{s}$ = .0016	
$CV_{\bar{x}}$	= .062	y-intercept	= .0207
$R_{\bar{x}}$	= .9940	slope	= -.00006
		r	= -.6367

Table C.4a: Statistical Summary of Sensitivity Experiments - Phosphorus Concentration Predictions (g/m^3)

Random Seed

All Normal Distributions:

RS = 123457.DO

y-int = .0207 r = -.6367

slope = -.00006 $\overline{CV}$ = .0837

RS = 987225.DO

y-int = .0204 r = .4575

slope = .00004 $\overline{CV}$ = .0802

1000 Iterations Per Year:

RS = 123457.DO

y-int = .0186 r = .2282

slope = .00002 $\overline{CV}$ = .143

RS = 135432.DO

y-int = .0186 r = .18712

slope = .00001 $\overline{CV}$ = .1441

RS = 987225.DO

y-int = .0187 r = .1589

slope = .00001 $\overline{CV}$ = .1482

500 Iterations Per Year:

RS = 123457.DO

y-int = .01878 r = -.0414

slope = -.0032 $\overline{CV}$ = .144

RS = 987543.DO

y-int = .01917 r = -.045054

slope = -.0000015 $\overline{CV}$ = .14277

Table C.5: Statistical Summary of Scenario Experiments - Phosphorus Concentration Predictions (g/m^3)

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

Table C.5a: Statistical Summary of Scenario Experiments - Phosphorus Concentration Predictions (g/m^3)

Land Use Shifts

Double Rate of Fraction Shifts	$\bar{x}$ = .01823	$\overline{CV}$ = .1406
	$s_{\bar{x}}$ = .00049	$\bar{s}$ = .0026
	$CV_{\bar{x}}$ = .0272	y-intercept = .01865
	$R_{\bar{x}}$ = .9989	slope = -.00002
		r = -.4895
Triple Rate of Fraction Shifts	$\bar{x}$ = .01822	$\overline{CV}$ = .1380
	$s_{\bar{x}}$ = .00050	$\bar{s}$ = .0025
	$CV_{\bar{x}}$ = .0273	y-intercept = .01868
	$R_{\bar{x}}$ = .9989	slope = -.00002
		r = -.5197
Constant Rate Shift To All Urban	$\bar{x}$ = .02199	$\overline{CV}$ = .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	$\overline{CV}$ = .2244
	$s_{\bar{x}}$ = .00257	$\bar{s}$ = .00519
	$CV_{\bar{x}}$ = .1139	y-intercept = .01833
	$R_{\bar{x}}$ = .9809	slope = .00021
		r = .9399

Table C.5a: Continued

Constant Rate Shift To All Forest	$\bar{x}$ = .01577	$\overline{CV}$ = .1039
	$s_{\bar{x}}$ = .00163	$\bar{s}$ = .00162
	$CV_{\bar{x}}$ = .1034	y-intercept = .01858
	$R_{\bar{x}}$ = .9842	slope = -.00014
		r = -.98295

Table C.6: Statistical Summary of Single Year Prediction Distributions - Phosphorus Concentration (mg/l)

100 Iterations Per Year

1990	$\bar{\bar{x}} = .01884$	$s_{\bar{x}} = .00376$	$CV_{\bar{x}} = .1998$	$R_{\bar{x}} = .9430$
2000	$\bar{\bar{x}} = .01839$	$s_{\bar{x}} = .00276$	$CV_{\bar{x}} = .1500$	$R_{\bar{x}} = .9672$
2010	$\bar{\bar{x}} = .01858$	$s_{\bar{x}} = .00329$	$CV_{\bar{x}} = .1771$	$R_{\bar{x}} = .9547$
2020	$\bar{\bar{x}} = .01867$	$s_{\bar{x}} = .00204$	$CV_{\bar{x}} = .1093$	$R_{\bar{x}} = .9823$

100 Iterations Per Year Using All Normal Distributions

1990	$\bar{\bar{x}} = .02051$	$s_{\bar{x}} = .00183$	$CV_{\bar{x}} = .0894$	$R_{\bar{x}} = .9881$
2000	$\bar{\bar{x}} = .01831$	$s_{\bar{x}} = .00170$	$CV_{\bar{x}} = .0927$	$R_{\bar{x}} = .9873$
2010	$\bar{\bar{x}} = .01849$	$s_{\bar{x}} = .00162$	$CV_{\bar{x}} = .0877$	$R_{\bar{x}} = .9886$
2020	$\bar{\bar{x}} = .01923$	$s_{\bar{x}} = .00170$	$CV_{\bar{x}} = .0883$	$R_{\bar{x}} = .9884$

Table C.7: Statistical Summary of Prediction Variability Over Time - Phosphorus Concentration (mg/l)

100 Iterations Per Year	$\bar{s}_{\bar{x}}$	= .0026	y-intercept$_{\bar{s}}$	= .0026
	s range	= .0017-.0054	slope	= -.000002
	$\overline{CV}_{\bar{x}}$	= .143	r	= -.0295
	CV range	= .096-.284		
500 Iterations Per Year	$\bar{s}_{\bar{x}}$	= .0026	y-intercept$_{\bar{s}}$	= .0028
	s range	= .0021-.0034	slope	= -.000013
	$\overline{CV}_{\bar{x}}$	= .144	r	= -.2692
	CV range	= .110-.180		
1000 Iterations Per Year	$\bar{s}_{\bar{x}}$	= .0027	y-intercept$_{\bar{s}}$	= .0028
	s range	= .0022-.0032	slope	= -.000004
	$\overline{CV}_{\bar{x}}$	= .143	r	= -.1784
	CV range	= .118-.167		
100 Iterations Per Year, All Normal Distributions	$\bar{s}_{\bar{x}}$	= .0016	y-intercept$_{\bar{s}}$	= .0017
	s range	= .0013-.0023	slope	= -.000005
	$\overline{CV}_{\bar{x}}$	= .084	r	= -.2593
	CV range	= .068-.110		
100 Iterations Per Year, Shift To All Agriculture	$\bar{s}_{\bar{x}}$	= .0052	y-intercept$_{\bar{s}}$	= .0025
	s range	= .0019-.0103	slope	= .00013
	$\overline{CV}_{\bar{x}}$	= .224	r	= .7505
	CV range	= .100-.389		

APPENDIX D

APPENDIX D

1000 ITERATION-PER-YEAR PERIODICITY

The persistence of the remarkable periodicity which appears in the 1000 iteration-per-year runs is quite perplexing. On a random basis, the appearance of peaks is not surprising. However, the occurance of three distinct peaks 12-1/2 to 13 years apart, regardless of the changes to the simulation, is more than just a little surprising.

After thoroughly checking the computer code for possible errors and examining the random number generators for possible cycling, the conclusion remained that this is not representing a random process. Changing the random seed does not affect this persistent irregularity (see Figures B.5a and B.5d). A similar peak appears in the 500 iteration-per-year run, but does not persist when the random seed is changed (Figures B.4 and B.4a). The next solution involved altering the second Markov constant (MKC2) in the lag-one flow model used to generate random lag-one correlated flow values, representing the flow from Lake Erie. Since flux from Lake Erie is the single greatest influence on phosphorus concentration in Lake Ontario, changing the nature of flow value generation should change the nature of the predictions. By increasing MKC2 by 5 percent, the random portion of the flow value is increased. This change indirectly affects the

overall variability of the entire system, and might be expected to result in peaks of greater amplitude or greater frequency. Neither expectation is realized, however. The peaks persist with the same familiar amplitude and frequency (Figure B.5b). The last attempt at solving this mystery resorts to questioning the nature of the model expression itself. The model is explicitly lag-one in character. Implicitly, however, the model lags back, to a diminishing degree, many years. That is to say, if the simulation incorporates 75 percent of the previous year's predicted concentration into the current prediction, next year's prediction will not only contain 75 percent of this year's prediction, but implicitly will contain 75 percent of 75 percent of last year's prediction. In this manner, even the tenth year's prediction will have incorporated in it about 8 percent of the first year's prediction. The term that dictates how great a proportion of the previous year's prediction will be incorporated into the current year's prediction is the log term, $e^{-(1/\tau + v_s/z)}$. Since detention time, τ, is not highly variable (CV = .14) and lake mean depth, z, is represented as a constant, the settling velocity term, v_s, was reduced in magnitude by one-half. If the periodicity is inherent in the model expression, the period of this cyclic phenomenon should change. Once again, however, no change from the established pattern occurred (Figure B.5c).

This anomaly serves to point out how very sensitive complex simulations can be to the most unexpected elements. Unfortunately, the mystery is only manifested by the 1000 iteration run. Storage

requirements demand 240,000 binary bits of central memory for each of these runs, making it a very expensive mystery to solve. Future efforts should focus on the possible cyclic nature of the synergistic effect of two or more randomly generated variates. In specific, there is one point which begins each peak. Each of these initiating points leaves a wide gap between it and the previous point, seemingly disregarding the lag effect that should eliminate, or at least make highly improbable, the occurrence of such points.

There is, of course, a finite probability that any given random number generator will generate a cluster of high values. This probability, however, is certainly not great enough to cause a one-year increase of such magnitude. Several random number generators would have to generate clusters of high values coincidentally, thereby driving the overall annual prediction artificially high. In addition, a low log term value would contribute to this gap by decreasing the percent of the previous year's prediction that is incorporated into the current year's prediction, and in this way indirectly giving greater weight to these new artificially high values. This hypothesis can only be viewed as a desperate attempt to explain a perplexing problem. Even this, however, explains only how these peaks might occur. The question of why they occur with such regularity remains unanswered.

APPENDIX E

APPENDIX E

SIMULATION DETAILS

A list of parameters and two flow charts accompany the computer listing of the FORTRAN code for the simulation. Cost reduction measures should be incorporated for longer runs (more years or more iterations per year) or for extended application. The major cost-saving device which might be considered is the elimination of the 28 dimensioned variables. At only 100 iterations per year, this necessitates the storage of 2800 values each year. The actual values are of little use other than for use in de-bugging the program. To simplify, and eliminate the need for this storage, a simple summation statement would require only one storage location for each parameter while retaining the ability to describe each year with summary statistics. In addition, some savings may be realized by employing on-line optimization routines and selecting less costly queues and rate groups for the lengthier runs.

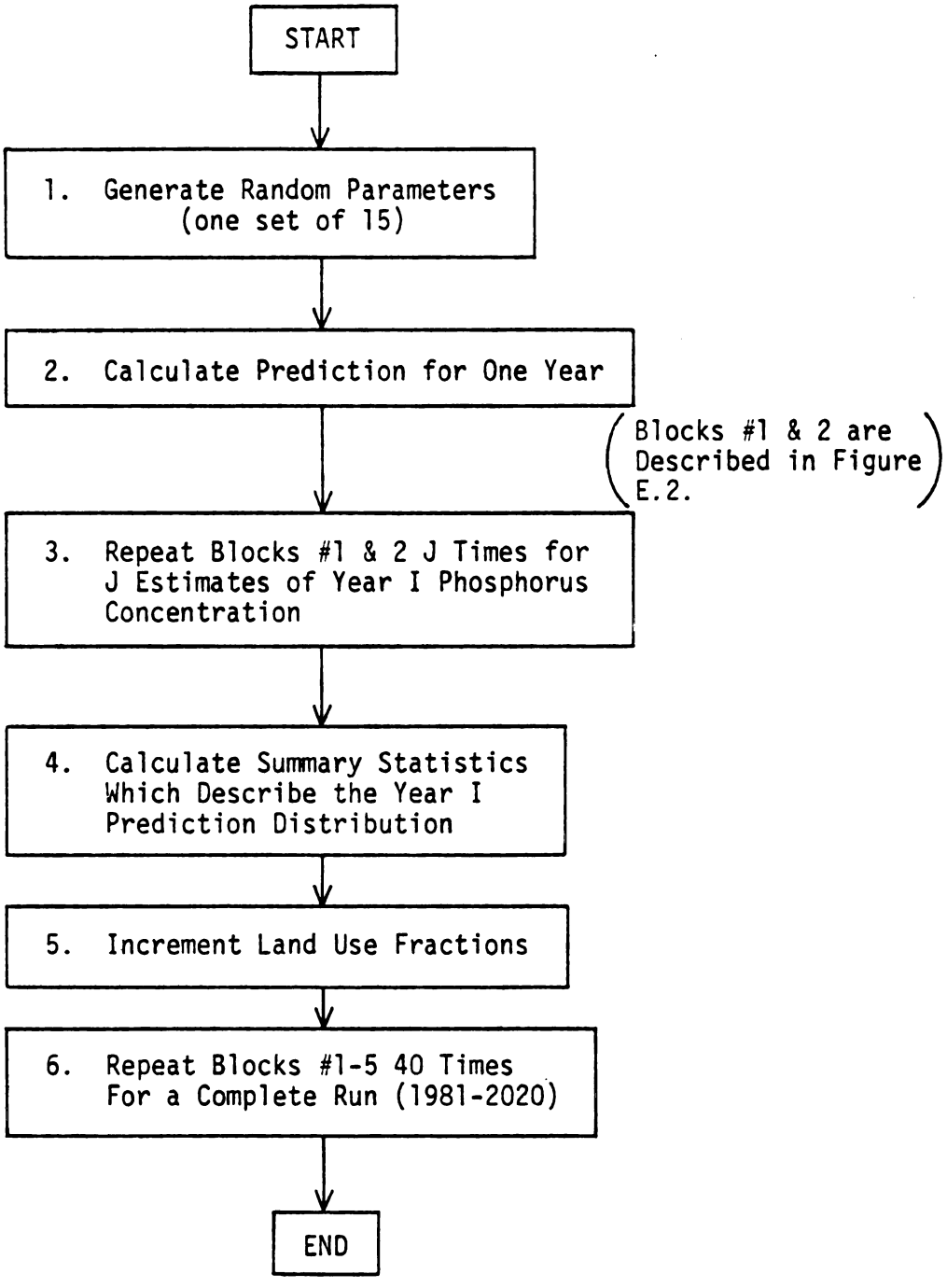

Figure E.1. Flow Diagram of Simulation.

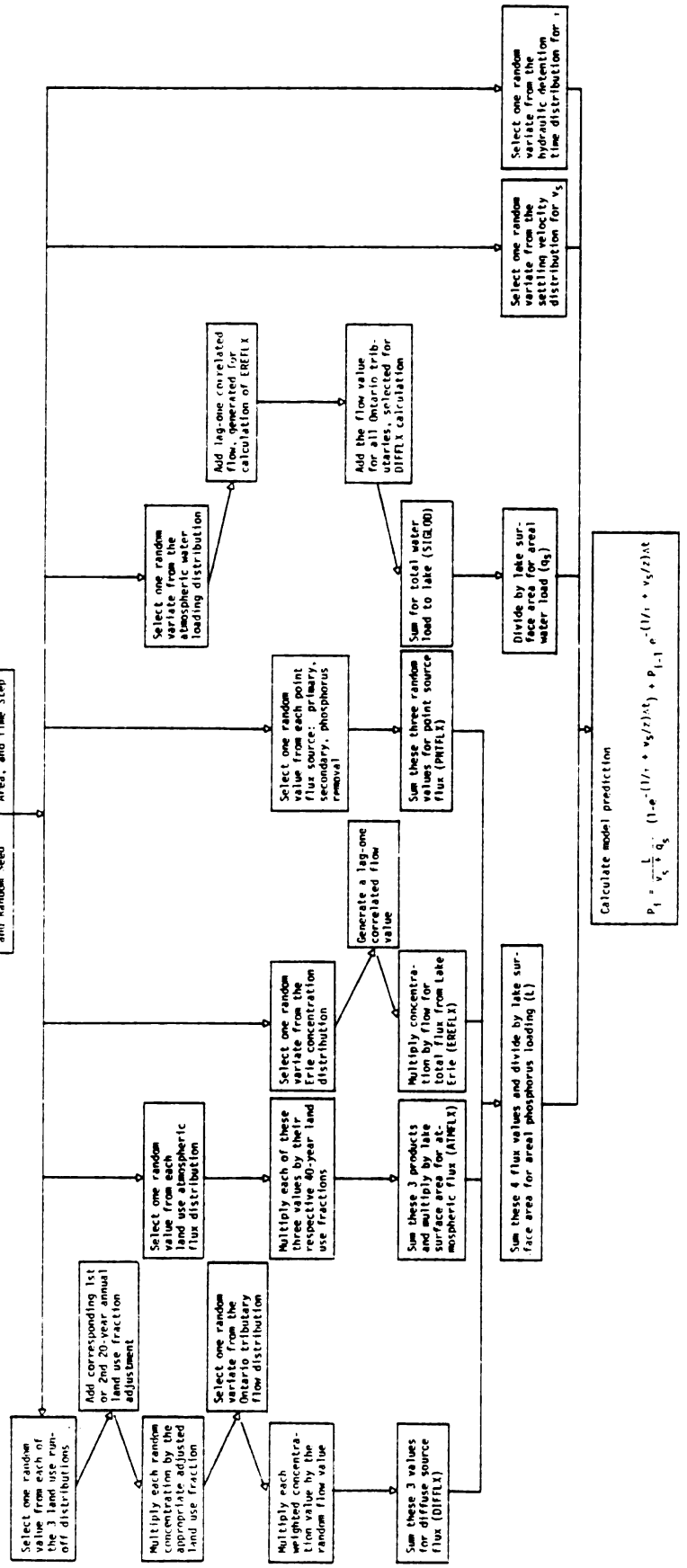

Figure E.2. Diagramatic Representation of Prediction Algorithm.

PARAMETER LIST

SOURCE PROGRAM-MCS

Parameter Name	Units	Notes
XSEED	Dimensionless	Changes Random Seed
Z	Meters	Lake Mean Depth
T	Dimensionless	Counts Iterations
PLAST	Grams/m^3	Previous Year's Concentration Prediction
LUFCFU	Dimensionless	Annual % Change of Urban Land Area
LUFCFA	Dimensionless	Annual % Change of Agricultural Land Area
LUFCFF	Dimensionless	Annual % Change of Forested Land Area
DSEED	Dimensionless	Random Seed
N	Dimensionless	Counts IMSL Iterations
LSA	Square Meters	Lake Surface Area
K	Dimensionless	Year Counter
I	Dimensionless	Iteration Counter
R	Dimensionless	Random Number Output Vector
XM, M	Appropriate	Mean of Parameter
S	Appropriate	Standard Deviation of Parameter
URBCON	Grams/m^3	Runoff Concentration-Urban
AGRCON	Grams/m^3	Runoff Concentration-Agricultural
FORCON	Grams/m^3	Runoff Concentration-Forest
AFXURB	g/m^2-yr	Atmospheric Flux-Urban
AFXFOR	g/m^2-yr	Atmospheric Flux-Forest

SOURCE PROGRAM-MCS - Continued

Parameter Name	Units	Notes
AFXAGR	g/m^2-yr	Atmospheric Flux-Agricultural
ATMFLX	Grams/yr	Atmospheric Flux-Total
ONTFLO	m^3/yr	Water Load-Ontario Tributaries
PFXPRI	g/yr	Point Source Flux-Primary Treatment
PFXSCD	g/yr	Point Source Flux-Secondary Treatment
PFXPRE	g/yr	Point Source Flux-Phosphorus Removal
PNTFLX	g/yr	Point Source Flux-Total
EREFLO	m^3/yr	Water Load-Lake Erie
LSTFLO	m^3/yr	Previous Year's Flow
MKC1	m^3/yr	Markov Constant
MKC2	m^3/yr	Markov Constant
ERECON	g/m^3	Influent Concentration-Lake Erie
VS	m/yr	Settling Velocity
TAU	Years	Hydraulic Detention Time
ATMLOD	m^3/yr	Water Load-Atmospheric
DIFFLX	g/yr	Diffuse Source Flux
EREFLX	g/yr	Flux From Lake Erie
TOTFLX	g/yr	Total Phosphorus Flux
SIGLOD	m^3/yr	Total Water Load
QS	m/yr	Areal Water Load
L	g/m^2-yr	Areal Phosphorus Flux
P	$Grams/m^3$	Areal Phosphorus Flux

SOURCE PROGRAMS-MCS - Continued

Parameter Name	Units	Notes
LOGTRM	Dimensionless	Lag Magnitude Term
TRMONE	g/m^3	Present Year Contribution
TRMTWO	g/m^3	Previous Year Contribution
PDYN	g/m^3	Time Dependent Concentration Prediction
PLAST	g/m^3	Previous Year Concentration Prediction
STDEV	g/m^3	Standard Deviation of 40 Annual Prediction Averages

Subroutine - SORTIT

N	Dimensionless	Array Length
I	Dimensionless	Array Counter
J	Dimensionless	Array Counter

Subroutine - GROUPIT

J	Dimensionless	Do-Loop Counter
JCLS	Dimensionless	Class Size Array

Subroutine - PPSELCT

I	Dimensionless	Do-Loop Counter
X	g/m^3	Prediction Total

Subroutine - SUMSTAT

I	Dimensionless	Do-Loop Counter
T	Appropriate	Prediction Total
XMEAN	Appropriate	Mean
Y	Appropriate	Squared Deviations
X	Appropriate	Sum of Squared Deviations
VAR	Appropriate	Variance
CV	Appropriate	Coefficient of Variation
R	Appropriate	Mode to Mean Ratio

```
C LAKE MODEL ERROR ANALYSIS
C PROGRAM MCS                          DECEMBER 1980
C BY V. DAVID LEE
C DEPARTMENT OF RESOURCE DEVELOPMENT
C MICHIGAN STATE UNIVERSITY
C MCS IS THE SOURCE PROGRAM WHICH SIMULATES 40 YEARS OF LAKE ONTARIO
C BASIN BEHAVIOR. LAND USE - WATER QUALITY INTERACTIONS ARE MODELED
C USING MONTE CARLO TECHNIQUES.
C THIS SOURCE PROGRAM MAY OPT TO CALL ANY OR ALL OF THE FOLLOWING
C SUBROUTINES  =  SORTIT, GROUPIT, PPSELCT, SUMSTAT
C
      PROGRAM MCS(INPUT,OUTPUT)
      REAL LAGONE,MKC1,MKC2,LSTFLO,M,LSA,L,LOGTRM,LUFCFU,LUFCFA,LUFCFF
      COMMON PDYN, K, JCLS, PLAST, STDEV
      DIMENSION R(2),URBCON(100),AGRCON(100),FORCON(100),ERECON(100),
     + ONTFLO(100),EREFLO(100),ATMLOD(100),SIGLOD(100),QS(100),
     + DIFFLX(100),PNTFLX(100),ATMFLX(100),EREFLX(100),TOTFLX(100),
     + LOGTRM(100),TRMONE(100),TRMTWO(100),PDYN(100),VS(100),L(100),
     + TAU(100),P(100),AFXAGR(100),AFXURB(100),AFXFOR(100),PFXPRI(100),
     + PFXSCD(100),PFXPRE(100)
      DOUBLE PRECISION DSEED
      XSEED=0.
      Z= 89.
      T=1.
      DSEED = 123457.00+XSEED
      PLAST = 0.02
      LUFCFU = 0.
      LUFCFA = 0.
      LUFCFF = 0.
      DO 600 K = 1,40
      DO 500 I = 1,100
      N=1
      LSA=19000000000.
C
```

```
C
C  GENERATE A LOG NORMAL RANDOM VARIATE FOR EACH OF THE SIX FOLLOWING
C  MODEL PARAMETERS.
C  NOTE =  THIS IMSL RANDOM NUMBER GENERATOR REQUIRES THE USE OF MEANS
C  AND STANDARD DEVIATIONS FROM LOG-TRANSFORMED DATA.
C
C  SAMPLE URBCON (G/M3)
      XM=-1.023
      S=0.803
      CALL GGNLG(DSEED,N,XM,S,R)
      URBCON(I)=R
C  SAMPLE AGRCON (G/M3)
      XM = -1.773
      S = 1.017
      CALL GGNLG(DSEED,N,XM,S,R)
      AGRCON(I)=R
C  SAMPLE FORCON (G/M3)
      XM=-4.266
      S=0.204
      CALL GGNLG(DSEED,N,XM,S,R)
      FORCON(I)=R
C  SAMPLE AFXURB (G/M2/YR)
      XM = -2.30
      S = .766
      CALL GGNLG (DSEED,N,XM,S,R)
      AFXURB(I) = R
C  SAMPLE AFXFOR (G/M2/YR)
      XM = -3.719
      S = .510
      CALL GGNLG (DSEED,N,XM,S,R)
      AFXFOR(I) = R
C  SAMPLE AFXAGR (G/M2/YR)
      XM= -3.356
      S = .773
      CALL GGNLG (DSEED,N,XM,S,R)
      AFXAGR(I) = R
C  CALCULATE ATMFLX (G/YR)
      ATMFLX(I)=(AFXURB(I)*.0708+AFXFOR(I)*.6333+AFXAGR(I)*.2962)*LSA
C
```

```
m
C GENERATE A NORMAL RANDOM VARIATE FOR EACH OF THE FOUR FOLLOWING
C MODEL PARAMETERS.
C NOTE =  THIS IMSL RANDOM NUMBER GENERATOR GENERATES STANDARD NORMAL
C VARIATES WHICH MUST BE "DE-STANDARDIZED".
m
C SAMPLE ONTFLO (M3/YR)
      M = 27500000000.
      S = 6240000000.
100   CALL GGNML(DSEED,N,R)
      ONTFLO(I)=R*S+M
      IF (ONTFLO(I).LT.0.) GO TO 100
C SAMPLE PFXPRI
      M = 66500000.
      S = 21900000.
 110  CALL GGNML (DSEED,N,R)
      PFXPRI(I) = R*S+M
      IF(PFXPRI(I).LT.0.) GO TO 110
C SAMPLE PFXSCD
      M = 784000000.
      S = 251000000.
 115  CALL GGNML (DSEED,N,R)
      PFXSCD(I) = R*S+M
      IF(PFXSCD(I).LT.0.) GO TO 115
C SAMPLE PFXPRE
      M = 827000000.
      S = 430000000.
 120  CALL GGNML (DSEED,N,R)
      PFXPRE(I) = R*S+M
      IF(PFXPRE(I).LT.0.) GO TO 120
C CALCULATE PNTFLX (G/YR)
      PNTFLX(I) = PFXPRI(I)+PFXSCD(I)+PFXPRE(I)
m
```

```
C
C EREFLO IS REPRESENTED BY THIS ALGORITHM WHICH GENERATES RANDOM LAG-ONE
C CORRELATED FLOW VALUES.
C
C SAMPLE EREFLO (M3/YR)
      N=1
  125 CALL GGNML(DSEED,N,R)
      LSTFLO = 184810000000.
      MKC1 = 36993000000.
      MKC2 = 6290100000.
      LAGONE=.7998
      SNRN=R
      IF(SNRN.LT.0.) GO TO 125
      EREFLO(I) = MKC1+(LAGONE*LSTFLO)+(MKC2*SNRN)
      LSTFLO = EREFLO(I)
C
```

```
C
C  THE FOUR REMAINING PARAMETERS ARE ALSO REPRESENTED BY NORMAL RANDOM
C  VARIATES.
C
C  SAMPLE ERECON (G/M3)
       M=.022
       S=.002
C  NIAGARA-WELLAND MEAN .022, SD= .002    E.ERIE  MEAN =.017, SD= .004

130    CALL GGNML(DSEED,N,R)
       ERECON(I)=R*S+M
       IF (ERECON(I).LT.0.) GO TO 130
C  SAMPLE VS (M/YR)
       M=16.02
       S=4.67

140    CALL GGNML(DSEED,N,R)
       VS(I)=R*S+M
       IF (VS(I).LT.0.) GO TO 140
C  SAMPLE TAU (YEARS)
       M=7.595
       S= 1.065
  150  CALL GGNML (DSEED,N,R)
       TAU(I) = R*S+M
       IF(TAU(I).LT.0.) GO TO 150
C  SAMPLE ATMLOD (M3/YR)
       M=16500000000.
       S=3040000000.
  160  CALL GGNML (DSEED,N,R)
       ATMLOD(I) = R*S+M
       IF(ATMLOD(I).LT.0.) GO TO 160
C
```

```
C
C  MODEL TERMS ARE AGGREGATED AND A DEPENDENT VARIABLE PREDICTION IS
C  CALCULATED.
C
      DIFFLX(I) = ((.0629+LUFCFU)*URBCON(I)+(.3028+LUFCFA)*AGRCON(I)+
     +(.6343+LUFCFF)*FORCON(I))*ONTFLO(I)
       EREFLX(I) = EREFLO(I)*ERECON(I)
       TOTFLX(I) = DIFFLX(I)+PNTFLX(I)+ATMFLX(I)+EREFLX(I)
       SIGLOD(I) = ONTFLO(I)+EREFLO(I)+ATMLOD(I)
       QS(I)=SIGLOD(I)/LSA
       L(I)=TOTFLX(I)/LSA
       P(I)=L(I)/(VS(I)+QS(I))
       LOGTRM(I) = EXP(-1/TAU(I)-VS(I)/Z)
       TRMONE(I) = P(I)*(1-LOGTRM(I))
       TRMTWO(I) = PLAST*LOGTRM(I)
       PDYN(I) = TRMONE(I)+TRMTWO(I)
C
C
C  INCREMENT THE RANDOM SEED SUPPLEMENT.
C
      XSEED=XSEED+10
C
C
C  THIS ENDS THE INNER LOOP WHICH IS REPEATED 100 TIMES FOR 100
C  PREDICTIONS OF A SINGLE YEAR'S IN-LAKE PHOSPHORUS CONCENTRATION.
C
  500 CONTINUE
C
```

```
C
C  SUBROUTINE PPSELCT                DECEMBER 1980
C  BY V. DAVID LEE
C  CALLED BY SOURCE PROGRAM MCS
C  DEPARTMENT OF RESOURCE DEVELOPMENT
C  MICHIGAN STATE UNIVERSITY
C
C
C  THIS SUBROUTINE IS CALLED BY PROGRAM MCS TO GENERATE A MEAN FROM THE
C  PREVICUS YEAR"S DISTRIBUTION FOR USE AS PLAST IN THE SUCCEEDING
C  YEAR"S PREDICTION CALCULATIONS.
C  CALCULATION OF SUMMARY STATISTICS
C
      SUBROUTINE PPSELCT (PDYN,K,JCLS,PLAST,STDEV)
      DIMENSION PDYN(100)
      X = PDYN(1)
      DO 102 I=1,99
      X=X+PDYN(I+1)
  102 CONTINUE
      PLAST = X/100
      RETURN
      END
C
```

```
C
C ANNUAL INCREMENTATION OF LAND USE FRACTIONS = FIRST SET FOR YEARS
C 1981 - 2000, SECOND SET FOR YEARS 2001 - 2020.
C
      IF (K.GT.20) GO TO 165
      LUFCFU = LUFCFU+.00049
      LUFCFA = LUFCFA-.00051
      LUFCFF = LUFCFF+.00002
      GO TO 167
  165 LUFCFU = LUFCFU+.00030
      LUFCFA = LUFCFA-.00015
      LUFCFF = LUFCFF-.00015
C
C
C SUBROUTINE PPSELCT MUST BE CALLED.
C
  167 CALL PPSELCT (PDYN,K,JCLS,PLAST,STDEV)
      CALL SORTIT(PDYN,K,JCLS,PLAST,STDEV)
      CALL GROUPIT (PDYN,K,JCLS,PLAST,STDEV)
      CALL SUMSTAT (PDYN,K,JCLS,PLAST,STDEV)
C
C
C REDEFINES PLAST AS MEAN OF DISTRIBUTION OF PREVIOUS YEAR"S PREDICTION
C DISTRIBUTION.
C
C SAMPLE PLAST
      M = PLAST.
      S = STDEV
  625 CALL GGNML (DSEED,N,R)
      PLAST = R*S+M
      IF (PLAST.LT.0.) GO TO 625
C
C
C OPTIONAL PARAMETER PRINTS AND SUBROUTINE CALLS MAY BE INSERTED HERE.
C
C
C THIS ENDS THE OUTER LOOP WHICH IS REPEATED 40 TIMES TO GENERATE
C PREDICTIONS FOR YEARS 1981 - 2020.
C
  600 CONTINUE
      END
C
```

```
C
C  SORT PREDICTIONS BY MAGNITUDE.
C  SUBROUTINE SORTIT                    DECEMBER 1980
C  BY V. DAVID LEE
C  CALLED BY SOURCE PROGRAM MCS
C  DEPARTMENT OF RESOURCE DEVELOPMENT
C  MICHIGAN STATE UNIVERSITY
C
C  THIS SUBROUTINE SORTS PREDICTIONS BY MAGNITUDE, SMALLEST TO LARGEST.
C
      SUBROUTINE SORTIT (PDYN,K,JCLS,PLAST,STDEV)

      DIMENSION PDYN(100)
      N = 100
      DO 240 I = 1,100
      DO 230 J=1,N
      IF(PDYN(J).LE.PDYN(I)) GO TO 230
      SWTCH = PDYN(I)
      PDYN(I) = PDYN(J)
      PDYN(J)= SWTCH
  230 CONTINUE
  240 CONTINUE
      PRINT 257,K
  257 FORMAT (* THE ORDERED PREDICTIONS FOR YEAR *,I3,* ARE*)
      PRINT 265,PDYN
  265 FORMAT (20(3X,5(F17.4,3X)/))
C  THESE ORDERED PREDICTIONS ARE ORDERED BY MAGNITUDE, THEY ARE NOT IN
C  TEMPORAL SEQUENCE.
      RETURN
      END
C
```

```
      SUBROUTINE GROUPIT
C
C GROUP PREDICTIONS
C SUBROUTINE GROUPIT                    DECEMBER 1980
C BY V. DAVID LEE
C CALLED BY SOURCE PROGRAM MCS
C DEPARTMENT OF RESOURCE DEVELOPMENT
C MICHIGAN STATE UNIVERSITY
C
C
C THIS SUBROUTINE GROUPS PREDICTIONS IN ONE MICRO-GRAM PER LITER CLASSES.
C CALCULATION OF DISTRIBUTION MEAN
```

```
      SUBROUTINE GROUPIT (PDYN,K,JCLS,PLAST,STDEV)
      DIMENSION PDYN(100),JCLS(10)
      DO 100 L=1,10
  100 JCLS(L)=0
      DO 500 I=1,100
      IF(PDYN(I).LT..015) GO TO 480
      IF(PDYN(I).GE..016) GO TO 200
      JCLS(1)=JCLS(1)+1
      GO TO 500
  200 IF(PDYN(I).GE..017) GO TO 220
      JCLS(2)=JCLS(2)+1
      GO TO 500
  220 IF(PDYN(I).GE..018) GO TO 240
      JCLS(3)=JCLS(3)+1
      GO TO 500
  240 IF(PDYN(I).GE..019) GO TO 260
      JCLS(4)=JCLS(4)+1
      GO TO 500
  260 IF(PDYN(I).GE..020) GO TO 280
      JCLS(5)=JCLS(5)+1
      GO TO 500
  280 IF(PDYN(I).GE..021) GO TO 300
      JCLS(6)=JCLS(6)+1
      GO TO 500
  300 IF(PDYN(I).GE..022) GO TO 320
      JCLS(7)=JCLS(7)+1
      GO TO 500
  320 IF(PDYN(I).GE..023) GO TO 340
      JCLS(8)=JCLS(8)+1
      GO TO 500
  340 IF(PDYN(I).GE..024) GO TO 360
      JCLS(9)=JCLS(9)+1
      GO TO 500
  360 IF(PDYN(I).GE..025) GO TO 490
      JCLS(10)=JCLS(10)+1
      GO TO 500
  480 PRINT*,"A LOWER OUTLIER EXISTS."
      GO TO 500
  490 PRINT*,"AN UPPER OUTLIER EXISTS."
  500 CONTINUE
      PRINT 510, K
  510 FORMAT (* THE CLASS FREQUENCIES FOR YEAR *,I3,* ARE*)
      PRINT 515,JCLS
  515 FORMAT(30X ,10(I5,5X))
      RETURN
      END
```

```
C SUBROUTINE SUMSTAT                DECEMBER 1980
C BY V. DAVID LEE
C CALLED BY SOURCE PROGRAM MCS
C DEPARTMENT OF RESOURCE DEVELOPMENT
C MICHIGAN STATE UNIVERSITY
C
C
C THIS SUBROUTINE IS CALLED BY PROGRAM MCS TO GENERATE A MEAN, STANDARD
C DEVIATION, COEFFICIENT OF VARIATION, AND MODE TO MEAN RATIO FOR THE
C FIRST PARAMETER STATED IN THE SUBROUTINE ARGUEMENT.
C
      SUBROUTINE SUMSTAT (PDYN,K,JCLS,PLAST,STDEV)
      DIMENSION PDYN(100),Y(100)
      T = PDYN(1)
      DO 102 I = 1,99
      T = T+PDYN(I+1)
  102 CONTINUE
      XMEAN = T/100
      DO 104 I=1,100
      Y(I) = (PDYN(I)-XMEAN)**2
  104 CONTINUE
      X = Y(1)
      DO 106 I=1,99
      X = X+Y(I+1)
  106 CONTINUE
      VAR = X/99
      STDEV = SQRT(VAR)
      CV = STDEV/XMEAN
      R = (1+CV**2)**(-1.5)
      PRINT 108,K
  108 FORMAT(*- THE MEAN,SD,CV,AND R VALUES FOR YEAR *,I3)
      PRINT 110,XMEAN,STDEV,CV,R
  110 FORMAT(40(3X,4(F15.5,3X)/))
      RETURN
      END
```