© 1977

WILLIAM WILMOT WALKER, Jr.

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they ere spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed .the photographer followed e definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

SOME ANALYTICAL METHODS

APPLIED TO LAKE WATER QUALITY PROBLEMS

A thesis presented

by

William Wilmot Walker, Jr.

to

The Division of Engineering and Applied Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Engineering

Harvard University

Cambridge, Massachusetts

February, 1977

THESIS ABSTRACT

"Some Analytical Methods Applied to Lake Water Quality Problems"

William W. Walker, Jr.

A variety of analytical methods are demonstrated and evaluated in the context of assessing lake water quality problems. The techniques are drawn from the general areas of exploratory data analysis, parameter estimation, sensitivity analysis, and error analysis. Both empirical and theoretical approaches are taken in examining and modelling the behavior of a cross-section of lakes, as well as the behavior of a single lake in temporal and spatial domains. Proper use of these techniques is suggested as a means of improving efficiencies and assessing inaccuracies in converting scientific principles and observed data into environmental impact projections and management strategy designs.

Data characterizing the phosphorus balances, morphology, and hydrology of 105 northern temperate lakes provide a basis for demonstrating empirical approaches. Using nonlinear regression techniques, a two-parameter model for the phosphorus retention coefficient is shown to have stable parameter values, but increasing standard errors of estimate across the oligotrophic, mesotrophic, and eutrophic states. An error analysis estimates the independent variable, parameter, and model error components of retention coefficient and lake phosphorus concentration predictions. Stepwise discriminant and principal component analyses are employed to develop a model for estimating lake trophic state, expressed in probabilistic terms, from uncertain estimates of phosphorus loading, mean depth, and hydraulic residence time. The implications for the design of monitoring programs to provide data for use of these models are discussed. The limitations of the models are assessed in relation to various characteristics of the data base upon which they are developed.

Extensive water quality data from Onondaga Lake, New York, and its tributaries provide a basis for demonstrating the roles of and methods for exploratory data analysis in preliminary assessments. Theoretical approaches are taken in developing a model for vertical stratification in the lake. The model is estimated using nonlinear programming algorithms for locating maximum likelihood estimates of parameters in dynamic systems. The stability of optimal parameter estimates under different hydrologic and meteorologic conditions is demonstrated using data from different years of the survey. Error and sensitivity analyses are performed in applying the model to assess the effects of an outfall design on vertical mixing in the lake.

Also discussed are the potential roles of the methods demonstrated in addressing some of the problems and deficiencies which have been characteristic of efforts at modelling natural systems, specifically including: parameter estimation problems (particularly in dynamic models); imbalances with regard to model complexity and data availability; misuse of model projections, as related to the scarcity of attempts to estimate confidence limits and to quantify error sources; misuse of models, as related to misinterpretations of "verification" tests and results. Control of these types of problems is characterized as being more difficult in systems where relatively complex models must be employed and where the quality and quantity of independent variable data are relatively low. The stability of optimal parameter estimates across lakes and along temporal dimensions within lakes is suggested as an effective indicator of model generality and a useful criterion for model verification.

PREFACE

Our civilization is such that environmental quality objectives often
apparently conflict with other short- and long-term interests and require-
ments. We cannot express the costs and benefits of environmental quality
management on the same scales and rely upon our political process to weigh
and compare them in selecting, in theory, the best options for society as
a whole. In contributing to this process, the scientific and engineering
community has assumed responsibilities to assess the cultural impacts upon
the environment and to suggest specific, feasible alternatives for achiev-
ing quality standards. These efforts require functional representations
of natural systems, or "models" which transcend the descriptive analyses
historically characteristic of environmental studies.

If the environmental modelling community were assembled along a gradi-
ent of empiricism, the extreme left might be found performing stepwise lin-
ear regressions, without much regard for physics, chemistry, or biology,
and the extreme right, integrating hundreds of simultaneous differential
equations, without much regard for observed data. At an inquisition, the
left could be accused of curve-fitting and proposing models with limited
realism and generality, and the right, of being narrow-minded and ignoring
the truth in the data. Most of the community would attempt to use both
data and theory as bases for developing models. For analysis of most envi-
ronmental systems, efficiency would be maximized somewhere in the middle,
as determined by quantity and quality of data and by extent and validity
of scientific theory. However, the middle is also an area with many pit-
falls, not the least hazardous of which is to accept correlation as suf-
ficient proof of causation.

At its conception during undergraduate years, my approach to environ-
mental modelling was, I suppose, somewhere to the right of the middle, a
position attributed chiefly to a background in physical science. Experience
in the Environmental Systems Program at Harvard has developed my ability
to move to the left, according to the needs of a particular problem. Hope-

fully, it has also increased my awareness of the many pitfalls in this area. Interactions with faculty and students working on problems that were, for the most part, quite different from my own provided exposure to analytical techniques and approaches which were generally new to me, but which seemed promising as means of fortifying my own approaches. Identifying those techniques which were both valid and useful for application to the types of problems of my concern was a difficult and frustrating educational exercise, aided considerably by conversations with faculty members and fellow students.

The guidance of Professor Joseph J. Harrington has been invaluable. Based upon his experience with a wide variety of techniques and upon his general grasp of the environmental field, he has provided many stimulating suggestions and thoughtful responses. His proposal that my thesis focus on the concept of technique rather than on a particular theory or model eventually eliminated much doubt and confusion as to possible directions for my graduate studies. His demonstrations of faith and patience have earned him my life-long gratitude and respect.

Professor Harrington's guidance has been more-than-adequately supplemented by the comments and suggestions of Professors Ralph Mitchell, Frederick E. Smith, and Harold A. Thomas, Jr. The time, interest, and insights of these gentlemen are gratefully acknowledged.

The opportunity to exchange ideas with fellow students in the program has been a valuable aspect of my graduate experience. Associations with Kenneth H. Reckhow, who shares an interest and involvement in lake water quality problems, have been particularly meaningful.

Extensive data from Onondaga Lake, New York, have provided an excellent basis for many of the analytic efforts in the thesis. The Onondaga County program is one of the most intensive monitoring efforts ever undertaken in and around a lake of this size. I salute the County for its en-

deavors and hope that proper use of the data they have gathered will some-
day contribute to improvements in the lake's unfortunate condition. The
following people have been especially helpful in providing information and
insight relative to Onondaga Lake and its environs : Dr. Cornelius B. Mur-
phy, Jr. and Mr. Greg Welter (O'Brien and Gere, Engineers); Mr. Randy R.
Ott and Mr. Donald Lawler (Onondaga County Department of Drainage and Sani-
tation); Mr. William Embree, Mr. Kenneth Darmer, and Mr. Al Randal (U.S.
Geological Survey); Dr. Jay Bloomfield (New York State Department of Envi-
ronmental Conservation); Mr. James Rooney (U.S. E.P.A., Region II); Mr.
Richard Clough (Allied Chemical Corporation). Their interest and coopera-
tion are gratefully acknowledged.

I am grateful to the National Science Foundation for its financial
support of these efforts.

The moral and clerical support of Ms. Martha J. Ploetz have been
an essential factor in the draft and final preparation of this volume.

Finally, I would like to dedicate this work to my parents.

William W. Walker, Jr.
Harvard University
February 1977

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

SYNOPSIS

Our ability to assess the impacts of cultural interventions upon natural systems provides a partial basis for design of resource management programs and for policy-making in the midst of other societal interests. While, historically, much of environmental management has been dictated by trial and error, we now require more rational methods which facilitate both the organization of experience at the system level (empirical approaches) and construction of mechanistic models with projective capabilities (theoretical approaches). These approaches entail functional understanding which transcends the purely descriptive analyses historically characteristic of environmental studies. Rational assessments rely upon two general information sources: data and scientific principles. This thesis is intended as a demonstration and evaluation of some techniques for synthesizing these information sources in formulating environmental impact assessments. A variety of quantitative techniques are demonstrated and evaluated in the context of analyzing lake water quality problems. These techniques are drawn from the general areas of exploratory data analysis, parameter estimation, sensitivity analysis, and error analysis. They are appropriate for use at various stages of the analytic process and are employed in examining the behavior of a cross-section of lakes, as well as that of a single lake in temporal and spatial domains.

Assesaments of cultural eutrophication problems in lakes are dependent upon two types of models: source models and lake models. The factors driving these systems and thus determining lake response can be classified as naturally or culturally mediated. The former are uncontrollable, while

SYNOPSIS (continued)

the latter often represent decision variables or control points for the
design and implementation of management strategies. A review of the
literature suggests that, while much data have been compiled, (e.g.,
nutrient export versus land use), the development of generalized source
models is considerably behind that of lake models. The latter can be
classified as "theoretical" or "empirical", according to whether or not
they attempt to eimulate internal mechanisms. These types of models are
reviewed and compared with regard to complexity, applicability, accuracy,
and generality.

Quantitative assessments of possible errors involved in use of source
or lake models are generally lacking, suggesting that, as yet, there is
an insufficient basis for rational lake water quality management program
designs. In predicting lake response to a given management strategy,
errors are introduced through uncertainty in the independent variables
and parameter estimates, as well as through inadequacies in the models
themselves. Given estimates of these individual error sources, some fairly
simple methodologies can be employed to track these errors through the
analysis and to compare their contributions to total prediction error.
This can identify particular data or model deficiencies and thus provide
a basis for efficient allocation of monitoring, experimental, and modelling
effort. The effects of projection error can also be considered in final
designs or policy recommendations. The difficulty and data requirements
of such an exercise increase with model complexity.

Based upon data for 105 northern temperate lakes, empirical lake
models are developed and compared with existing models for steady-state
predictions of phosphorus retention coefficients and lake trophic states
as functions of three factors: total phosphorus loading, L (g/m^2-year) ,
lake mean depth, Z (m) , and mean hydraulic residence time, T (years) .
The approach is based upon five assumptions: (1) mass balance; (2) com-
pletely-mixed conditions; (3) phosphorus limitation of lake ecosystems;
(4) first-order kinetics for phosphorus sedimentation; (5) the possible
influence of lake morphometric and hydrologic factors upon nutrient dynamics
and trophic state response. The accuracy and value of these models are
shown to be limited by various features of their data base, including
measurement or estimation errors in the independent or dependent variables,
multicollinearity, nonsteady-state conditions existing in the lakes during
sampling, and the subjectivity involved in classifying lakes. In this
context, it is difficult to identify model deficiencies due to effects of
other controlling factors or aspects of system behavior which are ignored.
Because of the data-dependency of these empirical models, their general
superiority over existing models is not claimed. The primary emphases
are upon the approaches taken and techniques employed in their development
and evaluation and upon the strategies proposed for their application. A
general focus is upon the development and use of error estimates.

Modifying Vollenwieder's[9] model based upon first-order kinetics, the
following expression is proposed for use in predicting the phosphorus
retention coefficient:

SYNPOSIS (continued)

$$1 - RE = \frac{1}{1 + KT}$$

$$K = a\left(\frac{1}{T}\right)^b z^c$$

Using nonlinear regression techniques, the model is shown to have stable parameter values, but increasing standard errors of estimate across the oligotrophic, mesotrophic, and eutrophic states. The parameter expressing depth dependence, c, is not found to be significantly different from zero in any of the data groups tested. The best estimates of parameters a and b suggest the following equation:

$$1 - RE = \frac{1}{1 + 0.824T^{.454}}$$

Through an error analysis, the accuracy of a lake phosphorus prediction is shown to be data-limited (or source-model-limited) in lakes with low hydraulic residence times and retention-model-limited in lakes with high residence times. It is demonstrated that model errors would be statistically detectable relative to measurement errors in only about one third of the lakes employed in the analysis. By coupling this model with Dillon and Rigler's[3] equation relating chlorophyll concentration to lake phosphorus concentration, the accuracy of a chlorophyll prediction is shown to be limited by errors in predicting chlorophyll from lake phosphorus, not by errors in predicting lake phosphorus from estimates of L, Z, and T. This suggests that the retention model is adequate for use with state-of-

SYNPOSIS (continued)

the-art empirical chlorophyll models and that future work should be focused
on improving the latter, possibly by incorporating the effects of algal
growth limitation by light or other nutrients.

Discriminant analysis is employed to evaluate and compare existing
models for trophic state prediction and to derive an optimal linear classi-
fication model for this data set. The analysis suggests the general
inferiority of Vollenweider's[8] first model, in which lakes are classified
in the L versus Z axes, as compared with Vollenweider's[10] second model
(L versus QS), with Dillon's[2] model (CO), or with an optimal model
derived from stepwise discriminant analysis (L versus COEST). It is
more difficult to distinguish among the last three models, however. The
stepwise model misclassifies an average of 14.5% of the lakes, with most
of the errors centered in the mesotrophic group.

The primary advantage of posing the classification problem as a
formal discriminant analysis is that it permits predictions of trophic
states on probabilistic terms. Using the results of the stepwise analysis,
equations are developed which permit estimation of trophic state probabi-
lities as functions of L, T, and Z. The "rational" phosphorus loading
allocation for a given lake is suggested as that loading which satisfies
a certain maximum probability (risk) of eutrophic classification, typically
5%:

$$L_{.05} = 0.0310 \left[\frac{Z}{T} (1 + 0.824T^{.454}) \right]^{.815}$$

SYNPOSIS (continued)

Means of incorporating the effects of uncertainty in the independent variables (Z , T) and in the source models (used to estimate L as a function of management strategy) into the estimation of trophic state probabilities and into the rational loading allocation scheme are also developed. Efficiencies of loading allocations derived from the approach are shown to be limited in part by independent variable errors and dependent variable (classification) errors in the original data set.

Extensive data from Onondaga Lake, New York,[4,5] and its tributaries are used to illustrate potential roles and methods for exploratory data analysis and data reduction in preliminary assessments. Onondaga is a saline, eutrophic lake used primarily for municipal and industrial waste disposal purposes by Metropolitan Syracuse. The lake is introduced by describing major water quality issues, plans for pollution abatement, and water quality and quantity data availability. The variations of major water quality components are summarized and displayed along temporal and spatial dimensions in order to elucidate trends and seasonal variations and to determine the extents of horizontal and vertical mixing in the lake. The use of cubic splines[7] as a technique for examining seasonal and long-term variations in time series data is demonstrated along with a variety of computerized mapping and display technologies. Original methods for estimating and displaying mass balances on a continuous basis are also employed.

Statistically significant horizontal variations in some water quality

SYNOPSIS (continued)

components are explained by the location of a sewage outfall in the lake.
Horizontal differences are small, however, compared with vertical and
seasonal variations, which are interpreted relative to the effects of
dilution, density, and differences in the chemical and biological reactions
occurring in the epilimnion and hypolimnion. Some important trends observed
during the 1968-75 monitoring period include: reductions in phosphorus,
chromium, and silica levels, reductions in vertical temperature and chloride
gradients, increases in hypolimnic dissolved oxygen levels, and the near
disappearance of a formerly dominant blue-green algal population. Mass
balances on several chemical constituents are formulated at monthly inter-
vals over a five-year period and interpreted (1) in light of the chemical,
physical, and biological processes considered to be of importance in the
lake ecosystem; (2) in relation to specific pollution abatement programs
which were implemented during that period; (3) with regard to the impli-
cations for the potential improvement of lake water quality through future
implementation of specific point and non-point source control measures.

A modelling problem is designed to estimate the impact of the design
of an outfall for future discharge of a saline industrial/municipal waste
upon density stratification and vertical mixing rates in Onondaga Lake.
The dynamic model developed to address this issue serves as a context in
which to demonstrate the use of nonlinear programming algorithms for esti-
mation of parameters in dynamic models, sensitivity analysis, and error
analysis. In this effort, considerable modifications to Bard's[1] program
for parameter estimation in nonlinear dynamic systems are made to improve

SYNOPSIS (continued)

performance, flexibility, and the interpretative value of its output.

The modelling exercise demonstrates that essential aspects of vertical mixing in the lake can be simulated using a spatially and temporally aggregated model, consisting of two mixed compartments driven by monthly-average meteorologic and hydrologic boundary conditions. The model simulates temperature and chloride variations, based upon balances of mechanical energy, thermal energy, and mass. Parameter estimates indicate that density-dependent vertical exchange in the lake can be simulated by assuming that a constant percent (47%) of the turbulent kinetic energy introduced at the surface by wind shear stress is used to increase the buoyant potential energy of the system by mixing the hypolimnic and epilimnic waters. Empirical parameters are shown to be stable when estimated separately based upon data from different years of the survey. An error analysis indicates that prediction variance associated with uncertainty in the parameter estimates is generally small compared with residual variance, suggesting the adequacy of the data for parameter estimation purposes. The standard errors of temperature predictions are shown to be comparable to those typical of much more complex and disaggregated thermal stratification models[6].

A sensitivity analysis compares the effects of wind mixing, salt discharge, and industrial cooling water consumption on lake density gradients. Simulated vertical mixing rates correlate with observed increases in hypolimnic dissolved oxygen levels over the 1968-74 monitoring

SYNOPSIS (continued)

period. Simulations indicate that discharge of the combined municipal/
industrial effluent into the hypolimnion would induce permanent density
stratification under the meteorologic conditions of 1968-74, whereas
discharge into the epilimnion at an initial dilution ratio of 16 or greater
would result in an average of nearly two periods of vertical circulation
annually, thus minimizing the impact of salt disposal on lake mixing.
Variations in simulated density gradients and mixing rates attributed
to parameter uncertainty are small relative to those attributed to year-
to-year fluctuations in hydrologic/meteorologic conditions and to those
induced by alternative management strategies. The variety of economic
and environmental aspects of the outfall design issue are also discussed.

A final chapter summarizes the roles and limitation of the various
methods and approaches demonstrated in the thesis. Exploratory data
analyses are useful in providing important descriptive information, but
independent evidence generally forms the basis for the functional under-
standing required to predict system behavior. The potential of nonlinear
programming algorithms for use in estimating parameters in general models
is high, subject to economic constraints on model complexity and require-
ments for reasonably unbiased estimates of independent variables. The
importance of considering the quantity and quality of available independent
variable data in specifying models is stressed. For a variety of reasons,
error analyses yield very approximate results, particularly in cases of
complex models. Comparisons of error sources can still yield useful infor-
mation for assessing model and data adequacies, because the terms of the

total error equation often differ by orders of magnitude.   Estimates of
prediction errors should not be relied upon too heavily as bases for prob-
abilistic projections and rational designs, particularly if a complex model
is employed.

Generality is a key model attribute.   Empirical lake models have de-
monstrated reasonable generality in associating the average conditions of
lakes in a given geographic region with nutrient loading and morphometric
factors.   Theoretical lake models have successfully simulated behavior of
individual lakes in time.   The generality of empirical models along tempo-
ral dimensions and that of theoretical models among different lakes have
not been demonstrated, due to inadequacies in the data and/or in the models
themselves.   These weaknesses suggest that future efforts should strive
to demonstrate generality by estimating and verifying models based upon
time series data from more than one lake system.   Using the estimation
techniques demonstrated here, the stability of optimal parameter estimates
along temporal and system dimensions could be examined and used as a partial
basis for assessment of generality.   Applied to key parameters, this stra-
tegy would partially eliminate the necessity of using parameters derived
from laboratory experiments, the results of which are often of limited
validity under field conditions.   Such an approach would depend upon the
availability of adequate data and upon the feasibility of expressing es-
sential functional relationships in concise terms, based upon current under-
standing of lake ecosystems.

LIST OF SYMBOLS - SYNOPSIS

Z = lake mean depth   (m)

T = mean hydraulic residence time   (years)

QS = Z/T = surface overflow rate (m/year)

L = total phosphorus loading per unit area $(g/m^2 year)$

CI = average inlet total phosphorus concentration (g/m^3)

CO = average outlet total phosphorus concentration (g/m^3)

R = $1 - (CO/CI)$ = observed phosphorus retention coefficient (dimensionless)

RE = estimated phosphorus retention coefficient (dimensionless)

$COEST$ = $(1 - RE) CI$ = estimated average outlet total phosphorus concentration (g/m^3)

K = effective first order sedimentation coefficient for total phosphorus $(year^{-1})$

a,b,c = regression parameters

REFERENCES - SYNPOSIS

1. Bard, Y., "Nonlinear Parameter Estimation and Programming", Program No. 360D-13.6.003, Share Program Library Agency, Triangle Universities Computation Center, Research Triangle Park, North Carolina, December 1967.

2. Dillon, P.J., "The Phosphorus Budget of Cameron Lake, Ontario: The Importance of Flushing Rate to the Degree of Eutrophy of Lakes", Limnology and Oceanography, Vol. 20, No. 1, pp. 28-45, January 1975.

3. Dillon, P.J. and F.H. Rigler, "The Phosphorus-Chlorophyll Relationship in Lakes", Limnology and Oceanography, Vol. 19, No. 1, pp. 767-773, September 1974.

4. O'Brien and Gere, Engineers, "Onondaga Lake Monitoring Program", Annual Report to Onondaga County, Department of Public Works, 1970-74.

5. Onondaga County, New York,"Onondaga Lake Study", US EPA, Water Quality Office, Publication 11000 FAE 4/71, April 1971.

6. Parker, F.L., B.A. Benedict, and C. Tsai, "Evaluation of Mathematical Models for Temperature Prediction in Deep Reservoirs", National Environmental Research Center, Office of Research and Development, US EPA, Document No. EPA-660/3-75-038, June 1975.

7. Reinsch, C.H., "Smoothing by Spline Functions", Numerical Mathematics, Vol. 10, No. 117, 1967.

8. Vollenweider, R.A., "Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, with Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication", Organization of Economic Cooperation and Development, Paris, September 1970.

9. Vollenweider, R.A. "Possibilities and Limits of Elementary Models Concerning the Budget of Substances in Lakes", Arch. Hydrobiol., Vol. 66, No. 1, pp. 1-36, April 1969.

10. Vollenweider, R.A., "Input-Output Models", unpublished manuscript, 1973.

1.0 INTRODUCTION

1.1 Objectives

Increasing demands imposed upon our natural resources as a result of population growth, urbanization, and industrialization have made it necessary to take specific steps for protection of these resources. Our ability to manage them depends in part upon our understanding of the important structural and functional aspects of the natural systems which we are trying to protect. This understanding provides a basis for impact assessment and input to a decision-making process involving economic and social, as well as environmental interests. Methods for interpretation and synthesis of data and theory in formulating impact assessments are of primary concern in this work.

Environmental impact assessment generally entails the use of models and data. Models, defined as "abstract representations of form and function"[6], may range from "rules of thumb", which qualitatively capture essential aspects of system dynamics, to complex, mathematical simulation models. In a management context, the role of a model is to relate "response" variables to "stimulus" variables. Some of the latter typically include potentially controlled boundary conditions or manageable aspects of the system itself. In this way, a model provides a basis for control strategy designs and impact assessments. Data can be grouped into two general categories according to whether they describe boundary conditions or system states and, as such, represent either stimulus or response variables. Both types of data are required for the detection and assessment of problems and for the estimation and application of models.

1.1 Objectives   (continued)

It is evident that, in analyzing a given system, a planner, engineer, or manager must make numerous decisions in the process of gathering, developing, and synthesizing data and theory in order to formulate and apply a model. In evaluating a given policy, the goal of the analyst should be to utilize available monitoring and modelling resources most effectively to creat his "best" estimate of the state of the system under the conditions imposed by that policy. A variety of factors could be considered in the definition of "best", including that the estimate should have minimum bias, minimum variance, and be robust to errors in the data or theory or to errors in judgment or implementation committed by the analyst himself. The hope is that policy recommendations derived from this process are independent of the particular analytic decisions made, provided that the choices are among apparently equally valid approaches and are internally consistent.

Generally, the availability of a wide selection of techniques to the experienced analyst can increase his freedom of choice, perspective, and ability to make most effective uee of the resources available to him. This thesis is intended as a demonstration and evaluation of some techniques appropriate for ecosystem analysis. It is suggested that facility with such techniques should serve as a supplement to, rather than a substitute for familiarity with the physical, chemical, and biological aspects of the systems being monitored and modelled. Various methods are demonstrated in the context of the general problem of lake water quality management, but could be applied as well in analyses of other types of environmental systems.

1.2 Causation and Correlation

In some respects, the general problems confronting modellers of environmental systems are perhaps more closely related to those confronting social scientists than to those confronting physical scientists. Because of the size, complexity, and relatively uncontrollable nature of the systems being studied, research tends to be more non-experimental than experimental in form. While various aspects of the systems can be studied individually in controlled laboratory or semi-controlled microcosm experiments, rarely is there an opportunity to experiment with the system as a whole. Under such conditions, causation can be difficult to establish. Regardless of the level of empiricism of a model, a statistical fit of un-controlled system behavior establishes association, but not causation[4,6].

In systems whose internal forms and functions are well-understood theoretically, causal assumptions are perhaps less risky than in systems whose components are less well-understood and which are represented by more empirical models, particularly if the independent variables of such systems are correlated with each other[4]. Regardless of the type of model, causality cannot be absolutely established in a purely logical sense unless there is an opportunity for controlled experimentation. We can observe changes in environmental systems following accidents or implementation of management strategies, but these may occur in the contexts of other, natural variations and are not controlled situations.

1.2 Causation and Correlation (continued)

On the other hand, environmental systems are real and must conform to the laws of physical and biological science. These types of absolutes do not have analogues in many other non-experimental fields. They include, for example, the law of continuity (mass balance) and the laws of thermodynamics. Models at all levels of empiricism must conform to such laws. These constraints on the systems and models can greatly facilitate analyses and inferences of causation at a practical level. Because of these aspects, a purely statistical approach to the analysis of natural systems would generally not be acceptable. A better approach would be founded upon physical, chemical, and biological principles and flavored with statistical notions of "systems","models", and "data".

1.3 Types of Models

Models exist along a gradient of empiricism. All are ultimately based upon observations, but they differ with regard to the level at which the observations are made. Less empirical, or more theoretical models attempt to simulate mechanisms and are based upon observations of the behavior of isolated components of the system under controlled laboratory conditions and/or upon accepted principles. These models tend to be complex and often require the estimation of relatively large numbers of parameters whose values cannot always be derived directly from laboratory studies. Design of controlled experiments to test mechanistic theories and/or develop parameter estimates for use in large-scale simulation models can be difficult, because of the potential for over-simplification. For this reason, such experimental results often cannot be directly applied to simulations of field conditions. More empirical, or "black box" types of models are based upon observed associations of system behavior with properties of the system itself or of its boundary conditions. Such models cannot be properly applied outside of the range of conditions in which they have been calibrated. Despite this lack of generality, the "black-box" models are usually relatively simple and do not suffer as much from the parameter estimation difficulties characteristic of the more complex, theoretically-based systems models.

Another type of gradient is complexity. Above, it was implied that the complexity and empiricism axes are not independent of each other, in that empirical models tend to be less complex than theoretical

1.3 Types of Models (continued)

ones. However, this is not necessarily the case. Regardless of the
level of empiricism, each model is applied at a level of temporal and
spatial aggregation, which should be selected based upon characteristics
of the system, types and amounts of data available, and types of manage-
ment strategies being evaluated. Another type of complexity pertains
to model characteristics within each spatial element and time domain.
In a theoretical model, this is determined by the number of processes
or mechanisms simulated. In an empirical one, it is a function of the
number of independent and dependent variables employed.

There are numerous trade-offs involved in the specification of
model complexity. If important mechanisms or independent variables
are not included, biased parameter estimates and a distorted view of
system dynamics can result[4,6]. On the other hand, overly-complex
models are cumbersome, not "parsimonious"[5], and have demanding data
requirements, both for specification of boundary conditions and for
estimation of parameters. Errors in such data can lead to distortion
and bias, just as can omission of important variables or processes.
The expense of implementing a complex model can limit the analyst'e
opportunities to fully explore and exploit the model in various ways.
Optimally, a model should probably be as simple as possible, given the
important processes in the system and the intended uses for the model.

1.4 Model Specification, Estimation, and Verification

There are essentially three steps involved in the development of a model: specification, estimation, and verification. These steps, outlined in Figure 1.4-1, rely upon four types of information: (1) established theory; (2) extraneous data obtained from experiments or other natural systems; (3) historical data characterizing the boundary conditions of the system(s) being modelled; (4) historical data characterizing the states of the system(s) being modelled.

Model specification involves definition of (1) the system control boundaries; (2) the variables of importance; (3) equations relating the variables. This step relies upon established theory, as well as upon analysis and interpretation of data characterizing the system and its boundary conditions. The roles of data at this stage can be significant, particularly in highly empirical models. An extreme example would be the use of stepwise linear regression, in which the selection of independent variables is dictated by the data. In another situation, data might be analyzed to determine the levels of spatial and temporal aggregation which are appropriate. The selection of independent or forcing variables at this stage can be critical. If important ones are omitted, biased parameter estimates and a distorted representation of system dynamics could result. This type of problem has led some investigators to believe that, in non-experimental research, model specification is the most important of the three steps described in this section[6]. The hypothesized model

Figure 1.4-1

STAGES IN MODEL DEVELOPMENT

THEORY

MODEL
SPECIFICATION

HYPOTHETICAL
MODEL

EXTRANEOUS
DATA

MODEL
ESTIMATION

CALIBRATED
MODEL

BOUNDARY
CONDITION
DATA

MODEL
VERIFICATION

VERIFIED
MODEL

SYSTEM
DATA
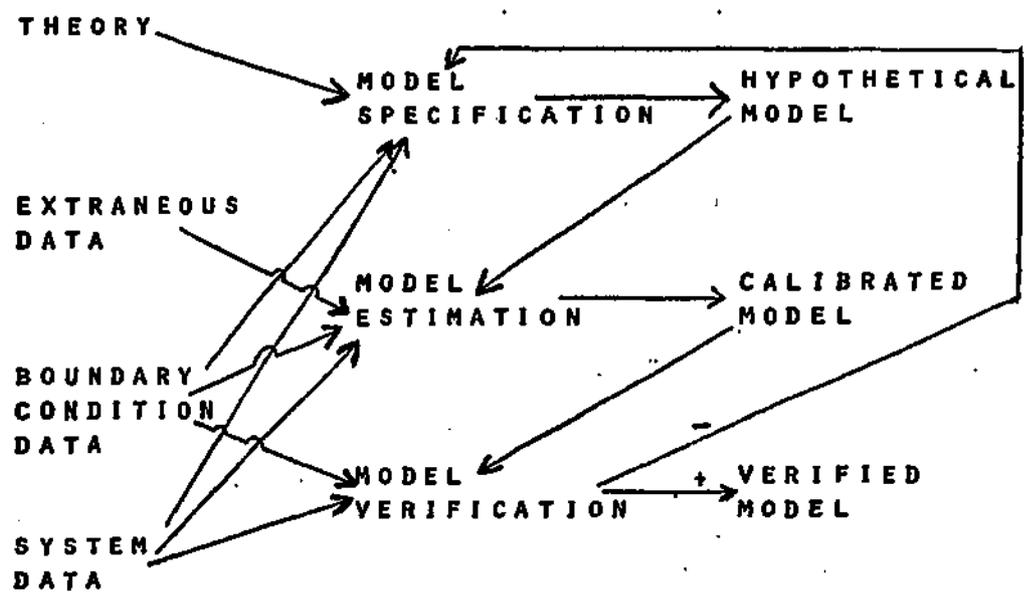
1.4 Model Specification, Estimation, and Verification (continued)

is of the general form*:

$$\underline{y}_i = f(\underline{x}_i, \underline{p}), \quad i = 1, n_i \qquad (1.4-1)$$

where,

$\underline{y}_i$ = vector of state variables for case or observation i, dimension n_y

$\underline{x}_i$ = vector of independent variables of boundary conditions, dimension n_x

$\underline{p}$ = vector of parameter values, dimension n_p

n_i = total number of cases or historical measurements

If the model is being developed to represent a number of different systems, n_i corresponds to the number of cases for which observations are available. If the model is dynamic, and being used to simulate one system, n_i corresponds to the number of times at which system measurements are available. In the latter case, the equation must

---

* Notation: A, $\underline{A}$, and $\underline{\underline{A}}$ represent a scalar, a vector, and a matrix, respectively.

$\underline{A}^T$ = transpose of vector $\underline{A}$

$\underline{\underline{A}}^{-1}$ = inverse of matrix $\underline{\underline{A}}$

det $(\underline{\underline{A}})$ = determinant of matrix $\underline{\underline{A}}$

1.4 Model Specification, Estimation, and Verification (continued)

be integrated through time, so that the elements of y_i are also

functions of the historical values of elements of the y_i and x_i

vectors.

The process of parameter estimation transforms a hypothesized

model into a calibrated one. This involves the use of extraneous data

and/or a matched set of observations made on the boundary conditions

and on the system. Generally, complex, theoretically-based models

have relied more heavily upon the former, while the simpler, empirical

models have relied primarily upon the latter as bases for parameter

estimates. There is no reason, however, why theoretical models could

not rely upon data of both types, provided that model verification is

carried out carefully. The relative difficulty of estimating parameters

of theoretical models based upon system data has probably been the

primary reason for the scarcity of such attempts. The complexity,

and often time-variable nature of these models renders the estimation

problem somewhat more involved, but, in many cases, equally as feasible

as linear least squares, given the availability of computing resources.

Parameter estimation involves specification of the elements of the

vector $\underline{p}$. For those elements which are estimated from observations

of the system, the objective is to select the parameter values which

maximize the agreement between observations and model predictions. In

the case of a one-variable model, ($n_y = 1$) , a commonly employed

objective is to minimize the sums of squares of the residuals:

·1.4 Model Specification, Estimation, and Verification (continued)

$$e_i = y_i - f(\underline{x}_i, \underline{p}) \qquad (1.4 - 2)$$

$$\phi = \sum_{i=1}^{n_i} e_i^2 \qquad (1.4 - 3)$$

Bard[2] has shown that, for the general case, maximum-likelihood estimates of the parameter values can be obtained by maximizing a function of the following form:

$$G_{ML} = -\frac{n_y n_i}{2} \ln(2\pi) - \frac{n_i}{2} \ln[\det(\underline{V}_y)] - \frac{1}{2} \sum_{i=1}^{n_i} \underline{e}_i^T [\underline{V}_y^{-1}] \underline{e}_i \qquad (1.4 - 4)$$

where,

G_{ML} = generalized maximum likelihood function

$\underline{V}_y$ = covariance matrix of residuals

$\underline{e}_i$ = vector of residuals for observation i

G_{ML} represents the logarithm of the posterior probability of the observations, given the predictions derived for a set of parameter estimates. While other general criteria have been employed, maximum likelihood estimators have generally been the most popular for use

1.4 Model Specification, Estimation, and Verification (continued)

in parameter estimation problems[2], because of the desirable properties discussed below.

Parameter estimates derived from maximization of the above function will be unbiased, possess minimum variance, and be normally distributed when the following conditions hold[2]:

(1) The residuals have zero means.*

$$E (e_{ij}) = 0. \qquad \begin{aligned} i &= 1, n_i \\ j &= 1, n_y \end{aligned} \qquad (1.4\text{-}5)$$

(2) The residuals from different observations (cases) are uncorrelated:

$$E (e_{ij} e_{\ell k}) = 0. \qquad \begin{aligned} i &\neq \ell \\ j,k &= 1, n_y \end{aligned} \qquad (1.4\text{-}6)$$

(3) The residuals from each observation have identical normal distributions with covariance matrix $\underline{V}_y$:

$$E (e_{ij} e_{ik}) = [\underline{V}_y]_{jk} \qquad j,k = 1, n_y \qquad (1.4\text{-}7)$$

---

* $E(x)$ = expected value of x

1.4 Model Specification, Estimation, and Verification (continued)

In order to apply equation (4) some assumptions must be made about the appropriate form of the covariance matrix of residuals, $\underline{V}_y$. In one respect, $\underline{V}_y$ can be known beforehand, known to within a multiplicative constant, or unknown. In another, $\underline{V}_y$ can be assumed to be a general, symmetric matrix or a diagonal one. In the diagonal case, the residuals for different observed variables are assumed to be uncorrelated. The appropriate form for the objective function derived from the above general equation depends upon the assumptions made about $\underline{V}_y$. In the general case, $\underline{V}_y$ is assumed to be unknown and with potentially non-zero off-diagonal elements, and equation (4) reduces to:

$$G_{ML} = -\frac{n_i n_y}{2}[1 + \ln(2\pi)] - \frac{n_i}{2}\ln\left[\det\left(\frac{1}{n_i}\underline{A}\right)\right] \qquad (1.4-8)$$

$$\underline{A} = \sum_{i=1}^{n_i} \underline{e}_i^T \underline{e}_i \qquad (1.4-9)$$

where,

$\underline{A}$ = moment matrix of residuals

Likelihood functions appropriate for other assumed forms for the residual covariance matrix, $\underline{V}_y$, are given in Bard.[2] In situations

1.4 Model Specification, Estimation, and Verification (continued)

where prior estimates of the parameters are available, based, for example, upon extraneous experimental data or data from other natural systems, Bayes Theorem[2] can be applied and another term added to G_{ML} to account for the prior information, which is assumed to be given in the form of a vector of mean p^o and covariance matrix V_{p^o} :

$$G_B = G_{ML} - \frac{1}{2} (p - p^o)^T V_{p^o}^{-1} (p - p^o)$$

(1.4-10)

In this way, information obtained both from extraneous sources and from the system itself can be incorporated into the parameter estimates. These extraneous sources could also include the opinions of experts familiar with aspects of the systems represented by various parameters, provided that such "opinions" can be translated into mean parameter values and corresponding standard deviations.

For some models, the parameter vector maximizing G_{ML} can be solved for explicitly, e.g, multiple linear regression. In many cases, however, the models are nonlinear and/or dynamic and finding the maximum of the objective function takes on the form of a nonlinear programming problem. Various search[2] or gradient[1,2] methods can be employed to locate the solution. Since the cost of evaluating the objective function for a given set of parameter values can be appreciable, particularly in the case of a dynamic model, methods which require a minimal number of

1.4 Model Specification, Estimation, and Verification (continued)

function evaluations are generally preferred. Bard[2] recommends the use of gradient methods, which require estimates of the first and, in some cases, second derivatives of the objective function with respect to the parameter values in order to guide convergence. These derivatives can be computed from analytic derivatives of the model equations or by using finite-difference techniques. Starting with an assumed set of parameters, gradient methods are essentially iterative, hill-climbing algorithms with select step length and direction in parameter space to guide convergence to the solution. Typically, the process is stopped when fractional changes in each element of the parameter vector with each iteration become less than some specified amount (~ 0.0001)[2]. The "solution" obtained in such a way cannot be guaranteed to be global, because the response surface may be irregular, with several local optima. Because of this aspect, the algorithm can usually be started from a number of different locations in parameter space to test for the significance of alternative solutions.

The use of nonlinear parameter estimation routines in dynamic modelling of environmental systems has been demonstrated by a number of investigators. Radha Krishnan et al.[14] have done extensive work in evaluating various algorithms for parameter estimation in dynamic models of streams using simulated data. Hornberger et al.[9] have attempted to apply Bard's algorithms[1,2] to experimental data for the purpose of estimating kinetic parameters in algal growth rate formulations. Yih and Davidson[16] have evaluated three algorithms for use in estimating effective longitudinal disper-

1.4 Model Specification, Estimation, and Verification (continued)

sion coefficients in a dynamic salinity intrusion model of the Delaware River Estuary. Johnston and Pilgrim[10] have demonstrated the use of formal optimization methods in estimating the parameters of a watershed model. A recursive estimation algorithm (the Extended Kalman Filter) has been used by Beck and Young[3] in systematically identifying the structure of a model for dissolved oxygen in England's River Cam.

One important advantage in posing the parameter estimation problem in a formal way is that approximate estimates of the covariance matrix of the parameters can be derived[2]:

$$[\underline{\underline{H}}]_{jk} = (\frac{\partial^2 G}{\partial p_j \partial p_k}) \tag{1.4-11}$$

$$\underline{\underline{V}}_p \simeq (\underline{\underline{H}}^{-1})_{\underline{p} = \underline{p}^*} \tag{1.4-12}$$

where,

$\underline{\underline{H}}$ = "Hessian" = matrix of second partial derivatives of the objective function with respect to the parameters

$\underline{\underline{V}}_p$ = covariance matrix of parameter estimates

$G = G_{ML}$ or G_B = log-likelihood function

$\underline{p}^*$ = parameter vector of solution

The required derivatives are estimated in the process of locating the solution, if a gradient method is employed. The covariance matrix

1.4 Model Specification, Estimation, and Verification (continued)

is useful (a) in interpreting the relative significance of various parameters and the processes or mechanisms they represent; (b) in assessing the adequacy of the amount of data used for parameter estimation purposes; (c) in estimating the portion of model prediction error attributed to uncertainty in the parameter estimates; (d) in estimating parameter confidence regions. It should be noted that estimates of the covariance matrix obtained according to equation (12) are only approximate in the case of a nonlinear model.

If unknown, the covariance matrix of residuals can also be estimated from the following formula [2]:

$$\underline{\underline{V}}_r = \underline{\underline{V}}_y = \frac{1}{n_i - \frac{n_p}{n_y}} \underline{\underline{A}} \tag{1.4-13}$$

The covariance matrix is useful in assessing the predictive abilities of the model. The covariances of the residuals can b attributed to measurement errors in the observations and to model error.

Once the parameters have been estimated, regardless of whether or not the formal procedure outlined above is employed, the model is considered calibrated. The next step, verification, (Figure 1.3-1) tests the adequacy of the model for its intended uses. If the model fails the test, the entire specification, estimation, and verification procedure

1.4 Model Specification, Estimation, and Verification (continued)

may be repeated with a revised model. If it passes, the model is considered suitable for application.

The term "verification" has been used and misused with a variety of different meanings in the environmental modelling field. As noted above in Section 1.2, we rarely establish the truth of a model, but rather we test whether or not that truth can be rejected. Because we cannot conduct controlled experiments, "verification" is usually just a test for association and not, in itself, sufficient. We usually rely upon independent evidence or established principles as bases for assumptions of causation, a necessary condition for true "verification". It is important to consider that verification also depends upon the intended uses for the models. Highly empirical models cannot be verified for use under conditions outside of the range in which they have been calibrated, i.e. the range in which the systems themselves have been observed. Whereas theoretical models of relatively well-understood systems may be verified for such use, provided that all of the relevant processes are incorporated and independent means of estimating parameters are available.

The key to testing whether or not a model can be rejected is in analyzing the residuals. A variety of tests can be applied to the residuals in order to determine whether the principle assumptions of the parameter estimation exercise (equations (5) to (7)) have been violated[7]. The residuals can be plotted against the independent variables in the model to reveal possible inadequacies in assumed

1.4 Model Specification, Estimation, and Verification (continued)

functional forms or against variables not included in the model to
determine whether potentially important factors have been ignored.
For dynamic models, serial correlation in the residuals may reflect
the effects of errors in specification of boundary conditions or the
effects of factors not considered in the model. The accuracy of the
model, as gauged by the standard deviations of the residuals, can be
compared with expected measurement errors (which would represent lower
limits for residual standard errors), with the desired accuracy for
model applications or with the accuracies of alternative models. The
last could serve as a partial basis for model selection.

A supplemental criterion for verification is parameter stability.
To test for this, data can be divided into two or more groups and
optimal parameters estimated separately for each group. The parameters
could then be tested for statistically-significant differences among
groups. If differences are not detected, a single set of parameter
values could be considered to be equally appropriate over the range of
conditions represented by the various groups. This would tend to
support the generality of the model and the consistent quality of
the independent variable data. For maximum efficiency, the parameters
should probably be re-estimated based upon all the data, once the model
is considered "verified". This strategy is similar to that of Beck and
Young[3] who used the temporal stability of recursive parameter estimates
as evidence for correct identification of a model for dissolved oxygen
in a river.

1.5 Model Applications

There are a variety of approaches and techniques which can facilitate full and proper use of a model. Three aspects of concern in this work are sensitivity analysis, simulation, and error analysis. These are introduced briefly below.

Potentially useful information about some of the important controlling factors and processes in a system can be derived from sensitivity analysis. Patten[13] has suggested that such an analysis be used as a means of identifying control strategies and has outlined a general approach for application to lakes. Miller[11] has shown that sensitivity analysis can yield useful information, even under conditions of uncertainty in the model or parameter estimates. Sensitivities of model simulations to parameter values can be used as a basis to assess important controlling processes within the system, while sensitivities to independent variables can indicate the relative controlling influences of external factors, some of which may be manageable, possibly acting as "decision variables" in a design problem. This type of analysis can also be used as a basis for simplification of the model. A useful way of expressing sensitivities are as normalized first partial derivatives of the model dependent variables with respect to the factor being studied:

$$ \underline{S}_j = w_j \frac{\partial \underline{Y}}{\partial w_j} \qquad (1.5\text{-}1) $$

1.5 Model Applications (continued)

where,

w_j = factor j (parameter or independent variable)

$\underline{S}_j$ = vector of sensitivities of state variables with respect to factor j

Multiplication by w_j essentially corrects for differences in factor scales and permits direct comparisons among different factors. The coefficients can be determined analytically or by using finite-difference techniques. In a dynamic model, the sensitivity coefficients are time-variable.

Control strategies can be viewed as modifications to the vector of independent variables. Most often, many of the independent variables are not influenced by management strategies. Meteorologic regimes are usually not controllable, but can have significant effects upon the systems being studied. For this reason, control strategies should be evaluated in the context of the variability induced by fluctuations in such factors. In some situations, "critical" conditions can be easily defined and employed in simulating management strategies. A classic example of this is evaluating the effects of biochemical oxygen demand discharges on dissolved oxygen in a river under low-flow conditions. In others, "critical" conditions may not be as evident, and simulations of historical records under the conditions imposed by each management

1.5 Model Applications (continued)

strategy can be used as a basis for evaluating and comparing alternative schemes. If sufficient historical records are not available, a synthetic record might be generated with time series models identified and estimated from data obtained from the geographical region being studied or from one similar to it. The range of system variability induced by fluctuations in the uncontrollable factors can be compared with the range induced by alternative management strategies as a means of estimating the potential detectability of changes induced by such strategies.

Error analysis can be a useful and revealing aspect of model application. Generally, covariance in model predictions can be attributed to three error components:[2]

$$
\begin{array}{ccccc}
\text{Total} & = & \text{Parameter} & + & \text{Independent} \\
\text{Error} & & \text{Error} & & \text{Variable} \\
& & & & \text{Error}
\end{array} + \begin{array}{c} \text{Residual} \\ \text{Error} \end{array}
$$

$$
\underline{V}_{\hat{y}} \approx \left(\frac{\partial \hat{y}}{\partial \underline{p}}\right) \underline{V}_{p} \left(\frac{\partial \hat{y}}{\partial \underline{p}}\right)^{T} + \left(\frac{\partial \hat{y}}{\partial \underline{x}}\right) \underline{V}_{x} \left(\frac{\partial \hat{y}}{\partial \underline{x}}\right)^{T} + \underline{V}_{r} \tag{1.5-2}
$$

In order to evaluate the total error, we need to estimate three covariance matrices: $\underline{V}_{p}$, $\underline{V}_{x}$, $\underline{V}_{r}$. The last can be estimated by comparing model predictions with system observations and is usually

1.5 Model Applications (continued)

obtained in the process of estimating the model parameters. The parts
of $\underline{V}_p$ corresponding to those parameters which have been estimated
directly from system observations can be estimated according to equation
1.3-12. The covariances of the remaining parameters and $\underline{V}_x$ must be
estimated independently, often with only very approximate results.
The required sensitivity vectors can be obtained from differentiation
of the model equations, again using analytic derivatives or finite-
difference methods. The error should be evaluated for each case or
management strategy studied. Although the individual covariance matrices
may be invariant, the sensitivities and total prediction error may change
from one case to the next. For a dynamic model, the total error and the
first two terms in equation (2) are also time-variable.

The residual error component includes both measurement and model
errors:

$$\begin{matrix}\text{Residual} \\ \text{Error}\end{matrix} = \begin{matrix}\text{Measurement} \\ \text{Error}\end{matrix} + \begin{matrix}\text{Model} \\ \text{Error}\end{matrix}$$

$$\underline{V}_r = \underline{V}_m + \underline{V}_e \qquad (1.5\text{-}3)$$

These two components usually cannot be distinguished without some
independent evidence. For instance, measurement error could be

1.5 Model Applications (continued)

estimated from replicate observations.

If the model and measurement error terms can be separated, a
comparison of the parameter error term with the model error term can
provide a basis for deciding whether future efforts should be directed
at improving the model or at obtaining additional data. The covariances
of the parameters can usually be reduced by gathering additional data
for model calibration, whereas changes in the model would generally be
required in order to reduce the model error component. In the interest
of minimizing total prediction variance, future efforts would be directed
toward data acquisition or model improvement according to whether the
parameter error term is greater or less than the model error term, res-
pectively. The relative sensitivities of these terms to investments in
reducing them might also be considered, as discussed below.

In cases where parameter or independent variable errors are signi-
ficant, Thomas[15] has suggested that an error analysis could also be used
as a basis for guiding data acquisition and research efforts. For the
one-variable case, and combining the parameter and independent variable
vectors into a general vector $\underline{w}$, the variance attributed to uncertainty
in the elements of $\underline{w}$ is given by:

$$\sigma_{\hat{y}}^2 \simeq (\frac{\partial \hat{y}}{\partial \underline{w}}) \; \underline{V}_w \, (\frac{\partial \hat{y}}{\partial \underline{w}})^T \qquad (1.5\text{-}4)$$

1.5 Model Applications (continued)

If the off-diagonal elements of $\underline{V}_w$ are assumed to be zero:

$$\sigma_{\hat{y}}^2 = \sum_{i=1}^{n_w} \left(\frac{\partial \hat{y}}{\partial w_i}\right)^2 \sigma_{w_i}^2 \qquad (1.5\text{-}5)$$

Given the objective of minimizing the total prediction variance, one strategy would be to invest additional resources into reducing the variance of the component contributing most to the total. According to this scheme, the various terms of equation (5) would be compared and the factor corresponding to the largest term would serve as a focus for additional monitoring and/or research efforts. Another strategy would consider cost-effectiveness by taking into account the fact that the variances of the parameters or independent variables may have different sensitivities to investment, i.e., some eetimates might be more difficult to improve than others. According to this scheme, Thomas suggested that the variables could be ranked according to:

$$\left(\frac{\partial \hat{y}}{\partial w_i}\right)^2 \left(\frac{\sigma_{w_i}}{\sigma_{\hat{y}}}\right) \left(\frac{\partial \sigma_{w_i}}{\partial c_i}\right) \qquad (1.5\text{-}6)$$

1.5 Model Application (continued)

where,

$$c_i = \text{investment in further data collection to improve the estimate of } w_i$$

Harrington[8] has reviewed applications of these and similar strategies to the design of ambient monitoring programs. In a similar vein, Moore[12] has proposed the use of Kalman Filtering algorithms to permit estimation of system states based both upon model simulations and upon field measurements. Using a minimum variance criterion, he has shown how this scheme can lead to rational designs for aquatic ecosystem monitoring programs. Strategies such as these can be used to guide research and/or monitoring programs and represent potentially fruitful areas of model application.

Another benefit derived from an error analysis is the capability of placing confidence limits on model predictions. Estimates of the first and second moments of predictions can be used, for example, to calculate the probability that implementation of a given management strategy will result in satisfaction of a system standard or quality criterion. Alternatively, management strategies can be designed to satisfy a given probability of achieving system standards. A probabilistic projection, as compared with a deterministic one, provides the policy-maker with a better decision basis and the analyst with an escape route.

1.6 Introduction to Illustrations in Subsequent Chapters

Chapter 2 introduces the general problem of cultural eutrophication
in lakes and reviews some of the roles of monitoring data and models in
this application. Using data from a cross-section of lakes, some
empirical approaches are demonstrated in the development, estimation,
and evaluation of models for prediction of phosphorus concentration and
lake trophic state as functions of phosphorus input rates, and hydrologic
and morphometric characteristics. Techniques employed include nonlinear
parameter estimation and discriminant analysis. Error analyses are
employed to (1) propose a scheme for rational allocation of sampling
effort in a monitoring program designed to gather data for lake nutrient
budget estimation; (2) assess the relative sizes of independent variable,
parameter, model, and measurement errors involved in prediction of lake
phosphorus and chlorophyll concentrations; (3) develop a means of incor-
porating the effects of uncertainty in decision variables on rational
design values for lake nutrient loading allocation. It concludes with
a discussion of model selection criteria. The study is cross-sectional
in nature and provides a basis for a more detailed, longitudinal case
study of Onondaga Lake, New York, in Chapters 3 and 4.

Chapter 3 introduces Onondaga Lake: its environment, historical
problems, and proposed solutions. The bulk of the chapter describes
and explores the existing data base on the lake and its tributaries,
one of the most extensive data bases of its type known to this author.
Techniques for reduction and display of the data are employed to reveal

1.6 Introduction to Illustrations in Subsequent Chapters (continued)

variations and associations in time and space. A method for examining
time series behavior is explored. Techniques for estimating and
displaying mass balances on a continuous basis are developed and demon-
strated. While attempts are made to mechanistically interpret some of
the observed relationships, the formal use of models is not stressed.
A summation of proposed management strategies for the lake is used in
combination with the results of data analyses to define the modelling
problem addressed in Chapter 4.

In Chapter 4, a model for vertical stratification in Onondaga Lake
is developed, estimated, and applied. The objective is to evaluate the
potential impact of the design and location of a waste outfall upon
general features of lake mixing. In the context of this problem, the
use of nonlinear programming algorithms for estimation of parameters
in dynamic models is demonstrated and evaluated. Other techniques
employed include sensitivity analysis, simulation, and error analysis.
The chapter concludes with some general comments on the adequacies of
the model, the data, and the techniques employed to address the defined
problem, as well as some recommendations for outfall design.

REFERENCES - CHAPTER 1

1. Bard, Y., "Comparison of Gradient Methods for the Solution of Nonlinear Parameter Estimation Problems", *Society of Industrial and Applied Mathematics, Journal of Numerical Analysis*, Vol. 7, No. 1, March 1970.

2. Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.

3. Beck, B. and P. Young, "Systematic Identification of DO-BOD Model Structure", *Journal of the Environmental Engineering Division, American Society of Civil Engineers*, Vol. 102, No. EE5, pp. 909-927, October 1976.

4. Blalock, H.M., "Causal Inferences in Non-Experimental Research", University of North Carolina Press, Chapel Hill, North Carolina, 1961.

5. Box, G.E.P. and G.M. Jenkins, *Time Series Analysis, Forecasting, and Control*, Holden-Day, San Francisco, 1970.

6. DeNeufville, R. and J.H. Stafford, *Systems Analysis for Engineers and Managers*, McGraw-Hill, New York, 1971.

7. Draper, N.R. and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1966.

8. Harrington, J.J., "Monitoring Program Design - A Systems Approach", presented at "A Seminar and Workshop on Water Monitoring Technology", University of Delaware, May 1976.

9. Hornberger, G.M., M.G. Kelly, and T.C. Lederman, "Evaluating a Mathematical Model for Predicting Lake Eutrophication", Virginia Water Resources Research Center, Virginia Polytechnic Institute and State University, Bulletin 82, November 1975.

10. Johnston, P.R. and D.H. Pilgrim, "Parameter Optimization for Watershed Models", *Water Resources Research*, Vol. 12, No. 3, pp. 477-486, June 1976.

11. Miller, D.R., "An Experiment in Sensitivity Analysis on an Uncertain Model", *Simulation*, Vol. 23, No. 3, September 1974.

12. Moore, S.F., "Estimation Theory Applications to Design of Water Quality Monitoring Programs", *Journal of the Hydraulics Division, American Society of Civil Engineers*, Vol. 99, No. H45, pp. 815-831, 1973.

REFERENCES - CHAPTER 1 (continued)

13. Patten, B.C., "The Need for an Ecosystem Perspective in Eutrophication Modelling", in Middlebrooks, E.J., D.H. Falkenborg, and T.E. Maloney, Eds., Modeling the Eutrophication Process, Ann Arbor Science, Ann Arbor, Michigan, 1974.

14. Radha Krishnan, D.P., J.J. Lizcano, L.E. Erickson, and L.T. Fan, "Evaluation of Methods for Estimating Stream Water Quality Parameters in a Transient Model from Stochastic Data", Water Resources Bulletin, American Water Works Association, Vol. 10, No. 5, pp. 899-913, October 1974.

15. Thomas, H.A., Jr., "Operations Research in Disposal of Liquid Radioactive Wastes in Streams", Harvard Water Resources Group, December 1965.

16. Yih, S. and B. Davidson, "Identification in Nonlinear, Distributed Parameter Water Quality Models", Water Resources Research, Vol. 11, No. 5, pp. 693-704, October 1975.

2.0 METHODS FOR ASSESSMENT OF LAKE WATER QUALITY PROBLEMS

2.1 Introduction

A lake can be viewed as a system which responds in various ways to culturally or naturally induced changes in its boundary conditions. Assessment of lake water quality problems is a good context in which to illustrate some of the analytic techniques and strategies discussed in Chapter 1. In the following chapter, the roles of data and models in managing lake water quality are discussed with an emphasis on the general problem of cultural eutrophication. Introductory sections review aspects of lake classification, nutrient balance estimation, and lake modelling. In Section 2.5, data from over 100 northern-temperate lakes are used to evaluate and compare some of the simpler lake models which have been proposed. Nonlinear regression and discriminant analyses are employed to develop empirical models for phosphorus retention and trophic state projection. Strategies for application of these models are discussed with a particular emphasis on parameter estimation, error analysis, and data requirements. The chapter concludes with a brief overview and discussion of model selection criteria in Section 2.6. Besides providing a context in which to illustrate some of the various techniques discussed in Chapter 1, this chapter serves as a partial basis for a more detailed case study of Onondaga Lake in Chapters 3 and 4.

2.2 Lake Classification

Lakes have been classified according to various physical, chemical, and biological characteristics. Since lakes exist on a continuum, classification into discrete states is rather artificial, but serves as a basis for organization and comparison. While classifications are defined on arbitrary scales, some may be more theoretically-based than others. Schemes for use in water quality management should relate to established criteria, which, in turn, should reflect potential for beneficial use and, possibly, some fundamental measures of ecosystem health.

An important aspect to consider is that lakes are generally not static entities. The term "eutrophication" is generally used to describe the natural evolution of lakes from deep, nutrient-poor systems with low biological productivity to shallow, nutrient-rich systems with high productivity, and, eventually, into swamps and meadows.[38] The phrase "cultural eutrophication" more accurately reflects water quality management concerns because it describes the acceleration of this natural process as a result of cultural interventions. This acceleration is generally considered to be a result of enhanced flow of nutrients into the lake ecosystem.

Traditionally, biological productivity has served as a primary basis for classification of lakes into three basic types: "oligotrophic," "eutrophic," and "dystrophic." The history and basis of this classification scheme have been reviewed by Welch[98]

2.2 Lake Classification (continued)

and are summarized in Table 2.2-1. Oligotrophic lakes are relatively
low in productivity, organic matter and nutrient content, and rela-
tively high in hypolimnetic dissolved oxygen and transparency, while
eutrophic lakes have opposite characteristics. The term "mesotrophic"
refers to an intermediate state between the above. Dystrophic lakes
are distinguished by their high humic acid content, which renders
a characteristic yellow or brown color. Organic matter in these
lakes is generally allochthonous (originating in the watershed),
whereas, organic matter in oligotrophic and eutrophic lakes is
autochthonous (produced within the lake). The effect of light
limitation limits the submerged, biological productivity of dystrophic
lakes to relatively low levels, although oxygen levels are similar
to those found in eutrophic lakes, due to decay of organic matter
entering the lake from external sources. The natural succession
of oligotrophic lakes is toward the eutrophic state, that of
eutrophic lakes, toward swamps, and that of dystrophic lakes, toward
peat bogs. While there are some quantitative aspects to this tradi-
tional classification scheme, it is primarily descriptive.

A number of investigators[73,75,88,92] have suggested some more
quantitative classification schemes in attempts to remove some of
the subjectivity from the traditional method while retaining
essentially the same interpretation. These schemes, reviewed by
Uttormark and Wall[92], have been based upon such indices as minimum
hypolimnio dissolved oxygen, transparency, chlorophyll-a, available

Table 2.2-1

Principal Characteristics of Oligotrophic, Eutrophic, and Dystrophic Lakes According to Welch[98]

Characteristics	Oligotrophic type	Eutrophic type	Dystrophic type
General form relations...............	Deep; shoals narrow or absent; volume of hypolimnion greater than epilimnion	Shallow; shoals broad; volume of hypolimnion smaller than epilimnion	Deep to shallow; in bog surroundings or in old mountains
Color of water.....................	Blue to green; characteristic color always predominant	Green to yellow and brownish green; characteristic color often concealed by vegetation color	Yellow to brown; characteristic color always predominant
Transparency.....................	High	Lower; eventually very low	Low
Chemical content of water..........	Electrolytes variable; relatively poor in N and P; humic materials absent; Ca variable	Electrolytes variable; rich in N and P; humic materials slight; Ca abundant, seldom scanty	Electrolytes low; poor in N and P; humic materials abundant; Ca scanty
Suspended detritus.................	Minimal quantities	Large quantities	Humic materials abundant
Deep deposits.....................	Poor in organic materials; not putrescent	Rich in autochthonous,[*] putrescent organic materials	Poor in autochthonous, humic materials; rich in allochthonous[*] materials
Dissolved oxygen content: a. In summer...............	Oxygen content relatively uniform from surface to bottom; an increase in thermocline; abundant in hypolimnion; deep water with 60-70 % saturation; few or no putrefactive processes in bottom deposits	In deep lakes of this type, oxygen content in thermocline suddenly and strongly increased; little or no oxygen in hypolimnion; oxygen decline of deep waters 0-40% saturation, seldom above 40 %; vigorous putrefaction in bottom deposits	As in eutrophic type
b. In winter under ice.............	As in summer	In deep lakes as in oligotrophic type; in shallow lakes, oxygen may fall to 0 % saturation at bottom	As in eutrophic type
Cause of oxygen decline.............	....	In summer, due to plankton and bottom muds; in winter, due to bottom muds	Due to allochthonous suspended and deposited humic deposits; combination with Fe
Littoral plants.....................	Scanty	Abundant	Scanty
Plankton..........................	Quantitatively poor; present up to large depths; diurnal vertical migration wide in range. Water bloom rare; Chlorophyceae and to some extent the diatoms dominate over the Schizophyceae; true hypolimnion phytoplankton absent	Quantitatively rich; diurnal vertical migration restricted to small distance. Water bloom abundant; Schizophyceae and diatoms together with Chrysomonadales and the Peridiniaceae predominating over the Chlorophyceae; true hypolimnion phytoplankton abundant (±)	Phytoplankton quantitatively scanty. Water bloom absent or very rare; Schizophyceae and diatoms, Chlorophyceae, Chrysomonadales, Peridiniaceae, and desmids present; true hypolimnion phytoplankton abundant (±).
Distinction between littoral and profundal zone.	Indistinct; indicated only by termination of vegetation	Sharply defined; due chiefly to change of dissolved oxygen relation	As in eutrophic type
Profundal fauna: a. Qualitative...............	Rich in species; requiring high oxygen content; Tanytarsus fauna; Corethra never present	Poor in species; with wide toleration of oxygen-content changes; Chironomus fauna; Corethra almost always present	Still poorer in species; wide toleration of oxygen-content changes; Chironomus fauna; bottom deposits eventually devoid of fauna; Corethra present (occasionally only Corethra)
b. Quantitative.................	Relatively rich; in general, less than 1,000 but above 300 animals per sq. m. = 1-4 g. dry weight	Rich; 2,000-10,000 animals per sq. m. = 20-100 g. dry weight	Poor; at most, 10-20 animals per sq. m.; bottom layer often none
c. Distribution with depth........	Uniform	Not uniform	
Qualitative correlation between plankton and the profundal fauna.	Present	Absent	Bottom fauna always poor; phytoplankton poor; zooplankton often rich
Deep-dwelling coregonine fishes......	Abundant	Present only in exceptional instances	Always absent in advanced dystrophic lakes
Succession.........................	Into eutrophic type	Into pond, swamp, meadow moor	Into peat bog

[*] Limnologische Terminologie Abderhalden's Handbuch der biologischen Arbeitsmethoden, by permission of Urban and Schwarzenberg, Berlin.
[*] Autochthonous: produced within the same basin; allochthonous: originating elsewhere.

2.2 Lake Classification (continued)

and total nitrogen and phosphorus concentrations, and alkalinity. For collections of lakes, these indices are often highly correlated with one another, refleoting their common association with the eutrophication process. Because of the partioular effects of light limitation, dystrophic lakes are often excluded or treated separately.[73] Likewise, lakes in which primary production is dominated by aquatic macrophytes generally do not fit into classification schemes developed on lakes dominated by phytoplanktonic production.[73,88] Some schemes have been based upon rankings and have been used primarily for comparisons within a given group of lakes.[88] Other investigators have attempted to derive objective scales for more general application.[73,75,88,92] Both types of schemes, however, depend directly upon the particular lake data set employed as a basis and are accordingly restrictive in scope of application.

Shannon and Brezonik[73,74] have led the work on subtropical lakes with a unique, multivariate - statistical approach. Their data base consisted of 55 North and Central Florida lakes. Cluster analysis of water quality data was used to classify the lakes into distinot groups, which roughly corresponded to the classical trophic states. The first principal component of the correlation matrix of traditionally-employed trophic state indicator variables was used to derive a formula for a so-called "trophic state index" (TSI):

2.2 Lake Classification (continued)

$$TSI = 5.19 + 0.919 \frac{1}{SD} + 0.800 \ COND + 0.896 \ TON + 0.738 \ TP \qquad (2.2\text{-}1)$$
$$+ 0.942 \ PP + 0.862 \ CHA + 0.634 \frac{1}{CR} \ .$$

where

			Mean	Standard Deviation
$\frac{1}{SD}$	=	Inverse Transparency $(m)^{-1}$	0.84	0.77
COND	=	Specific Conductance (micromhos/cm)	93.1	101.3
TON	=	Total Organic Nitrogen (mg/l)	1.02	0.82
TP	=	Total Phosphorus (mg/l)	0.125	0.177
PP	=	Planktonic Primary Production $(mg\text{-}C/m^3\text{-}hr)$	44.8	82.3
CHA	=	Chlorophyll-a (mg/m^3)	16.9	19.8
$\frac{1}{CR}$	=	Inverse Cation Ratio	1.47	1.63
	=	$[(Na + K)/(Ca + Mg)]^{-1}$		

The coefficients in equation (1) are appropriate for use with standardized values of the respective variables. This expression explained about 70% of the variance in the correlation matrix of the indicator variables. Oligotrophic lakes generally had TSI

2.2 Lake Classification (continued)

values for less than 3, mesotrophic lakes, between 3 and 7, and
eutrophic lakes, greater than 7.

The TSI offers some distinct advantages over traditional
classification schemes. It provides a more objectives means of
ranking lakes according to water quality using a continuous scale
which is both more sensitive and more realistic than the traditional,
discrete trophic states. The coefficients of variation of five of
the seven indicator variables in equation (1) were greater than one,
however. This reflects considerable skewness in the distributions
of these variables and suggests that a transformation of the
variables may have been appropriate prior to the principal component
analysis. By promoting normality in the distributions of the variables,
transformations would have rendered the analysis more in agreement
with the principal conditions for efficiency of the multivariate
techniques employed.

As a second phase of their work, Shannon and Brezonik demon-
strated correlations between TSI values and land use patterns,
population densities, and nitrogen and phosphorus loadings. Multiple
regressions of TSI on land use and population density factors explained
81.5% of the TSI variance, while regressions of TSI on estimated
phosphorus and nitrogen loadings explained up to 67%. These kinds
of relationships were suggested as means of evaluating watershed
management strategies for the control of eutrophiciation.

2.2 Lake Classification (continued)

In view of the results of Vollenweider[95] and Dillon[16], the results might have been revealing if the effects of lake morphometric and hydrologic properties had been taken into account. The significance of the regressions of TSI on nutrient loadings is questionable, since nutrient concentrations contributed directly to the calculated TSI values. In the interest of distinguishing between causes and effects, one might argue against the inclusion of nutrient concentrations in the list of trophic indicators. It would seem more logical to allow the TSI to depend only upon variables which reflect observed ecosystem response or which relate to the potential for beneficial use. Nutrients are generally viewed more as causes than effects and should therefore be excluded from the TSI, according to this rationale.

Despite these criticisms, the general approach of Shannon and Brezonik provides some potentially valuable management tools for application in regional planning. TSI values provide an objective means of ranking lakes with regard to observed water quality. The statistical relationships between TSI values and watershed characteristics provide means of evaluating management strategies and of suggesting feasible subjects for water quality restoration and protection programs. However, the approach is highly empirical and at most as good as the data used to develop the ranking schemes and relationships. Potential effects of data include the effects of original lake selection, measurement errors, and assumptions

2.2 Lake Classification (continued)

and techniques employed in reducing the data to useful form.
More confidence could be placed in results of this type if
essentially the same results could be derived from an independent,
randomly-selected set of lakes and watersheds in the same geo-
graphical region. In addition, the generality of their model
could perhaps be improved by the incorporation of more theoretical
factors in the analysis, such as lake flushing rate and depth.
Perhaps the major contribution of Shannon and Brezonik was the
demonstration that multivariate techniques can be applied
successfully to problems of this type in order to reduce dimen-
sionality and thereby provide a concise and useful summary of
the data.

Uttormark and Wall[92] developed a system for classifying lakes
based upon subjective information. Noting that data were generally
lacking to permit the application of a classification scheme based
directly upon water quality measurements, they devised the "Lake
Condition Index" (LCI), which was based upon the point system
outlined in Table 2.2-2. In classifying 1100 Wisconsin lakes,
Uttormark and Wall constructed a questionnaire which facilitated
estimation of LCI values by regional agency personnel who were
generally familiar with the lakes in question. In comparing the
LCI values with the traditional trophic state rankings for those
lakes for which data were available, oligotrophic lakes were
generally found to have LCI values of 4 or less, mesotrophic lakes,

2.2 Lake Classification (continued)

between 5 and 9, and eutrophic lakes, between 10 and 23. Comparing
multiple responses for the same lakes, Uttormark and Wall found
that estimated LCI values were reproducible to within ± 2 units
for 89% of the lakes tested. The source of this variation was
found to lie primarily in the use impairment component, which,
in turn, correlated with the particular professional concerns
or recreational interests of the individuals or agencies being
polled. Based upon a national survey, they concluded that the
LCI technique could be used to correctly classify 70-80% of the
lakes in the U.S. with surface areas greater than 40 hectares.

Table 2.2-2

Point System for Uttormark and Wall's Lake Condition Index[92]

Component	Range[a]
Dissolved Oxygen	0-6
Transparency	0-4
Fish Kills	0,4
Use Impairment	0-9
TOTAL	0-23

[a] Lower scores are more desirable.

2.2 Lake Classification (continued)

As a second phase of their study, Uttormark and Wall suggested
how LCI values might be related to nutrient loading and watershed
development factors to demonstrate that the classification scheme
could serve as a workable data base for lake renewal and management
programs. They suggested that a normalized nutrient loading
parameter could be estimated for each lake based upon measured
or estimated loadings and upon lake morphometric properties. The
normalized loading would be an indicator of eutrophication potential.
For instance, according to Vollenweider's first model* (Figure 2.4-1),
the normalized loading would be given by $L/z^{.6}$, or, according to
his second model* (Figure 2.4-2), it would be given by $L/QS^{.5}$. In
the absence of measured loading data, land use factors, such as
urban drainage area/lake surface area could be employed as surrogates.
A plot of the normalized loading against LCI would, in effect, depict
potential problems versus observed problems.

A major contribution of Uttormark and Wall was that such a
plot (Figure 2.2-1) could be used as a device for "dynamic" classi-
fication of lakes. Their basic assumption was that deviations from
predicted conditions were due to non-steady-state conditions.
Accordingly, four regions could be identified in Figure 2.2-1.
Lakes in Region I, lower left, would have low observed and low
potential eutrophication problems and would be considered to be in
no immediate danger. Convereely, lakes in Region III, upper right,

---

* These models will be discusssd in detail in Section 2.4.

Figure 2.2-1

Dynamic Lake Classification According to Uttormark and Wall[92]

Lake Condition Index
(Observed Water Quality Problems)
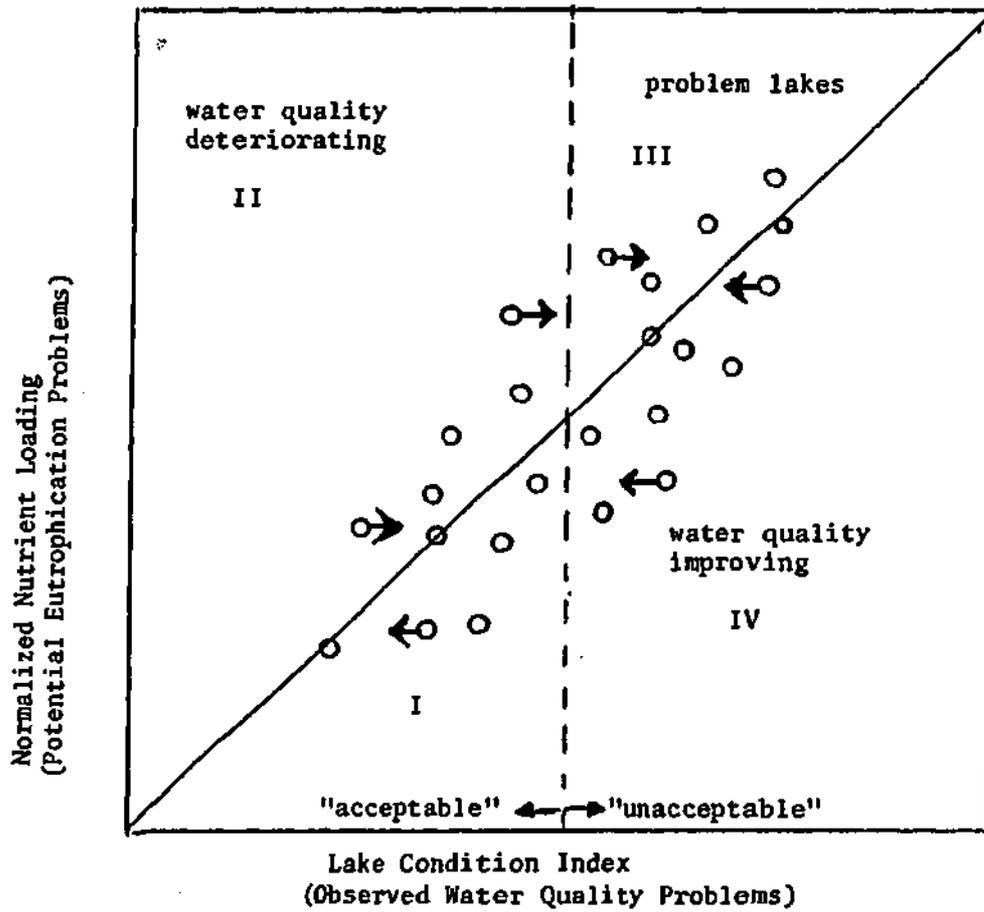
2.2 Lake Classification (continued)

would be problem lakes, in which extensive nutrient abatement may
be necessary in order to effect lasting improvements. Lakes in
Region II, upper left, would be considered high priorities for
protective actions, since their location on the plot suggests that
currently "acceptable" water quality may be deteriorating. Con-
versely, lake in Region IV, lower right, would be considered likely
candidates for restoration programs, since long-term benefits may
be possible without extensive nutrient source abatement.

This scheme for dynamic classification of lakes has some
obvious applications in regional management of water quality.
The basic assumption is that deviations from the perfect fit
of potential versus observed eutrophication problems are due to
non-steady-state conditions. Of course, data errors would
contribute substantially to such deviations. Model error would
also be another source of deviations, since all of the factors
contributing to variations in the capacities of lakes to handle
nutrient influx without adverse effects would not be accounted
for in a simple model of the Vollenweider type. With these
qualifications in mind, Uttormark and Wall's dynamic classification
scheme appears to be promising as a screening method in designing
regional water quality management programs. Its potential applica-
tions are not restricted to cases in which the LCI is used as a
measure of observed lake water quality.

2.3 Nutrient Balance Rationale

The nutrient balance is an essential element in any analysis
of lake water quality problems from a eutrophication viewpoint.
Regardless of the complexity of the particular lake model being
employed, long-term predictions of water quality can be heavily
dependent upon accurate estimation of the boundary conditions
imposed by the lake's nutrient inputs, particularly if projections
are being made in the nutrient-limited state of the system. Under
such conditions, it would be impossible to calibrate or verify a
given lake model unless the nutrient inputs were estimated accurately.
By quantifying the relative contributions of various nutrient sources
in a lake's drainage basin, the balance also serves to identify
problem origins and potential control points.

In formulating a nutrient balance, an estimate must be obtained
for each potential source in the watershed, averaged over the time
scale of interest. The latter is determined by the requirements of
the particular lake model being employed. Nutrient sources can be
grouped initially into general categories according to the types of
mechanisms involved, including hydrologic, cultural, meteorologic,
geologic, and biologic processes[52]. Hydrologic fluxes enter or
leave the lake in surface- or ground-water flows. Cultural fluxes
represent point discharges of municipal, industrial, or agricultural
wastes. Meteorologic fluxes are mediated by atmospheric phenomena.
Inputs include dustfall and precipitation; outputs, entrainment
of nutrients in spray from the water surface. Geologic
fluxes represent transfers to or from the lake's sediment, as

2.3 Nutrient Balance Rationale (continued)

governed by water and sediment chemistry. Biologic fluxes include
contributions from migrant waterfowl, insect emergence, plant or
animal harvesting, nitrogen fixation, denitrification, and sediment
exchanges.

Because the geologic and many of the biologic fluxes involving
the sediment are difficult to estimate or measure directly, the
lake's sedimant is usually included within the control boundaries of
the system. The accumulation term represents the net sum of all
fluxes not estimated directly. Hydrologic, cultural, and, to some
extent, meteorogic fluxes are relatively easy to measure and generally
represent the primary terms in the balance.

2.3.1 Factors Influencing Nutrient Sources

Point sources of nutrients are influenced by a number of factors, including tributary population, types and amounts of industrial or commercial activity, and types and degrees of waste treatment. The kinds of materials being cycled through the urban and industrial systems are also of importance (e.g., detergents). Point sources generally exhibit marked diurnal, weekly, and seasonal periodicities which reflect variations in cultural activities[58,82]. In contrast to nonpoint sources, they are relatively easy to measure and trace.

The flux of nutrients from a watershed can be viewed as leakage from the resident terrestrial ecosystems. Likens and Borman[52] have discussed the various processes which are responsible for the input, cycling, and loss of nutrients in terrestrial systems. Generally, cultural interventions tend to open the relatively closed nutrient cycles of natural terrestrial environments. This results in an increased flux of nutrients from the watershed and in an increased potential for water quality problems. The concept that management of water quality entails management of the terrestrial environments has become the fundamental theme of programs for controlling nonpoint sources of water pollution[2,52].

Some nutrients are more tightly conserved than others in terrestrial systems[52]. Biological uptake acts to conserve

2.3.1 Factors Influencing Nutrient Sources (continued)

nutrients by converting soluble forms to forms which are less labile in the hydrologic cycle. Soil chemistry is an important regulating factor. Erosion is a principal export mechanism for nutrients which can be strongly adsorbed to soil particles and are therefore filtered from percolating waters (e.g., phosphorus or ammonia). The degree of adsorption depends in part upon soil pH and clay composition[48]. In contrast, nitrate nitrogen generally does not adsorb and is readily leached from the soil in surface runoff and groundwaters [8,51]. The reduction in the potential for contact between surface waters and soils is considered to significantly increase nutrient mobility and export from urban areas [31].

The work of Likens et al.[8,34,51] at Hubbard Brook in New Hampshire has demonstrated the relatively closed nature of the nutrient cycles in an undisturbed forest. The measured export rate of phosphorus from one portion of the watershed was 21 g/ha-yr , compared with an input of 108 g/ha-yr in precipitation and an annual internal cycling of 1900 g-P/ha-yr in leaf fall alone.[8] Thus, only a relatively small fraction of the phosphorus cycled within the watershed escaped in the hydrologic outflows and the system appeared to be accumulating phosphorus at the rate of 87 g/ha-yr. A similar area which had been clear-cut had an average export rate of 203 g-P/ha-yr, an increase of about 10-fold over the undisturbed system. Most of this increase

2.3.1 Factors Influencing Nutrient Sources (continued)

was attributed to the particulate fraction (> 1 mm); the
corresponding increase in the dissolved and fine-particulate
phosphorus export after clear-cutting was about two-fold.
The export of nitrate nitrogen increased about 50-fold and
remained at that level for at least two years after clear-cutting.

Increases in nutrient export following agricultural or
urban development have also been well-documented[2,19,52,53,89,91]
Table 2.3-1 presents the results of an extensive literature
review conducted by Uttormark et al.[91] to summarize the available
data on the relationships between land use and nutrient export.
The wide distribution of export rates within each land use
classification can be attributed to variations in such factors
as land use intensity[52], geology[19,89], soil type[19,89,93], and
drainage basin morphology[43,89]. In addition, the types of
fertilizers, the timing and techniques of fertilizer application,
soil structure, geography, and animal population characteristics
are considered to be critical factors in regulating nutrient
losses from agricultural watersheds[32,53]. Population densities,
street sanitation practices, extents of paved surfaces, and
sewerage system characteristics regulate losses from urban
watersheds[31,97]. Errors in the mean flux estimates could also
contribute substantially to the variability of export rates
within each land use classification. Such errors would be
characteristic of estimates derived from sampling programs of

Table 2.3-1

Typical Values of Nutrient Runoff Coefficients

for Various Land Uses According to Uttormark et al. .91

Land use	NO$_3$-N+NH$_4$-N kg/ha/yr			Total-N kg/ha/yr		
	High	Low	Ave	High	Low	Ave
Urban	5.0	1.0	2.0	10.0	2.5	6.0
Forests	3.0	0.5	1.6	5.0	1.0	2.5
Agricultural	10.0	1.0	5.0	10.0	2.0	5.0

	Diss.inorg-P kg/ha/yr			Total-P kg/ha/yr		
	High	Low	Ave	High	Low	Ave
Urban	2.0	0.5	1.0	5.0	1.0	1.5
Forests	0.1	0.01	0.05	0.8	0.05	0.2
Agricultural	0.5	0.05	0.1	1.0	0.1	0.3

2.3.1 Factors Influencing Nutrient Sources (continued)

insufficient frequency or length to provide an adequate basis for calculation of a long-term-average flux. Hydrologic variations may also be important, since, as discussed below, concentration tends to vary less than flux in streams not dominated by point sources.

The factors discussed above are responsible for variations in long-term-average nutrient export rates from watershed to watershed. Temporal variations for a given watershed, as induced by climatologic changes on various time scales, are also of concern. An understanding of such variations is critical to the interpretation of measurements of export rates obtained over a particular time period. Because different types of terrestrial ecosystems respond differently to climatologic variations, one would expect that both the mean and the variability of the nutrient export rate would depend upon land use distributions and upon various geologic and morphologic characteristics of the watershed.

The effect of variation in streamflow on nutrient concentrations is of primary concern. Flow may influence concentration via several mechanisms[9,39,41,62], including: (a) energy effects (streambed or land surface scouring during periods of high flows, heavy precipitation, and/or surface runoff); (b) residence time effects (higher flows permitting less time for entrapment and

2.3.1 <u>Factors Influencing Nutrient Sources</u> (continued)

utilization of nutrients in upstream terrestrial or aquatic
ecosystems); (c) dilution effects (dilution of upstream
point-source discharges or groundwater base flows during high
flow periods). The interaction of all these mechanisms determines
the net response of nutrient concentrations in a given stream to
changes in flow. The responses of dissolved, suspended, and
particulate fractions are often quite different, reflecting the
relative importance of various export mechanisms.

Generally, rivers which are not dominated by upstream
point-source discharges tend to show less variation in total
nutrient concentration than in flux (concentration × flow). This
is true both in response to short- and long-term variations in
hydrologic regime. For example, in developing a 6-year nutrient
budget for Shagawa Lake, Mauleg et al.[58] monitored nutrient
concentrations at weekly intervals on the Burntside River, a
major tributary with a heavily forested watershed. The ranges
of the reported yearly-average flows, phosphorus fluxes, and flow-
weighted mean phosphorus concentrations were $34.6 - 83.5 \times 10^6$ m^3/yr,
$429 - 973$ Kg/yr, and $11.7 - 14.9$ mg/m^3, respectively. The corres-
ponding coefficients of variation were 0.30, 0.26, and 0.09,
indicating that year-to-year variations in concentration were
much less significant than variations in flow or mass flux.
No consistent relationship between yearly average concentration
and flow was evident. Similarly, results from Hubbard Brook[39,51]

2.3.1 Factors Influencing Nutrient Sources (continued)

and from the Burntside River[58] indicate a relatively small response of total phosphorus or total nitrogen concentration to seasonal flow variations in forested watersheds.

Less extensive data exist to determine whether other types of watersheds respond similarly to changes in hydrologic regime. Kilkus et al.[44] found that ortho-phosphorus and ammonia nitrogen concentrations correlated slightly positively with flow in some primarily agricultural watershed in Central Iowa. They suggested that this could be a result of an increased proportion of relatively high-concentration runoff during high flow periods. Cahill et al.[9] found a negative correlation between total and ortho-phosphorus concentrations and streamflow during steady-stage conditions on the Brandywine River, which drains a highly diverse watershed in southeastern Pennsylvania and Delaware. They suggested that dilution of upstream point sources during higher flow periods was primarily responsible for the observed behavior. Likewise, Wang and Evans[96] found an inverse relationship between ortho-phosphorus concentration and flow on the Illinois River. In a study of flow and concentration data from various water quality sampling stations in the U.S., Enviro Control[27] observed dilution effects in 8 stations out of 26 for ortho-phosphorus, but in only one station out of 42 for total phosphorus. The National Eutrophication Survey analyzed concentration and flow data from over 200 watersheds and concluded

2.3.1   Factors Influencing Nutrient Sources (continued)

that, on the average, total phosphorus and total nitrogen

concentrations varied with the -0.11 and -0.6 power of flow,

respectively.

The response of stream concentrations to storm events

is also of interest. Observed increases in suspended nutrient

concentrations during rising river stages have been attributed

to the scouring of land surfaces and river bottoms with increasing

flows [9,27,41]. In the Enviro Control study mentioned above [27],

total phosphorus was found to exhibit this behavior in 21 out

of 41 stations, whereas ortho-phosphorus exhibited it in 3 out

of 26. If the scouring of streambeds is an adequate explanation,

the response of a stream to a given storm would depend not only

upon the spatial and temporal rainfall intensity pattern, but

also upon the antecedent dry-weather period, as the latter would

determine the amount of material stored in the stream bed at the

beginning of the storm. The consequences of this phenomenon on

the design of tributary sampling programs will be discussed in a

subsequent section.

2.3.2 Estimation Methods

Strategies for estimating the terms of a nutrient balance can be classified as direct or indirect. Direct evaluation entails the integration of flow and concentration measurements obtained at or near the point of discharge of each source into the lake. Indirect evaluation is based upon established relationships between fluxes or concentrations and the causally-related factors discussed in the previous section.

The current state-of-the-art is such that direct estimation methods are generally preferable. Indirect methods are most useful in extrapolating estimates derived from direct measurements. For example, the nutrient fluxes from a given watershed might be estimated from measurements on an adjacent watershed with similar land use, geologic, and morphometric characteristics, assuming a drainage area proportionality or equivalent concentrations[62,86]. Nutrient fluxes from sewage treatment plants can be estimated from population and type of treatment. Per capita estimation of nutrient sources from shoreline residences with septic systems has been commonly employed[86]. The latter estimates are suspect due to the complexities of the factors determining the performance of such disposal systems. Unfortunately, direct measurement of contributions from shoreline septic systems is an also difficult task[28], and this can contribute substantial uncertainty to the total nutrient budgets of some lakes[69].

2.3.2 Estimation Methods (continued)

Comprehensive models for indirect estimation of nonpoint sources do not exist as yet. The relatively wide variations of nutrient export rates for various land use categories (Table 2.3-1) suggest that other factors will have to be included before sufficient accuracy and reliability can be achieved[53]. The development of these models is essential to the understanding and eventual control of nonpoint sources and thus to lake water quality management in general.

Direct evaluation of nutrient fluxes presents some particular sampling and estimation problems. In order to obtain an average flux estimate over a time scale Δt , the following integral must be evaluated:

$$\overline{W}_{\Delta t} = \frac{1}{\Delta t} \int_0^{\Delta t} q_t\, c_t\, dt \qquad (2.3-1)$$

where

q_t = instantaneous flow (vol/time)

c_t = instantaneous concentration (mass/vol)

$\overline{W}_{\Delta t}$ = average flux over time Δt (mass/time)

2.3.2   Estimation Methods (continued)

Typically, continuous measurements of flow are available.   If
continuous, flow-weighted composite sampling is employed to
determine concentration, the integral can be evaluated directly
and the only errors in the estimate are the usual ones associated
with the measurement and analytical process.   If only grab sample
concentration data are available, however, assumptions must be
made about the behavior of concentration between sampling times.
This requires some assumptions which introduce estimation errors,
as well as measurement errors.

The various methods of computing average fluxes from
continuous flow and grab-sample concentration data differ in
their assumptions regarding the behavior of concentration between
sampling times.   The characteristics of the stream, in particular,
the relationship between concentration and flow, determine which
method is appropriate.   Generally, calculation methods which assume
a constant flux are appropriate for point-source dominated systems,
while methods which assume constant concentration are more appro-
priate for use in systems dominated by nonpoint sources.   These
concepts are discussed in greater detail below.

The appropriate calculation technique also depends upon the
time scale of interest.   For instance, if a continuous flux estimate
is desired, the data can be treated as a time series and interpola-
tion techniques can be employed to estimate fluxes between sampling

2.3.2 Estimation Methods (continued)

times[69]. Interpolation can be done on the flux versus time
axes if flux is found to be less variable than concentration,
or on the concentration versus volume (cumulative flow) axes
if the opposite is true. If the time scale of interest is long
enough compared with the sampling frequency, the data can be
aggregated and treated as samples from a stationary population.
With appropriate adjustment in the equivalent sample size to account
for any serial correlation in the observations[4] , this permits
estimation of both a mean and its variance. The latter is useful
in establishing confidence limits and in identifying relative
needs for additional stream sampling.

In a study for the Army Corps of Engineers, Meta Systems[62]
developed nutrient budgets on various reaches of two major river
systems in north-central Florida: the Oklawaha and the Withlacoochee.
For each reach, total nitrogen, total phosphorus, and BOD_5 balances
were developed on monthly, yearly, and eight-year-average time
scales. Point sources of nutrients were generally unimportant in
these rivers. In this effort, this author investigated a number of
methods for estimating average fluxes from continuous flow and
grab-sample concentration data. For each station, nutrient, and
time scale, both the mean and variance of the flux estimate were
derived. The latter permitted evaluation of the statistical
significance of the nutrient accumulation terms in various reaches,
relative to the errors inherent in the input and output estimates.

2.3.2 <u>Estimation Methods</u> (continued)

Monthly balances were based upon regression models which related concentration to flow, season, and time (trend). In these models, it was necessary to incorporate interactions between season and·flow, since the slopes of the concentration versus flow regression lines were found·to vary significantly with season at many of the stations. These models explained between 23% and 70% of the variance in the log-transformed concentration data at various stations. They were used in combination with the continuous flow records to generate flux estimates and mass balances at monthly intervals.

To estimate yearly- and long-term-average fluxes, five methods were investigated and compared, based upon the computed variances of the respective estimates. The formulas employed are given in Table 2.3-2. Method 1 involved the direct averaging of the products of sampled concentrations and corresponding flows. According to Method 2, the concentration and flow data were averaged independently. Method 3 was based upon the flow-weighted average concentration and the average flow. Methods 4 and 5 were based upon regression models which related sampled concentration to a sampled flow. The model formulations were concentration versus inverse flow and log (concentration) versus log (flow), respectively.

On both time scales, the variances of the estimates derived

Table 2.3-2

Methods of Estimating Nutrient Fluxes

Investigated in the Cross-Florida Barge

Canal Study[62]

**Techniques of Estimating Nutrient Loadings from Grab Sample
Concentration Data and Continuous Flow Measurements**

<u>Definitions:</u>

c_{s_i} = grab sample concentration on sampling day i

q_{s_i} = flow corresponding to sample i

N_s = number of samples

$\bar{q}$ = average flow in period of interest

L_j = average estimated loading (q x c) using Method j

$V(L_j)$ = variance of L_j

$\bar{c}$ = average concentration

$\bar{c}'$ = flow-weighted average concentration

a,b = regression parameters

e_i = error term

S_e = standard error of estimate = standard deviation of e.

<u>Method 1: Average Loading</u>

$$L_1 = \frac{1}{N_s} \sum_{i=1}^{N_s} c_{s_i} q_{s_i}$$

$$V(L_1) = \frac{\sum_{i=1}^{N_s} \left(q_{s_i} c_{s_i} - L_1\right)^2}{N_s(N_s-1)}$$

Method 2: Average Concentration

$$L_2 = \bar{q} \, \bar{c} = \bar{q} \frac{1}{N_s} \sum_{i=1}^{N_s} c_{s_i}$$

$$V(L_2) = \frac{\sum_{i=1}^{N_s} q_{s_i} (c_{s_i} - \bar{c})^2}{N_s (N_s - 1)}$$

Method 3: Flow-Weighted Average Concentration

$$L_3 = \bar{q} \, \bar{c}' = \bar{q} \frac{\sum_{i=1}^{N_s} q_{s_i} c_{s_i}}{\sum_{i=1}^{N_s} q_{s_i}}$$

$$V(L_3) = \frac{\sum_{i=1}^{N_s} q_{s_i}^2 (c_{s_i} - \bar{c}')^2}{N_s (N_s - 1)}$$

Method 4: Regression Model A

$$c_{s_i} = a + \frac{b}{q_{s_i}} + e_i$$

$$L_4 = a \bar{q} + b$$

$$V(L_4) = \bar{q}^2 s_e^2 \left[\frac{1}{N_s} + \frac{\left(\frac{1}{\bar{q}} - \left(\frac{1}{N_s} \sum_{i=1}^{N_s} \frac{1}{q_{s_i}} \right) \right)^2}{\sum_{i=1}^{N_s} \left(\frac{1}{q_{s_i}} - \frac{1}{\bar{q}} \right)^2} \right]$$

Method 5: Regression Model B

$$\ln c_{s_i} = a + b \ln q_{s_i} + e_i$$

define,

$$\bar{x}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} \ln q_{s_i}$$

$$s_x^2 = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(\ln q_{s_i} - \bar{x}_s \right)^2$$

q_{s_j} = mean monthly flow for month j

N_s = total number of months in averaging period

$$L_5 = \frac{1}{N_s} \sum_{j=1}^{N_s} a \, q_{s_j}^{b+1} \left(1 + \frac{1}{2} \left(s_e^2 \left(1 + \frac{1}{N_s} + \frac{(\ln q_{s_j} - \bar{x}_s)^2}{N_s s_x^2} \right) \right) \right)$$

$$V(L_5) = L_5^2 \left(\frac{1}{N_s} + \frac{(\ln \bar{q} - \bar{x}_s)^2}{N_s s_x^2} \right) s_e^2$$

Table 2.3-2 (continued)

2.3.2 Estimation Methods (continued)

from Methods 1 and 2 were generally higher than those derived
using Methods 3, 4, or 5. Besides giving estimates with relatively
high variance, it can be shown that Method 2, in which flow and
concentration are averaged independently, gives biased flux
estimates if concentration is not independent of flow, whereas
the other methods all give asymptotically unbiased estimates. In
deriving long-term (eight-year) flux estimates, no general dis-
tinction could be made among the last three methods. Methods 3,
4, and 5 gave somewhat sharper estimates of phosphorus, BOD_5, and
nitrogen fluxes, respectively. In estimating yearly-average fluxes,
Method 3 was found to develop an advantage over Methods 4 and 5,
in that it required estimation of only one, as opposed to two
parameters.

Based upon considerations of minimum bias, variance, and
computation effort, the results of the Cross-Florida Barge Canal
Study[62] indicated that Method 3 was preferable overall. For large
sample sizes, the regression methods offered an advantage only if
the relationships between concentration and flow were significant.
Method 3 involved multiplying the flow-weighted average sample
concentration times the average flow over the entire period of
concern. This amounts to a "ratio estimate" of the mean and
assumes that, on the average, flux is proportional to flow.
Snedecor and Cochran[77] note that this type of estimate works best
in systems where the variance of the dependent variable (flux)

2.3.2 Estimation Methods (continued)

increases with the magnitude of the independent variable (flow).
They also note that the ratio estimate (Method 3) is superior
to direct averaging (Method 1) only if:

$$\rho_{XY} > \frac{C_X}{2C_Y} \qquad (2.3\text{-}2)$$

where

ρ_{XY} = correlation coefficient between independent variable X and dependent variable Y

C_X, C_Y = coefficients of variation of X and Y , respectively

In this case, the above criterion will be approximately
satisfied if the slope of a log (concentration) versus a log
(flow) regression is greated than -0.5. As dilution effects
become important, the slope would approach -1 and direct
averaging (Method 1) would become preferable. For systems
not dominated by point sources, however, concentration would
be expected to be a weak function of flow*, and the ratio
estimate would be generally preferable.

*See Section 2.3.1.

2.3.2   Estimation Methods (continued)

In computing the variances of the flux estimates in the Florida study, independence of the samples was assumed. Serial correlation in the observations would tend to decrease the effective sample sizes for calculation of the mean and variance[4]. As will be discussed in the next section, such effects would be expected to be small at the sampling frequences of less than one per month, typical of the data used in the Florida study. As sampling frequency increases, however, serial correlation may distort the mean and variance estimates derived directly from the equations in Table 2.3-2. Interpolation techniques which treat the data as time series may become superior to the methods discussed above. An approach to evaluating calculation techniques under these conditions is discussed below.

A Monte-Carlo study would be useful as a means of evaluating sampling strategies and calculation methods for estimating the mean and variance of nutrient flux in streams of various characteristics. Stochastic models for streamflow and concentration could be identified and estimated from any high-frequency sampling data available. The models would incorporate trends, seasonal effects, and stochastic variations[37,50,59] Relationships between flow and concentration could be built into the deterministic and/or stochastic elements of these models. They could be used to generate synthetic time series of flow and concentration which could be sampled at various frequencies. Various calculation

2.3.2 <u>Estimation Methods</u> (continued)

methods could be applied to estimate the average fluxes from the
sampled data. This would permit comparison and evaluation of the
methods with regard to bias and variance. The accuracy of the
variance estimates derived from various methods could also be
assessed. Alternative model formulations and/or parameter values
could be used to test the effects of different stream characteristics
on the relative performances of the estimation methods and upon
the sampling frequencies required to provide estimates with given
error bounds.

2.3.3 Monitoring Program Design

In designing monitoring programs to provide basic data for nutrient balance estimation, the particular characteristics and periodicities of the sources must be considered. In the case of cultural streams, concentration or flux data may exhibit cyclical variations at daily or weekly frequencies. Such periodicity has been observed in quantity and quality data from municipal sewage effluents[58,82]. Industrial effluents may be characterized by a high degree of variability. Because of these aspects, continuous, flow-weighted composite sampling has been most-often employed in streams of these types. However, continuous sampling of nutrient concentrations has been done only rarely in tributaries[13]. Accordingly, flux estimates in these cases are most-often derived from grab-sample concentration measurements. If this is the case, grab-sampling should be scheduled at roughly equal volume, rather than equal time increments. This would increase temporal frequency during high runoff periods (e.g. spring).

One problem in this regard is the potential impact of storm events on loading estimates derived from infrequent grab sampling. As discussed previously (Section 2.3.1), total phosphorus concentrations have been observed to increase during rising flows as a result of the scouring of land surfaces and/or bottom deposits[9,27,34,41]. Brief periods of high flows and high concentrations can introduce considerable skewness into

2.3.3 Monitoring Program Design (continued)

the distribution of total mass flux. As a result, average fluxes estimated from dry-weather grab-sampling could be considerably in error. One factor which would tend to reduce the significance of this phenomenon is that it is generally associated with the particulate nutrient fractions[34,41]. Keup[41] has suggested that these materials settle out fairly rapidly in downstream river reaches as storm flows subside or eventually in lakes. The likelihood that they enter into lake nutrient cycles is relatively small. However, periodic grab-sampling should probably be supplemented with some sampling during storm events in order to observe the variations of nutrient concentrations in the dissolved, suspended, and gross particulate fractions. If large increases in the dissolved and suspended fractions are evident over the course of a storm, then periodic grab sampling alone may not provide a sufficient basis for mean flux estimation. The particular storm events sampled should have fairly long antecedent dry-weather periods, so as to allow sufficient time for materials to accumulate in upstream river beds or on land surfaces.

The standard design for tributary sampling programs in lake management schemes has typically been to sample all tributaries at a given frequency, usually monthly or biweekly. Given the objective of obtaining a total nutrient loading estimate which is unbiased and has minimum viariance, it would

2.3.3 Monitoring Program Design (continued)

be of interest to consider whether appreciable benefits would
be derived from optimizing the allocation of sampling effort
among the various streams, subject to a fixed total monitoring
cost. This problem is initially posed as a problem in mathematical
programming and solved in Table 2.3-3. The objective is to obtain
a total loading estimate with minimum variance, subject to a fixed
cost constraint which is initially assumed to be expressed as a
fixed total number of samples, N . The allocation is assumed
to be made among M tributaries, each characterized by a
variability k_i and loading l_i . The solution is given by:

$$\frac{n_i}{N} = \frac{k_i c_i q_i}{\sum\limits_{i=1}^{M} k_i c_i q_i} = \frac{s_i}{\sum\limits_{i=1}^{M} s_i} \qquad (2.3-3)$$

Thus, three factors determine the fraction of the total
sampling effort optimally allocated to stream i : k_i ,
variability; c_i , average concentration; and q_i , average
flow. In designing a sampling program according to this
scheme, the parameters could be estimated a priori based
upon a preliminary survey, land use, and hydrologic data.

The potential benefits of optimizing the sampling
program can be estimated by comparing the variance of the

Table 2.3-3

Optimal Allocation of Samples Among
Tributaries for Nutrient Balance Estimation

Problem Definition:

Select: n_i, $i=1$, M

To Minimize: $\sigma_T^2 = \sum\limits_{i=1}^{M} k_i^2 \ell_i^2 / n_i$

Subject to: $\sum\limits_{i=1}^{M} n_i = N$

Where: n_i = number of samples allocated to stream i

M = total number of streams

σ_T^2 = variance of total flux estimate

N = total number of samples

k_i = variability of stream i

ℓ_i = average flux of stream i

Solution: LaGrange Method of Undetermined Multipliers

LaGrangian = $Y = \sum k_i^2 \ell_i^2 / n_i + \lambda(\sum n_i - N)$

$\dfrac{\partial Y}{\partial n_i} = -\dfrac{k_i^2 \ell_i^2}{n_i^2} + \lambda = 0$, $i=1,M$

$\rightarrow n_i = k_i \ell_i / \sqrt{\lambda}$

$\dfrac{\partial Y}{\partial \lambda} = \sum n_i - N = 0$

$\rightarrow \sum n_i = N = \sum k_i \ell_i / \sqrt{\lambda}$

$\rightarrow \sqrt{\lambda} = \sum k_i \ell_i / N$

$\therefore \dfrac{n_i}{N} = \dfrac{k_i \ell_i}{\sum k_i \ell_i} = \dfrac{k_i q_i c_i}{\sum k_i q_i c_i}$

2.3.3 Monitoring Program Design (continued)

estimate obtained from the allocation according to equation (3)
with the variance of the estimate for the case in which all
streams are sampled equally:

$$\sigma_{TE}^2 = \frac{M}{N} \sum_{i=1}^{N} s_i^2 \qquad (2.3\text{-}4)$$

$$\sigma_{T*}^2 = \frac{1}{N} \left(\sum_{i=1}^{N} s_i \right)^2 \qquad (2.3\text{-}5)$$

$$\frac{\sigma_{T*}^2}{\sigma_{TE}^2} = \frac{\left(\sum s_i \right)^2}{M \sum s_i^2} = \frac{1}{1 + \overline{CV}_s^2} \qquad (2.3\text{-}6)$$

where

σ_{TE}^2 = variance of estimate with equal
sampling allocations

σ_{T*}^2 = variance of estimate with optimal
sampling allocations

$\overline{CV}_s$ = coefficient of variation of s_i,
$i=1,M$

If all s_i are equal, $\overline{CV}_s = 0$, and the optimal and equal
allocations coincide. As the differences in the streams
become more pronounced, reflecting increasing levels of $\overline{CV}_s$,

2.3.3   Monitoring Program Design (continued)

the potential benefits of optimization increase.

The nutrient budgets of five lakes in the National Eutrophication Survey [86] have been examined to determine the possible effects of optimal sample allocations on the variances of the mean loading estimates for some typical tributary groupings. In estimating the optimal allocations, it was assumed that the weighting factors, s_i , were proportional to the total phosphorus loadings, i.e., the variabilities were assumed to be equal. Variance ratios according to equation (6) ranged from 0.15 to 0.88, corresponding to a range in the ratios of confidence limits of 0.39 to 0.94. The most pronounced effect was observed in a lake which had seven influent streams, one of which accounted for 96.7% of the total loading. The least pronounced effect was observed in a lake with two inlet streams at a loading ratio of about two to one. Thus, it appears that in some cases, appreciable benefits could be achieved from optimizing the sampling allocations.

One means of improving the above analysis would be to employ a more realistic cost constraint, which would not be strictly proportional to the total number of samples. One such constraint would be:

2.3.3 <u>Monitoring Program Design</u> (continued)

$$C = a\, n_i^{max} + b\, N \tag{2.3-7}$$

where,

$$C = \text{total cost}$$

$$n_i^{max} = \text{maximum value of } n_i$$

$$a,b = \text{cost parameters}$$

The first term represents the fixed costs of deploying a sampling crew a total of n_i^{max} days during the year of monitoring. The second term reflects the incremental sampling and analytical costs per sample.

A second improvement would be to account for possible effects of serial correlation in the observations on the variance of a mean flux estimate for a given stream. Bayley and Hammersly[4] have shown that serial dependence can be accounted for by using an equivalent sample size, n_i^e, in calculating the variance of the mean estimate:

$$\frac{1}{n_i^e} = \frac{1}{n_i} + \sum_{j=1}^{j=n_i-1} (n_i - j)\, \rho_{jt} \tag{2.3-8}$$

2.3.3   <u>Monitoring Program Design</u> (continued)

where,

$$t \; = \; \text{sample interval} \; = \; 365/n_i \; \text{days}$$

$$\rho_{jt} \; = \; \text{lag} \; jt \; \text{serial correlation coefficient}$$

A number of investigators[37,50,59,60] have shown that high-frequency components of stream water quality data can be modelled reasonably well as a first order, autoregressive (Markov) process with a serial correlation coefficient in the range of 0.7 to 0.9 at one day lag. For such models, Matalas and Langbein[57] have shown that equation (8) reduces to:

$$\frac{1}{n_i^e} \; = \; \frac{1}{n_i} + \frac{2}{n_i^2} \; (\frac{\rho^t \; (\rho^{nt} - n\rho^t + n - 1)}{(\rho^t - 1)^2}) \qquad (2.3-9)$$

where,

$$\rho \; = \; \text{lag 1 day autoregression coefficient}$$

This equation is plotted on log-log scales in Figure 2.3-1 for various values of ρ . At a monthly sampling frequency, the effect is relatively insignificant and $n_i^e \simeq n_i$. As

Figure 2.3-1

Relationship Between Effective and Actual Sample
Sizes in Calculating a Mean Generated by a Lag
One Markov Process
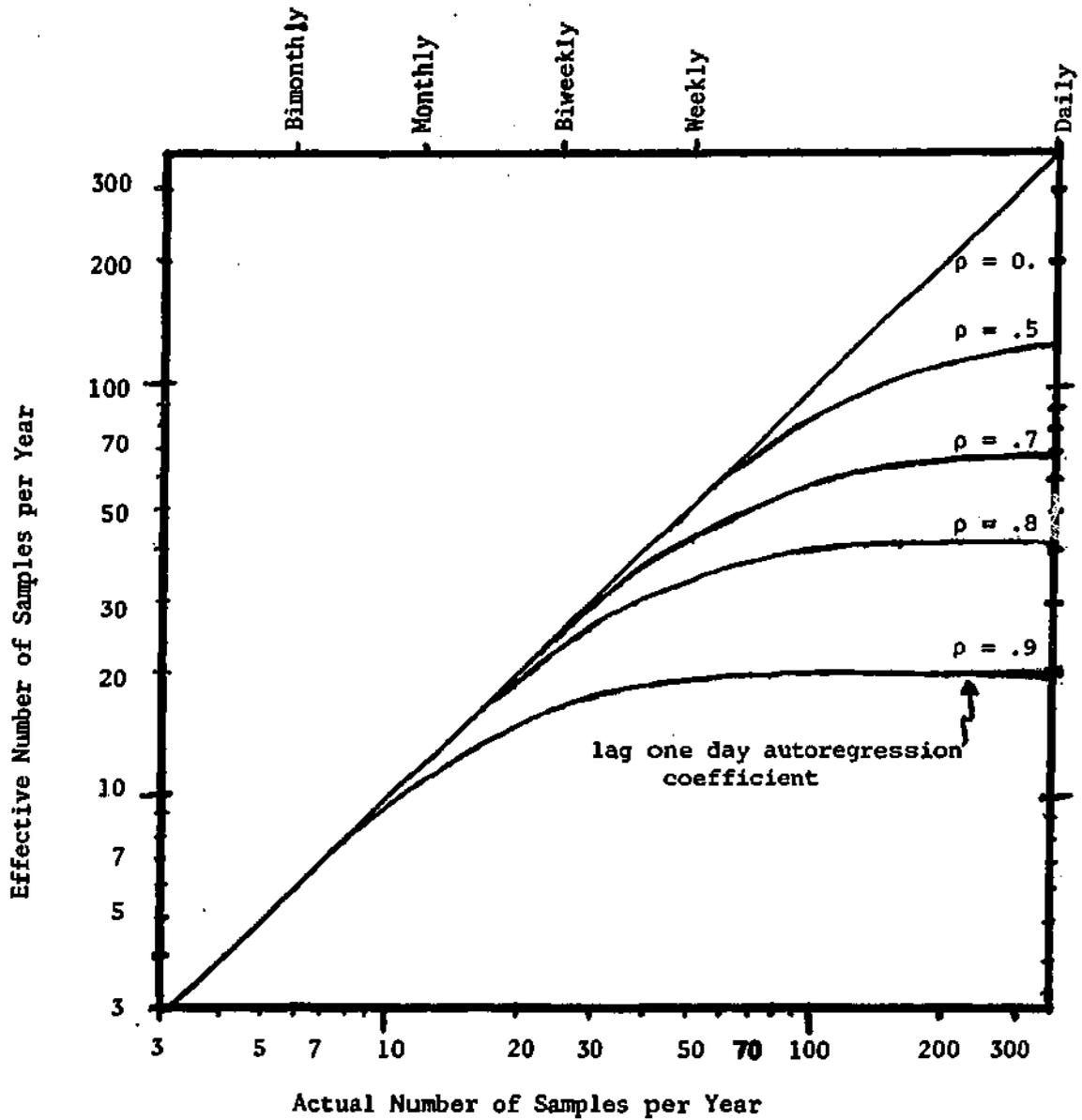
2.3.3 <u>Monitoring Program Design</u> (continued)

sampling frequency increases, the effect becomes more
pronounced. For values of between 0.8 and 0.9, there is
little difference in effective sample size between weekly
and daily sampling frequencies. Thus, if the objective is
to estimate a mean value, a weekly sampling program would
provide about as much useful information as a daily one.
Littenmaier[50] has discussed the implications of this on
monitoring program design and has employed the effective
sample size concept in studying the powers of nonparametric
tests to detect trends in water quality data. Incorporation
of these effects into the mathematical programming model would
tend to make the optimal allocations more equal than indicated
by equation (3).

In real applications, equation (3) might be used as a
rough indication of desirable sampling program design from a
nutrient budget standpoint. Incorporation of a more realistic
cost function and consideration of the effects of serial
correlation on the variance of the mean would tend to suggest
a more even allocation of sampling effort than that indicated
by equation (3). The benefit in terms of increased precision
in the loading estimate would tend to increase as the similari-
ties of the sampled streams decrease. In many cases, equal
allocation of sampling effort would not be far from the optimal
design. Since monitoring programs are usually multi-objective,

2.3.3 **Monitoring Program Design** (continued)

other factors may have to be considered before arriving at the final program design.

Rationally, the size of the sampling effort should depend upon the required accuracy of the nutrient budget estimates. This, in turn, should depend upon the requirements of the particular lake model employed and upon the potential impact of errors in the flux estimates upon the decisions made or management policies selected. Historically, however, little effort has been made to place confidence limits on nutrient budgets. The design of sampling programs has been dictated by "rules of thumb" and budgetary contraints. As the state-of-the-art advances, the increased interest in making probabilistic water quality projections should eventually provoke an awareness of the errors involved in nutrient budget estimates. Systematic approaches to lake monitoring, modelling, and managing efforts will, in turn, involve more rational monitoring program designs.

2.3.4 Interpretation Problems

A lake's nutrient budget which has been estimated based upon monitoring data obtained during a given period may not reflect average conditions. If a lake model is to be used to predict long-term water quality changes resulting from various management strategies, long-term-average nutrient loadings from various sources in the watershed must be estimated. Likewise, if there are lake-specific parameters in a steady-state model, the parameter values would have to be estimated from observations made on the lake or its outflow under steady-state conditions. Fluctuations in cultural, hydrologic, and meteorologic factors can cause year-to-year variations in the nutrient dynamics of watersheds and lakes. These variations introduce uncertainty into the extrapolation of monitored input and output fluxes to long-term-average conditions.

There are no data known to the author which would establish whether significant long-term variations in the export of nutrients from watersheds under culturally-stable conditions exist. The errors inherent in the measurement of these fluxes based upon grab sampling may have obscured our ability to detect such variations, however. The time scale of the natural evolution of terrestrial ecosystems and any resultant changes in nutrient export is probably beyond the planning horizon. Since concentration is generally a weak function of flow in tributaries not dominated by point sources, the flux for an average hydrologic year can be

2.3.4 <u>Interpretation Problems</u> (continued)

approximated as the product of the measured flux and the ratio
of average flow to the flow during the sampled year. This
assumes that the watershed is under fairly stable conditions
and that no significant interventions which would influence
nutrient trapping or release within the watershed occur during
the period of sampling. The National Eutrophication Survey[86]
has used essentially this procedure, with the exception that
the ratio of flows was raised to the .89 power, assuming that
average phosphorus concentration varies as the -.11 power of
flow. This slope was derived from an analysis of flow and
concentration data from over two hundred watersheds with fairly
heterogeneous characteristics. In the case of point sources,
which are not hydrologically-mediated, fluxes are usually assumed
to be independent of flow. If the distinction between point and
nonpoint sources cannot be easily made, the relationship between
concentration and flow can be examined graphically and, if a
significant slope is apparent, a regression model can be employed
to estimate loading during an average hydrologic year [9,33].
Assuming that watershed conditions are relatively stable during
sampling, the adjustment of loading to average conditions can
thus be made relatively easily.

Estimating the output and accumulation terms of a lake's
nutrient budget under average conditions is a much more difficult
task. In order to do this, assumptions must be made about the

2.3.4 Interpretation Problems (continued)

dynamics of the nutrient trapping mechanisms within the lake.
Estimates of the retention coefficient, defined as the fraction
of the influent phosphorus which does not leave in the outflows,
are required for the applications of Dillon's model[16,17], which
will be discussed in the next section. In converting observed
phosphorus and nitrogen outflows to an average hydrologic year,
the National Eutrophication Survey (NES)[86] has employed the
same techniques used in converting the nutrient inflows. The
following analysis suggests that this may have introduced some
significant errors in the accumulation rates, retention coeffi-
cients, and average outlet concentrations reported by the NES.
This is particularly true for the lakes which were sampled in
1972, which was a relatively wet year, characterized by a yearly
flow to average flow ratio of about two, for many of the midwestern
and northeastern lakes sampled. Tropical storm Agnes occurred
in late June of 1972 and could have had a profound effect on the
nutrient and hydrologic budgets of many of the lakes sampled that
year.

Assume that the lake's nutrient budget can be summarized in
the following four terms, which are measured during a given year:

$$L_d = \text{diffuse source loading} \quad (g\text{-}P/m^2\text{-}yr)$$

$$L_p = \text{point source loading} \quad (g\text{-}P/m^2\text{-}yr)$$

$$L_o = \text{outflow} \quad (g\text{-}P/m^2\text{-}yr)$$

$$L_a = \text{accumulation rate} \quad (g\text{-}P/m^2\text{-}yr)$$

2.3.4 <u>Interpretation Problems</u> (continued)

From a steady-state mass balance:

$$L_a = L_d + L_p - L_o \qquad (2.3\text{-}10)$$

In converting these terms to long-term averages, the NES has essentially employed the following conversion scheme:

$$
\begin{aligned}
L_d' &= L_d \, q^{.89} \\
L_p' &= L_p \\
L_o' &= L_o \, q^{.89}
\end{aligned}
\qquad (2.3\text{-}11)
$$

where

q = average-year flow/sampled-year flow.

The corresponding expressions for the sampled and average retention coefficients are given by:

$$R = \frac{L_a}{L_p + L_d} = \frac{L_p + L_d - L_o}{L_p + L_d} \qquad (2.3\text{-}12)$$

2-50

2.3.4 <u>Interpretation Problems</u> (continued)

$$R' = \frac{L_p + L_d \, q^{.89} - L_o \, q^{.89}}{L_p + L_d \, q^{.89}} \qquad (2.3\text{-}13)$$

For $L_p \ll L_d$,

$$R' = \frac{L_d - L_o}{L_d} = R , \quad \text{independent of } q \qquad (2.3\text{-}14)$$

For $L_p \gg L_d$,

$$R' = \frac{L_p - L_o \, q^{.89}}{L_p} \qquad (2.3\text{-}15)$$

The analysis shows that, for lakes dominated by nonpoint sources, the corrected retention coefficient, R' , equals the sampled retention coefficient, and is, therefore, independent of flow. This result is contrary to the empirical equation of Dillon and Kirchner[42], which expresses R as a function of QS , the surface overflow rate (= outflow/ surface area). This equation predicts that R should decrease from one to zero with increasing flows. The above result is

2.3.4 Interpretation Problems (continued)

also contrary to the intuitive argument that higher flows should permit less time for the physical, chemical, and biological reactions responsible for phosphorus trapping to occur within the lakes and should therefore result in lower retention coefficients. For the lakes dominated by point sources, the predicted response of the retention coefficient to changes in flow is correct in sign, but likely to be in error, since no consideration has explicitly been given to the factors controlling nutrient trapping in the lake.

Another difficulty in interpreting lake outflux data is that of assuming steady-state. One would expect the nutrient dynamics to respond faster to changing conditions in low residence time lakes. However, equilibration of lakewater/sediment exchanges may occur on a much longer time scale than indicated by the hydraulic residence time[48,56]. There is a general problem of deciding whether the nutrient outflux is in equilibrium with long-term or short-term hydrologic, and meteorologic, or loading conditions. The net result of these problems is that measurements of nutrient outflows, accumulation rates, and retention coefficients are relatively difficult to interpret in terms of average conditions, unless the measurements have been taken over an extended period of time. Such a period should probably encompass at least three hydraulic residence times, or one year, whichever is greater, and under hydrologic, meteorologic, and nutrient

2.3.4 <u>Interpretation Problems</u> (continued)

loading conditions which are relatively stable and not far from average. One primary consequence is that lake models which rely upon measured lake concentrations or retention coefficient values, such as Dillon's model[17], have more demanding data requirements than those which rely upon estimates of inlet fluxes alone. The data requirements are more demanding both with regard to the number of stations sampled and to the required length of the monitoring period.

2.4 A Review of Lake Models

Models which have been proposed for use in managing lake
water quality can be characterized along various dimensions,
including empirical versus theoretical, deterministic versus
stochastic, stratified versus mixed, and dynamic versus steady-
state. There is a correspondingly complex array of assumptions,
applicabilities, and data requirements. To facilitate discussion
of existing models, they have been separated into two categories:
"empirical" and "theoretical". Models which focus on nutrient
concentrations, relying upon observed correlations between nutrient
concentrations and eutrophic symptoms, have been included in the
"empirical" group. The relatively complex systems models which
attempt to directly predict lake response by simulating the various
physical, chemical, and biological processes have been included in
the "theoretical" group. Since the empirical-theoretical dimension
is continuous, the distinct classification of the models is
somewhat artificial and has been made only as a means of organizing
the discussion. Generally, models of the former group have relatively
low data requirements and are applicable to long-term management
criteria, while models of the latter group require more data and
are applicable to short-term, critical statss, as well as long-term
criteria.

2.4.1 A Review of Empirical Models

The models of Vollenweider[93,95] and Dillon[16-18] are based
upon empirical relationships among trophic states, phosphorus
input rates, and various morphometric or hydrologic variables.
Derived on data from northern temperate lakes, these models
are applicable to steady-state, yearly-average conditions and
focus on phosphorus as the limiting nutrient. They rely heavily
upon correlations between total phosphorus concentrations (mean
annual or at spring overturn) and lake trophic state, or indi-
cators thereof, such as mean midsummer chlorophyll-a concentrations.[14]
These models are characterized by their simplicity and relative
ease of application. They have been posed in graphical form and
require relatively little data to implement.

The areal loading of total phosphorus (g/m^2 lake surface-
year) is a critical variable in each of these models. Accordingly,
the kinds of data and methods employed in the formulation of a
lake's nutrient balance, as discussed previously (Section 2.3),
are particularly important to successful implementation. The
models will be discussed in the chronological order of the
development, which also corresponds roughly to decreasing empiricism
and increasing data requirements.

Vollenweider's first model[93] predicts trophic state as a
function of total phosphorus loading, L (g/m^2-yr) and mean
depth, Z (m) (Figure 2.4-1). In developing this model,

2.4.1   A Review of Empirical Models (continued)

Vollenweider plotted data from 18 lakes on log L versus log Z axes and drew two lines which roughly separated the three trophic states. Eutrophic lakes were generally located above a "dangerous" loading line, and oligotrophic lakes, below a "permissible" loading line. The slopes of these parallel lines are about 0.6, indicating that a doubling in mean depth would permit a 60% increase in loading without a change in trophic state. This is the first, simplest, and perhaps most widely cited of any general eutrophication model.

Vollenweider's second model[95] was developed on an expanded data base of 31 lakes. It predicts trophic state as a function of phosphorus loading and surface overflow rate, the latter given by:

$$QS = \frac{Q}{A} = \frac{Z}{T} \qquad (2.4\text{-}1)$$

where

QS = surface overflow rate (m/yr)

Q = mean annual outflow (m^3/yr)

A = lake surface area (m^2)

Z = mean depth (m)

T = mean hydraulic residence time (yr)
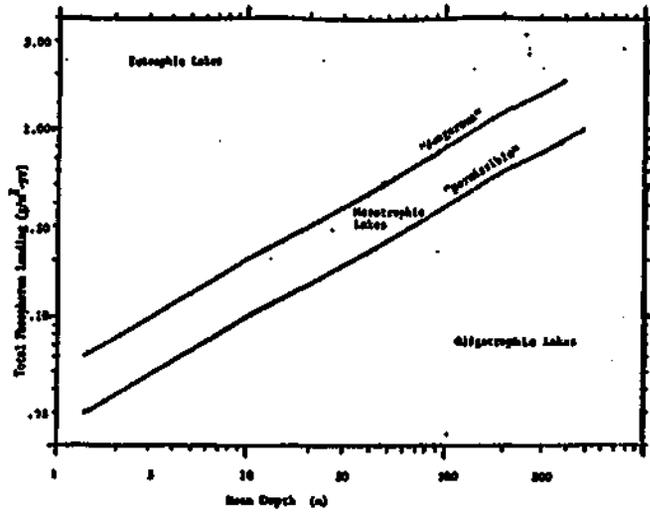
Figure 2.4-1
Vollenweider's First
Model for Predicting
Lake Trophic State
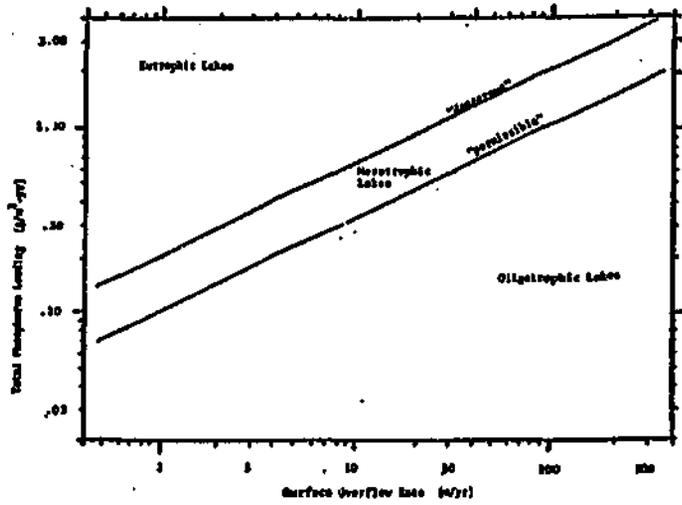
Figure 2.4-2
Vollenweider's Second
Model for Predicting
Lake Trophic State
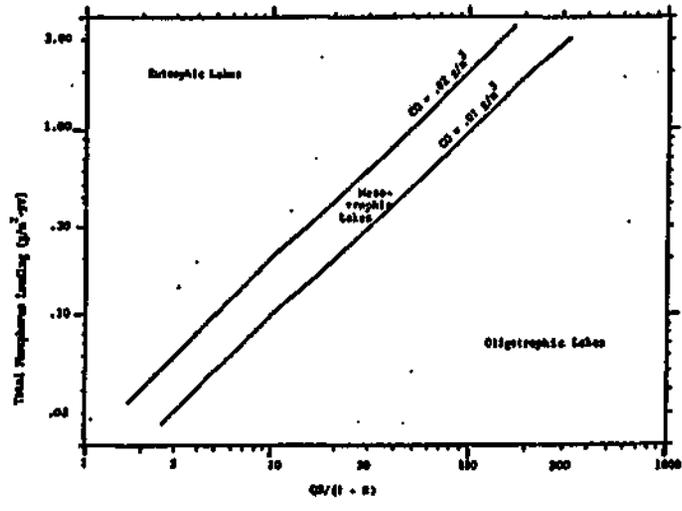
Figure 2.4-3
Dillon's Model for
Predicting Lake
Trophic State

2.4.1 A Review of Empirical Models (continued)

This model, depicted in Figure 2.4-2, incorporates the notion
that lake flushing rate, 1/T , in addition to mean depth,
determine the response of a lake's trophic state to a given
nutrient loading. The model is somewhat more theoretically-based
than Vollenweider's earlier attempt, although the locations of
the lines in Figure 2.4-2 were still empirically determined.
The slopes of these lines are about 0.5, indicating that a
doubling in surface overflow rate would permit a 50% increase
in phosphorus loading without a change in trophic state. This
model has been referred to extensively in the working papers
of the EPA's National Eutrophication Survey[86].

Dillon[16-18] has extended Vollenweider's second model by
coupling it more closely with the phosphorus mass balance.
He defined the retention coefficient, R , as the fraction of
the influent phosphorus which is trapped in lake sediments, or,
equivalently, the fraction which does not leave in the lake's
outflow. According to mass balance equation, the retention
coefficient is given by:

$$R = 1 - \frac{QS \cdot CO}{L} = 1 - \frac{CO}{CI} \qquad (2.4-2)$$

2.4.1 A Review of Empirical Models (continued)

where

> R   =   retention coefficient
>
> CO =   average outlet concentration (g/m^3)
>
> CI =   average inlet concentration (g/m^3)

In Dillon's model (Figure 2.4-3), the lines delimiting trophic state correspond to solutions of the steady-state phosphorus balance equation for mean outlet concentrations of 0.01 and 0.02 g/m^3, respectively. In Figure 2.4-3, Dillon's original axes[18], $L(1-R)T$ versus Z , have been transformed to the equivalent form, L versus $QS/(1-R)$. This transformation isolates the design variable, L , on the y-axis and permits a more direct comparison with Vollenweider's models.

In applying Dillon's model, the retention coefficient must be determined, in addition to L, Z, and T. For a given lake, R can be estimated from input/output measurements, or, if such are not available, from an empirical equation derived by Kirchner and Dillon[42], which allows estimation of R as a function of the surface overflow rate:

$$RD \;=\; 0.426\, e^{-.271\,QS} \;+\; 0.574\, e^{-.00949\,QS} \qquad (2.4\text{-}3)$$

2.4.1   A Review of Empirical Models (continued)

This equation had a correlation coefficient of 0.94 with data
from fifteen oligotrophic and mesotrophic lakes, primarily in
southern Ontario.

In using the model to estimate "dangerous" and "permissible"
loading levels, the assumption is made that   R   (or   RD) is
independent of average inlet or average outlet concentrations.
This is equivalent to assuming that phosphorus decays from the
lake system in a first-order fashion. Such an assumption seems
contrary to the saturation kinetics typically employed in
modelling algal productivity as a function of phosphorus
concentration[47]. These kinetics predict a zero-order dependence
at high phosphorus concentrations, with a half-saturation param-
eter in the range of 0.01 g/m^3. Other, chemical or physical
mechanisms for phosphorus removal (e.g., precipitation, floccu-
lation, and particulate settling) may exhibit first-order
behavior, however. Time series. data on the response of the
retention coefficient to a reduction or increase in phosphorus
loading for a number of lakes would be helpful in clarifying
this issue. Alternatively, if an empirical model for the
retention coefficient (such as equation (3)) could be shown
to have parameter values which are independent of concentration
(or loading) , then the first-order assumption could be applied
with increased confidence in predicting a lake's ultimate
(steady-state) response to a reduction or increase in loading.

2.4.1   A Review of Empirical Models (continued)

Dillon[15] has reviewed early attempts at simple, theoretically
based input/output phosphorus models. The basic model of this
type was initially proposed by Vollenweider[94]. This scheme repre-
sents a lake as a completely mixed reactor from which phosphorus
is removed via a simple, first-order reaction. The steady-state
solution for the average outlet (= average lake) phosphorus
concentration is given by:

$$CO = \frac{L}{QS + \sigma Z} \qquad (2.4-4)$$

where

σ = effective first-order reaction
coefficient for phosphorus removal
(1/yr)

Vollenweider found that σ , the so-called "sedimentation
coefficient", was generally not constant across the limited
number of lakes for which data were available. He subsequently
made a series of modifications to equation (4) to attempt to
account for various inadequacies, in particular the completely-
mixed assumption[15,95].

A host of more theoretically-based models of phosphorus

2.4.1 A Review of Empirical Models (continued)

dynamics in lakes have grown out of the early attempts of
Vollenweider. Complexities of various forms have been
incorporated into equation (4), resulting in multi-box
versions[39,56,63,78] (epilimnion, hypolimnion, and/or sediment),
multi-component versions[63,78] (dissolved P and particulate P),
and non-steady-state versions [56,63,78], the latter intended to
simulate seasonal and/or long-term responses. The common
factor in all of these models is the explicit consideration
of phosphorus or forms thereof. This distinguishes these
models from the more complex dynamic ecosystem models, which
explicitly consider other chemical and biological components
and are often built upon elaborate hydrodynamic models[10,20,21-24,72]
In a management context, however, the more empirical models of
Vollenweider and Dillon have received the most attention and
application.

2.4.2 A Review of Theoretical Models

Relatively complex systems models have also been proposed for use in managing lake water quality. Their theoretical bases, flexibility, and ability to handle spatial and temporal heterogeneities offer potential advantages over the relatively simple models discussed above. In application, they can be used to make predictions which relate directly to short- or long-term water quality management criteria. This permits straightforward comparisons of management strategies[67]. Using simulation and sensitivity analysis techniques, these models can also be employed to investigate the relative importance of various mechanisms in controlling water quality. Such results can be used, in turn, to suggest effective control strategies[65].

The complexities of these models render them relatively difficult to apply. Their data requirements are extensive and lack of sufficient data often severely limits effective and proper use. The elaborate spatial and temporal resolutions of these models are often not warranted in view of the natures of the data bases used for verification, parameter estimation, and for specification of boundary conditions. Typically, large numbers of parameters must be estimated, based upon observations of experimental systems or of the lakes themselves. Because some of the important processes are not well-understood theoretically, numerous structural assumptions are also required. The multi-dimensionality and inter-dependence of the structural

2.4.2 A Review of Theoretical Models (continued)

and parametric assumptions make it difficult to assess potential errors in model projections. Finally, in contrast to some of the simple models discussed in the previous section, specialized professionals and computers are required for successful implementation.

Applications of these types of models can be grouped into two generaly categories: ecologically-oriented studies and water-quality-oriented studies. The former are characterized as relatively fundamental, scientific investigations, whereas the latter pertain more directly to specific lakes and management problems and are more in the realm of engineering. The theoretical components shared by these models have gradually evolved, based upon trial-and-error simulationa and upon independent experimental evidence gathered in laboratcry microcosms or in situ. With regard to autotrophic communities, recent theoretical contributions have been made by Bannister[2] (light effeots) and Grenney et al.[30] (nutrient storage effects). Lassiter[47] has conveniently summarized the functional forms frequently used in representing various chemical and biological processes in aquatic ecosystems. This work is a useful reference in model building, although an analogous compilation of parameter estimates is lacking.

Ecological investigations have focused on the structures

2.4.2   A Review of Theoretical Models (continued)

and interactions of the aquatic food chain[29,49,64,76]. The
general goal of this type of work has been to determine to what
extent the interactions between and within the various trophic
levels can be simulated based upon current understanding of the
individual processes and relationships comprising the system.
In this work, qualitative agreement between observed and
simulated behavior is generally sought, but direct, quantitative
comparisons of observations and predictions are rare. Typically,
the consequences of structural assumptions on the behavior of
the systems of equations have been examined and used to generate
hypothesis concerning important controlling factors [76]. An
increased emphasis on multispecificity in these models has evolved.
Simulating the seasonal succession of algal species as controlled
by physical, predatory, and nutritional factors has been the
subject of works by Lehman et al.[49] and Grenney[29]. "CLEAN" [64]
(Comprehensive Lake Ecosystem Analyzer) is a recent attempt at
an elaborate ecosystem model, having grown out of the efforts of
the International Biological Program.

The coupling of ecosystem models with water quality models
and their application in a management context have grown
principally out of the early work of Chen and Orlob[10] and
Thomann et al.[20]. The latter group has published modelling
studies of the Sacramento River[21], Potomac Estuary[23], and the
Great Lakes[22,24,25,83], among others. The underlying structural

2.4.2 A Review of Theoretical Models (continued)

assumptions of these models have not changed much. Increasing

emphasis has been placed upon an effective display of results

and sensitivity   analysis [24]. The work of Thomann et al. on

Lake Ontario[24] is a good example of the effectiveness of sensi-

tivity analysis in studying the dynamics of these systems. A

diagram of the control pathways typical of models of this sort

is given in Figure 2.4-4.

The evolution from single-specific to multi-specific

representations of the algal community is an important advance

in promoting realism and permitting more direct comparisons of

model predictions with observed seasonal succession in algal

populations [6,46,66,67,70]. In a management view, the capability

of differentiating among the various desirable and undesirable

algal groups would been obvious benefit resulting from the develop-

ment of multi-specific models. Scavia et al.[70] applied a one-

dimensional, single-specific model initially developed and cali-

brated for Lake Ontario to each of the Great Lakes. They found

that it was necessary to "tune" two important parameters in order

to achieve "acceptable" agreement between observed and predicted

chlorophyll and nutrient levels in each of the lakes. One of

the parameters was the half-saturation constant for algal growth

on phosphorus; the other determined the effect of food concen-

tration on zooplankton growth rate. Since the Great Lakes

roughly span the path of a single lake through the trophic

Figure 2.4-4

Biological and Chemical Systems Diagram for

Lake Ontario Model[24]
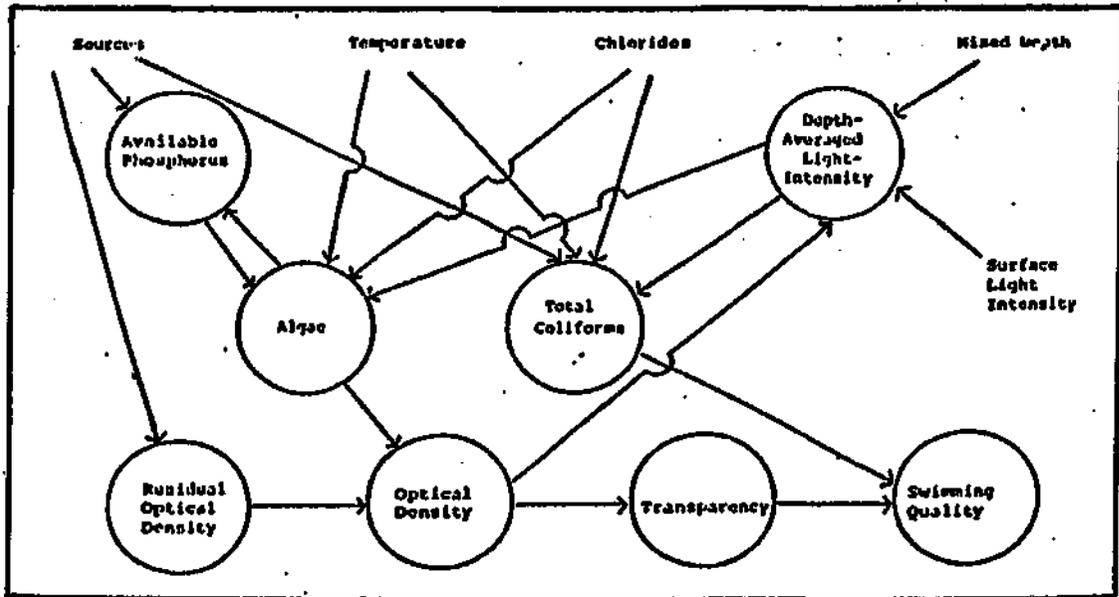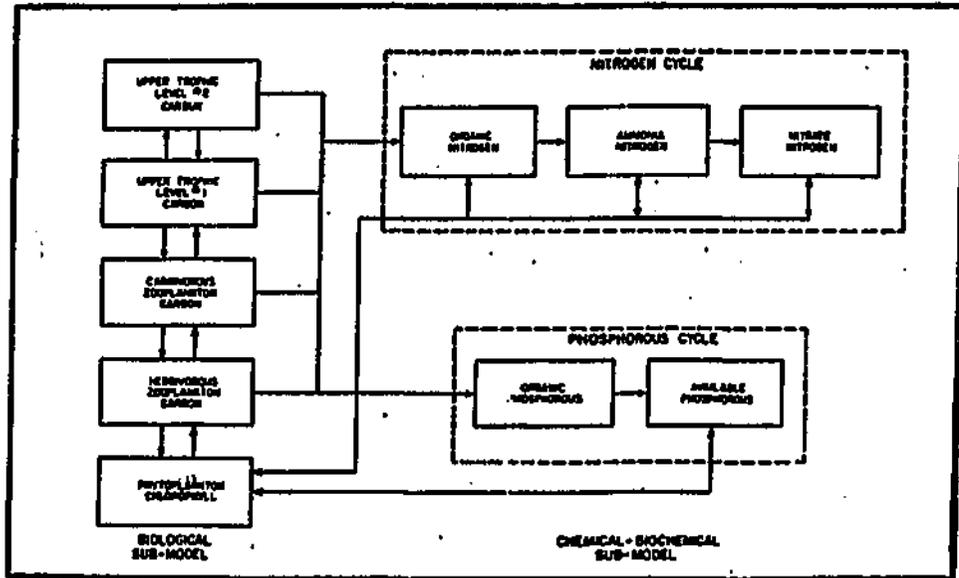
Figure 2.4-5

Control Pathways in the Lower Charles River Model[66,67]

2.4.2 <u>A Review of Theoretical Models</u> (continued)

states in time, they concluded that a single-specific model
with constant parameters would be inadequate to simulate the
eutrophication or recovery of a given lake:

> "For an ecological model to be able to predict
> ecological changes occurring during eutrophication,
> it must include at least several compartments in
> each level of the food chain to allow natural
> selection to be simulated. In this way, "recal-
> ibration" will take place automatically as succession
> and adaptation would in nature."

In this view, multi-specific models will be necessary for
application in predicting long-term responses to changes in
nutrient loading. The resulting enhanced complexities would
present no particular computational problem, but would multiply
parameter estimation difficulties.

This author was involved in monitoring and modelling
studies carried out by Process Research, Inc., on the Charles
River Basin[66,67]. The goal of this work was to develop a
model to predict the impact of various management programs
upon swimming quality in this nine-mile-long impoundment.
A diagram of the control pathways in the model is given in
Figure 2.4-5. An analysis of existing and projected future
conditions indicated that transparency and coliform bacteria
levels were the two factors most severely limiting use of the
basin waters for swimming. Due to the presence of relatively

high levels of background turbidity and color, light was an

important environmental factor, controlling both algal growth

rates and coliform decay rates. An examination of the algal

population data indicated significant shifts in algal species

and densities with river reach and season. These shifts

correlated with surface-water salinity changes induced by

penetration of sea water upstream in the form of a salt wedge.

It was necessary to include three algal types in the model in

order to achieve adequate agreement between simulated and

observed algae and transparency data. This is another instance

in which a single-specie model was found to be an inadequate

representation of the algal population dynamics and the resultant

impact on water quality.

In some applications, the essential aspects of the complex

models can be summarized into relatively simple formulations,

which retain realism, but require substantially less data and

effort to apply. An example of such work is that of Lorenzen

and Mitchell[54,55], who developed a model for predicting the

theoretical effects of artificial destratification on algal

production in lakes and impoundments. In this effort, a

general equation describing the rate of change of algal con-

centration in a lake was simplified in successive stages.

Assuming that the lake was well-mixed above the thermocline,

and that the primary term accounting for algal loss was

2.4.2 A Review of Theoretical Models (continued)

respiration, the growth equation was simplified to:

$$\frac{dC}{dt} = \mu C - rC \qquad (2.4\text{-}5)$$

where

μ = growth rate (1/day)

r = respiration rate (1/day)

C = algal concentration (g/m^3)

For eutrophic lakes, the effect of nutrient concentration on
μ was assumed to be negligible, and, accordingly, μ was con-
sidered to be a function only of available light. Respiration
rate was assumed to be constant. Both μ and r were to be
evaluated at typical, mid-summer epilimnion temperatures.
Employing traditional formulations for the effects of light
intensity on μ and for the effects of algal concentration
on the light extinction coefficient in the water column, Lorenzen
and Mitchell obtained an expression for the peak light-limited
biomass per unit area by setting the integral of equation [5]
over an averaging time period, ΔT , equal to zero and solving
for CZ:

2.4.2 **A Review of Theoretical Models** (continued)

$$C_L^{max} Z = \frac{\mu_{max}}{r\beta} P - \frac{\alpha}{\beta} Z \qquad (2.4\text{-}6)$$

$$P = \frac{1}{\Delta T} \int_0^{\Delta T} \ln(AI_o(t) + (1 + \overline{AI_o(t)}^2)^{\frac{1}{2}}) \, dt \qquad (2.4\text{-}7)$$

$$I_o(t) = I_o^{max} \frac{1}{2} (1 + \cos \frac{2\pi t}{\lambda}), \quad -\lambda/2 < t < \lambda/2 \qquad (2.4\text{-}8)$$
$$= 0, \qquad \text{otherwise}$$

$$\varepsilon = \alpha + \beta C_L^{max} \qquad (2.4\text{-}9)$$

where,

C_L^{max} = light-limited biomass (g/m^3)

Z = mixed depth (m)

μ_{max} = maximum growth rate (1/day)

β = incremental light extinction coefficient due to algae (m^2/g)

α = background light extinction coefficient (m^{-1})

ε = total extinction coefficient (m^{-1})

P = light/depth integral (dimensionless)

λ = day length (hrs)

ΔT = time interval for averaging (≥ 24) (hrs)

I_o^{max} = surface light intensity at noon (lux)

$I_o(t)$ = surface light intensity at time of day t (lux)

t = time of day, plus or minus from solar noon (hrs)

A = parameter describing the effect of light intensity on algal growth rate (lux^{-1})

2.4.2 A Review of Theoretical Models (continued)

It was further assumed that the peak nutrient-limited biomass was given by a concentration $C_N^{max}(g/m^3)$, which depended upon the availability of nutrients in the system. The value of X could be estimated from:

$$C_N^{max} = N_o/y \qquad (2.4-10)$$

where

N_o = limiting nutrient concentration at beginning of summer stratification period (g/m^3)

y = nutrient content of algal biomass (g-nutrient/g-biomass)

This essentially treats the lake as a batch reactor, ignoring any effects of nutrient inflow or outflow between the times of spring overturn and peak biomass.

For some typical parameters values, Lorenzen and Mitchell summarized the results in graphical form by plotting peak nutrient-limited and light-limited biomass against depth (Figure 2.4-6). The nutrient-limited line has a positive slope, while the light-limited line has a negative one. Peak
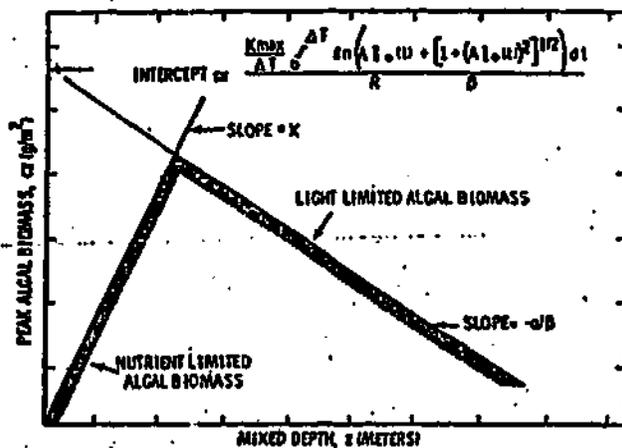
Figure 2.4-6

Effect of Mixed Depth upon Biomass
Potential According to Lorenzen and Mitchell [54]

2.4.2 A Review of Theoretical Models (continued)

biomass would be restricted to the region below both of these
lines. Accordingly, biomass would be nutrient limited if the
mixed depth were less than the Z value corresponding to the
intersection of the two lines, and light-limited, otherwise.

These results were used to explain observed variations
in the response of lakes to artificial mixing, which was
assumed to effectively increase mixed depth. The model
predicted that nutrient-limited lakes would show increases
in peak biomass as a result of mixing, whereas light-limited
lakes would show decreases. This was suggested as a theoretical
basis for estimating whether or not mixing would be beneficial
to a given lake.

A difficulty in applying this model is that associated
with the estimation of the various parameters and in evaluating
the light/depth integral, F , which had to be done numerically.
To simplify the latter, Sykes[81] employed Steele's [79] formulation
for algal growth rate as a function of light intensity:

$$\mu = \mu_{max} \frac{I(z,t)}{I_s} e^{(1 - \frac{I(z,t)}{I_s})}$$ (2.4-11)

where

I_s = saturation light intensity,
characteristic of algae (lux)

2.4.2 <u>A Review of Theoretical Models</u> (continued)

Integration of this expression over mixed depth, Z , and

incorporation into the Lorenzen-Mitchell model gave the

following expression for the light/depth integral:

$$F = \frac{2.718}{\Delta T} \int_0^{\Delta T} (e^{-\frac{I_0(t)}{Is} e^{-\epsilon Z}} - e^{-\frac{I_0}{Is}}) \, dt \qquad (2.4-12)$$

For parameter values over a reasonably wide range, Sykes

demonstrated that the first term in the integral was essentially

equal to one, while the second term equalled one during the

night and averaged near zero during the day. Thus, for a 24-hour

period and daylength λ , the integral can be evaluated as:

$$F = \frac{2.718}{24} [(1-0) \lambda + (1-1)(24-\lambda)]$$

$$(2.4-13)$$

$$= \frac{2.718 \lambda}{24}$$

Sykes neglected to include the daylength factor, $\lambda/24$, in

his result. Despite this oversight, he made a substantial

contribution by greatly simplifying the evaluation of the

light/depth integral term in the Lorenzen-Mitchell model.

2-75

2.4.2 A Review of Theoretical Models (continued)

With this simplification, according to equations (6) and
(13), application of the model can be made with estimates of
the following parameters: μ_{max}/r , α , β , N_o , γ , and
λ . Literature values for μ_{max}/r are generally in the range
of 10 - 20,[22,24] β ranges from 0.2 to 0.4 m^2/g [54], and α from 0.1 to
> 1 m^{-1} [54]. The first two are fundamental characteristics of
the algae and would not be expected to vary much from lake-to-
lake. The last parameter, α , is more lake-specific.

Further simplification of the parameter estimation problems
for a given lake might be facilitated with the use of Secchi disc
observations. A number of investigators[36,84] have shown that the
Secchi depth is inversely related to the extinction coefficient
in the water column:

$$Z_s \, \varepsilon \; = \; k \qquad\qquad (2.4\text{-}14)$$

where

Z_s = Secchi depth (m)

ε = extinction coefficient (m^{-1})

k = dimensionless parameter

2.4.2 A Review of Theoretical Models (continued)

Holmes[36] has shown that the value of k is approximately 1.44. Combining this equation with equation (9):

$$\frac{1}{z_s} = \frac{\alpha + \beta C}{k} \tag{2.4-15}$$

It is assumed that Secchi depth observations are available over the course of a year, and that α , the residual, or non-algal portion of the optical density is independent of season. The maximum Secchi depth would be observed when algal biomass is at an insignificant level. Accordingly:

$$\frac{1}{z_s^{max}} = \frac{\alpha}{k} \tag{2.4-16}$$

$$\alpha = \frac{k}{z_s^{max}} = \frac{1.44}{z_s^{max}} \tag{2.4-17}$$

Likewise, during peak algal biomass periods, the Secchi depth would be at a minimum:

$$\frac{1}{z_s^{min}} = \frac{1}{k} (\alpha + \beta C_L^{max}) = \frac{1}{z_s^{max}} + \frac{\beta C_L^{max}}{k} \tag{2.4-18}$$

2.4.2 A Review of Theoretical Models (continued)

If the peak biomass is not nutrient-limited, equations (6), (13),
and (18) can be combined to give:

$$\frac{1}{z_s^{min}} = \frac{1}{k} \cdot \frac{\mu_{max}}{rZ} \, F \qquad (2.4-19)$$

$$\frac{\mu_{max}}{r} = \frac{Z}{k \, F \, z_s^{min}} = \frac{12.71 \, Z}{\lambda \, z_s^{min}} \qquad (2.4-20)$$

Thus, this scheme permits application of the Lorenzen-Mitchell
model with knowledge of the maximum and minimum Secchi depths,
spring nutrient concentration (N_o) , daylength (λ) , and the
stoichiometric parameters β and y . The first four are
directly observable. The last two are fundamental characteristics
of algae and would not be expected to vary much from lake to lake.
μ_{max}/r is also a fundamental parameter, and agreement between
the predictions of equation (20) and the literature range of 10
to 20 could be viewed as a partial verification of this approach.
For Onondaga Lake, New York, for example, equation (20) has been
evaluated from mixed depth ($Z = 7.2$m) and Secchi observations
($z_s^{min} = 0.5$ m) to yield a value of 12.1 for μ_{max}/r , within
the range of reported values. Similarly, for this lake a value
of 0.58 for α has been estimated from equation (17), based
upon an observed maximum Secchi depth of 2.5 m. Application

2.4.2 <u>A Review of Theoretical Models</u> (continued)

of this scheme to Onondaga Lake will be discussed further in
Chapter 3. The above is intended as a demonstration of how
simplification of the general questions describing algal dynamics
can yield model formulations which have realism and which are
relatively easily applied to address specific problems.

2.5 <u>Analysis of Empirical Approaches</u>

The relatively widespread application of the models of
Vollenweider[93,95] and of Dillon[16,17] for predicting lake trophic
state has resulted primarily from their attractive simplicity,
relatively low data requirements, and initial successes. A
valid criticism of these models is that they have been derived
from generally small and somewhat restrictive data bases. This
has occurred more out of necessity than out of choice, because
of the limited availability of data. The empirical nature of
these models leads to questions about their validity in lake
systems other than those from which they have been derived.
While they are all based upon the sound concept that nutrients,
in particular, phosphorus, partially control rates of eutrophi-
cation, the parameter estimates, i.e. (the slopes and intercepts
of the "permissible" and "dangerous" loading lines) have been
empirically, even subjectively determined and may not be
appropriate for other lakes. Increased availability of data
on lakes from wider geographic, morphometric, and trophic ranges
has facilitated systematic evaluation of these schemes.

2.5.1 Lake Data Base Description

In order to provide a basis for evaluating and comparing these models, data from 105 northern temperate lakes have been compiled (Table 2.5-1). The 1972 sampling of the National Eutrophication Survey[86] has served as the primary data source. The lakes in this group are located primarily in Minnesota, Michigan, Wisconsin, Maine, New Hampshire, Vermont, and New York. These data have been screened by Reckhow[68] to eliminate lakes with acknowledged gross uncertainties in their measured or calculated nutrient or hydrologic budgets. A discussion of the limitations and errors likely to characterize the nutrient budgets estimated for the NES lakes has been presented previously. These data have been supplemented with lake data from the general literature[18,39,78,93]. The only restriction placed on the selection of the data base was that all lakes have mean depths greater than one meter.

Overall, the data set consists of 22 oliogotrophic, 23 mesotrophic, and 60 eutrophic lakes. The classifications have been based upon subjective evaluations of such indicator variables as transparency, chlorophyll, and hypolimnetic dissolved oxygen. Nutrient data should not have contributed directly to these classifications, so that valid evaluations can be made of models predicting trophic state as functions of morphometric, hydrologic, and nutritional factors.

Table 2.5-1

Tabulation of Lake Data

LAKE	S AREA KM2	D AREA KM2	Z M	T YRS	QS M/YR	L G/M2-YR	R —	CI G/M3	CO G/M3	TROPHIC STATE	Reference
SARATOGA	1.00	632.	7.94	0.411	19.295	1.600	0.530	0.0829	0.0390	E	86
SACALANDAGA	7.45	2583.	7.59	0.455	16.700	0.180	0.250	0.0108	0.0081	N	
KEUKA	2.89	424.	22.60	7.813	2.893	0.045	0.556	0.0156	0.0069	N	
LOWER ST REGIS	0.11	53.	5.10	0.301	16.932	0.400	0.111	0.0236	0.0210	E	
CROSS	0.54	8025.	5.50	0.019	286.458	33.500	0.226	0.1169	0.0905	E	
CONESUS	0.79	168.	8.90	1.701	5.233	0.380	0.560	0.0726	0.0320	E	
CAYUGA	10.50	1862.	54.50	11.236	4.850	0.540	0.570	0.1113	0.0479	M	
CARRY FALLS	1.59	2262.	5.40	0.104	51.323	0.720	0.280	0.0139	0.0100	M	
CARMONS FERRY	1.19	1157.	19.20	0.575	33.409	4.260	0.710	0.1275	0.0370	E	
CANANDAIGUA	2.62	434.	39.00	14.925	2.613	0.120	0.720	0.0459	0.0129	O	
RANGELEY	24.28	256.	14.33	2.800	5.118	0.090	0.440	0.0176	0.0098	O	
LONG	24.28	234.	13.41	3.200	4.191	0.120	0.560	0.0286	0.0120	O	
MATTAWAMKEAG	13.35	813.	3.66	0.115	31.826	0.600	0.320	0.0189	0.0128	N	
CLYDE	0.57	363.	3.35	0.010	335.000	8.320	0.150	0.0248	0.0211	N	
LAMOILLE	0.62	694.	1.68	0.003	560.000	25.210	0.200	0.0450	0.0360	E	
ARROWHEAD MT	3.34	1816.	3.14	0.010	314.000	11.260	0.270	0.0359	0.0262	E	
WATERBURY	3.60	290.	12.68	0.227	55.859	1.340	0.240	0.0240	0.0182	N	
WINNEPESAUKEE	180.44	940.	13.11	4.000	3.277	0.110	0.730	0.0336	0.0091	O	
ALTOONA	3.40	2100.	2.13	0.014	152.143	19.910	-0.080	0.1309	0.1413	E	
EAU CLAIRE	4.52	1541.	2.26	0.027	83.704	9.040	-0.020	0.1089	0.1102	E	
GRAND	0.95	253.	1.22	0.030	40.667	8.570	0.160	0.2107	0.1770	E	
SINISSIPPI	9.31	987.	1.40	0.090	15.556	6.350	0.010	0.4082	0.4041	E	
TOMAHLINE	0.61	12.	3.78	0.693	5.455	1.400	0.470	0.2567	0.1360	E	
WILLOW R	20.78	847.	4.57	0.326	14.018	0.470	0.110	0.0335	0.0298	N	
BUTTERNUT	4.07	119.	4.24	0.400	10.600	0.640	0.030	0.0604	0.0586	E	
NAGANICKA	4.15	116.	9.91	1.500	6.607	1.330	0.560	0.2013	0.0886	E	
TAINTER	6.85	4351.	4.05	0.027	150.000	20.300	0.120	0.1353	0.1191	E	
WISSOTA	21.97	14375.	8.81	0.046	191.522	8.870	0.150	0.0463	0.0394	E	
ALBERT-LEA	9.93	382.	1.07	0.200	5.350	6.310	0.290	1.1794	0.8374	E	
MC QUADE	0.66	64.	2.74	0.156	17.564	1.200	0.290	0.0683	0.0485	E	
PELICAN	44.29	179.	2.41	3.300	0.730	0.060	0.500	0.0822	0.0411	N	
SILVER	1.71	6.	1.22	2.300	0.530	0.580	0.320	1.0934	0.7435	E	
ANDRUSIA	6.11	1896.	7.92	0.130	60.923	4.020	0.370	0.0660	0.0416	E	
CASS	63.12	2927.	7.62	0.858	8.881	0.326	0.600	0.0367	0.0147	N	
CUKATO	2.20	117.	7.62	1.100	6.927	2.600	0.440	0.3753	0.2102	E	
NEST	3.82	319.	4.57	0.521	8.772	0.800	0.560	0.0912	0.0401	E	
WOLF	4.25	1772.	8.53	0.101	84.455	6.430	0.120	0.0761	0.0670	E	
BEMIDGI	25.98	1632.	9.75	0.734	13.283	0.440	0.360	0.0331	0.0212	E	
BUFFALO	6.11	114.	4.42	1.400	3.157	1.400	0.540	0.4434	0.2040	E	
ESTES	1.57	272.	3.05	0.030	101.667	9.630	0.050	0.0947	0.0900	E	
POYGOUN	44.48	13101.	2.13	0.036	59.167	5.550	0.030	0.0938	0.0910	E	
WISCCHSIN	36.02	23181.	1.83	0.011	166.364	15.210	0.050	0.0914	0.0869	E	
CLEARWATER	12.88	451.	5.18	1.400	3.700	0.670	0.780	0.1811	0.0398	E	
LEECH	453.30	2694.	4.72	5.200	0.908	0.040	0.500	0.0441	0.0220	N	
ALLEGAN	6.42	399.	3.35	0.019	176.316	31.400	0.240	0.1781	0.1353	E	
FORD	4.25	2108.	4.36	0.041	106.341	16.170	0.190	0.1521	0.1232	E	
HIGGINS	38.85	127.	14.94	15.600	0.958	0.030	0.670	0.0313	0.0103	O	
LONG (MICHIGAN)	0.85	84.	5.18	0.085	60.941	4.620	0.090	0.0758	0.0690	E	
HOUGHTON	81.16	494.	2.32	1.300	1.782	0.060	0.500	0.0337	0.0168	N	
STRAWBERRY	1.04	914.	6.74	0.036	189.264	9.190	0.221	0.0486	0.0378	E	
THELNAPPLE	1.66	930.	4.24	0.030	140.791	9.240	0.210	0.0656	0.0518	E	
WHITE	10.41	1305.	6.86	0.153	44.718	1.980	0.222	0.0443	0.0344	E	
ROSS	1.19	1261.	1.52	0.005	277.164	17.020	0.144	0.0614	0.0526	E	

Table 2.5-1 (continued)

LAKE	S AREA KM2	D AREA KM2	Z M	T YRS	QS M/YR	L G/M2-YR	R -	CI G/M3	CQ G/M3	TROPHIC STATE	Reference
SLATERSVILLE	0.84	232.	2.44	0.014	178.029	5.610	0.168	0.0315	0.0262	E	86
SEBAGO	116.49	1026.	30.79	5.490	5.702	0.080	0.500	0.0140	0.0070	O	
BAY OF NAPLES	30.85	305.	4.27	0.071	59.948	0.520	0.192	0.0087	0.0070	O	
MOOSEHEAD	303.22	2977.	16.46	3.000	5.488	0.080	0.375	0.0146	0.0091	O	
LONG	19.71	295.	10.37	1.200	8.638	0.140	0.500	0.0162	0.0081	M	
CHARLEVOIX	69.88	785.	16.77	3.200	5.240	0.120	0.667	0.0229	0.0076	O	
BLACKHOOF	0.74	21.	4.42	0.704	6.279	1.220	0.844	0.1943	0.0303	E	
CRYSTAL	2.93	15.	4.24	3.300	1.284	0.070	0.714	0.0545	0.0156	O	
CHEMUNG	1.26	14.	8.63	4.200	2.054	0.220	0.636	0.1071	0.0389	E	
HARRIMAN RES	8.84	478.	10.37	0.214	48.507	0.880	0.170	0.0181	0.0150	M	
POWDER MILL	1.51	373.	2.47	0.018	138.736	3.250	0.099	0.0234	0.0211	E	
KELLY'S FALLS	0.52	565.	2.26	0.004	550.268	28.810	0.327	0.0524	0.0352	E	
GLEN	0.61	523.	3.35	0.008	408.988	13.130	0.318	0.0321	0.0219	E	
ST CROIX	33.24	19919.	8.78	0.063	139.373	8.890	0.326	0.0638	0.0430	E	
MEMPHREMAGOG	94.62	1686.	15.55	1.700	9.146	0.500	0.420	0.0547	0.0317	M	
WAGONGA	6.54	66.	1.28	1.400	0.914	4.000	0.570	4.3750	1.8812	E	
TROUT	7.65	42.	15.24	17.400	0.876	0.372	0.920	0.4247	0.0340	E	18
CLEAR	—	—	12.50	7.692	1.625	0.040	0.800	0.0246	0.0049	O	
ERIE	—	—	18.00	2.500	7.200	1.060	0.840	0.1472	0.0236	E	
ONTARIO	—	—	84.00	6.579	12.768	0.650	0.780	0.0509	0.0112	M	
SUPERIOR	—	—	148.00	184.679	0.784	0.030	0.900	0.0382	0.0038	O	
MICHIGAN	—	—	84.00	31.250	2.688	0.229	0.900	0.0852	0.0085	M	
227	—	—	4.40	4.202	1.047	0.340	0.890	0.3247	0.0357	E	
OKANAGAN	—	—	75.30	58.824	1.280	0.390	0.950	0.3247	0.0152	M	
SKAHA	—	—	26.50	1.124	23.585	2.200	0.650	0.0933	0.0326	E	
KALAMALKA	—	—	58.00	111.111	0.522	0.320	0.900	0.6130	0.0613	E	
WOOD	—	—	21.00	108.696	0.193	0.500	0.900	2.5880	0.2588	E	
NENONA	—	—	7.80	1.200	6.497	2.140	0.710	0.3294	0.0955	E	
WAUBESA	—	—	4.80	0.300	15.984	9.930	0.0	0.6212	0.6212	E	
KEGUNSA	—	—	4.60	0.350	13.154	6.670	0.090	0.5071	0.4614	E	
TAHOE	—	—	303.00	714.286	0.424	0.042	0.930	0.0990	0.0069	O	93
AEGERISEE	—	—	49.00	8.696	5.635	0.160	0.680	0.0284	0.0091	O	
TURLERSEE	—	—	14.00	2.151	6.510	0.300	0.800	0.0461	0.0092	M	
HALLWILERSEE	—	—	28.00	3.846	7.280	0.550	0.360	0.0756	0.0484	M	
BODENSEE-OBERSEE	—	—	100.00	4.878	20.500	1.100	0.650	0.0537	0.0188	E	
PFAFFIKERSEE	—	—	18.00	2.597	6.930	1.360	0.770	0.1962	0.0451	E	
ZURICHSEE	—	—	50.00	1.471	34.000	1.320	0.250	0.0389	0.0291	E	
GREIFENSEE	—	—	19.00	2.041	9.310	1.565	0.620	0.1681	0.0639	E	
BALDEGGERSEE	—	—	34.00	4.545	7.480	1.750	0.610	0.2340	0.0912	E	
NORRVIKEN	—	—	5.40	0.571	9.450	4.030	0.490	0.4264	0.2175	E	
SEMPACHERSEE	—	—	46.00	16.949	2.714	0.770	0.940	0.2837	0.0170	E	16
CAMERON	12.72	3250.	7.10	0.063	112.698	7.960	0.340	0.0174	0.0115	M	
FOUR-MILE	7.73	45.	9.30	4.310	2.158	0.108	0.830	0.0501	0.0085	M	
BOB	2.27	29.	18.00	3.020	5.960	0.157	0.730	0.0263	0.0071	O	
12-MILE-BUSHKUNG	11.65	933.	18.10	0.480	37.708	0.330	0.360	0.0068	0.0056	O	
HALLS	5.32	270.	27.20	1.120	24.286	0.205	0.530	0.0084	0.0040	O	
BEECH	1.36	515.	9.80	0.052	188.462	1.520	0.060	0.0081	0.0076	O	
MAPLE	3.32	502.	11.60	0.148	78.378	0.830	0.320	0.0106	0.0072	O	
PINE	1.06	285.	7.40	0.058	127.586	1.055	0.020	0.0083	0.0081	M	
CRANBERRY	0.69	269.	3.50	0.017	205.882	1.300	-0.040	0.0063	0.0066	O	
EAGLE-MOOSE	5.15	251.	12.80	0.521	24.568	0.225	0.320	0.0092	0.0062	O	
OBLONG-HALIBURTON	11.03	143.	17.70	2.600	6.808	0.124	0.720	0.0182	0.0051	O	

2.5.1 Lake Data Base Description (continued)

As an initial step in the analysis, the statistical distributions of the variables in Table 2.5-1 were examined. It was evident that a logarithmic transformation was appropriate to promote symmetry in the distributions in all of the variables except R , the retention coefficient. Histograms of the transformed data, stratified by trophic state, were generated using the BMDP7D computer program [26], which also computed gross and within-group means, standard deviations, and ranges, as well as analyses of variance to test the statistical significance of variations across groups. Table 2.5-2 identifies the variables examined and displays the results.

The summary F table indicates that there were significant differences across groups for all variables, with significance levels ranging from greater than 99.99% for L , CI , CO , and L/Z to 93.02% for R . The highest F level observed was 66.2, in the case of the average outlet concentration, CO. Trophic states were most highly stratified, or, equivalently, most distinctly classified, on this variable. This is consistent with Dillon's model. The stratification with respect to depth is consistent with Vollenweider's first model, i.e., on the average, eutrophic lakes were shallower. However, the stratification with respect to surface overflow rate, QS , is contrary to Vollenweider's second model, which predicts that higher surface overflow rates should be associated with oligotrophic lakes. Because of

Table 2.5-2

Identification of Variables and Analyses of Variance Across Trophic Groups

Variable			Units	Transfor-mation	F Statistic[a]	Prob. F Exceeded	Table
1	Z	mean depth	(m)	$\log_{10}$	10.51	.0001	B
2	T	hydraulic residence time	(yr)	$\log_{10}$	7.82	.0007	C
3	QS	surface overflow rate	(m/yr)	$\log_{10}$	3.35	.0391	A
4	L	phosphorus loading per unit area	$(g/m^2\text{-yr})$	$\log_{10}$	52.90	<.0001	D
5	L/Z	phosphorus loading per unit volume	$(g/m^3\text{-yr})$	$\log_{10}$	43.44	<.0001	E
6	CI	average inlet phosphorus concentration	(g/m^3)	$\log_{10}$	37.84	<.0001	G
7	CO	average outlet phosphorus concentration	(g/m^3)	$\log_{10}$	66.21	<.0001	H
8	R	phosphorus retention coefficient	-	-	2.73	.0698	F

a - degrees of freedom = 2,102

Table 2.5-2 (continued)

A - Surface Overflow Rate

TABULATION OF VARIABLE 2 OS	WITH STRATIFICATION ON VARIABLE 9 TS		EXCLUDED VALUES TS
	OLIG	MESO	EUTR
MIDPOINTS			

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE.

MEAN	0.876	1.016	1.343			
S. DEV.	0.743	0.722	0.835			
N	22.000	23.000	60.000			
MAXIMUM	2.314	2.973	2.748			
MINIMUM	-0.372	-0.136	-0.714			

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

			SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	PROB. F EXCEEDED
MEAN	1.174	BETWEEN	4.2285	2	2.1143	3.3482	0.0391
S. DEV.	0.812	WITHIN	64.4097	102	0.6315		
N	105.	TOTAL	68.6382	104			
MAXIMUM	2.748						
MINIMUM	-0.714						

B - Mean Depth

TABULATION OF VARIABLE 1 2	WITH STRATIFICATION ON VARIABLE 9 TS		EXCLUDED VALUES TS
	OLIG	MESO	EUTR
MIDPOINTS			

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

MEAN	1.250	1.047	0.766			
S. DEV.	0.669	0.471	0.447			
N	22.000	23.000	60.000			
MAXIMUM	2.481	1.924	2.000			
MINIMUM	0.364	0.365	0.029			

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

			SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	PROB. F EXCEEDED
MEAN	0.931	BETWEEN	4.3123	2	2.1561	10.5046	0.0001
S. DEV.	0.473	WITHIN	20.9361	102	0.2053		
N	105.	TOTAL	25.2483	104			
MAXIMUM	2.481						
MINIMUM	0.029						

Table 2.5-2 (continued)

C - Mean Hydraulic Residence Time

```
TABULATION OF VARIABLE     2 T      WITH STRATIFICATION ON VARIABLE     9 TS
                                                                                    EXCLUDED
                                                                                    VALUES
            XIG                       NESO                      EUTR                 TS
.............................................................................................
MIDPOINTS
   3.000I
   4.0001*
   2.700I
   2.500I*
   2.100I                                                      1 00
   1.800I                       *
   1.500I                       *
   1.300I**                     *                              **
   0.900I**                     **
   0.600I********               ****                           ****
   0.300I*                      **                             ******0**
  -0.000I*                      *0*                            *******0*
  -0.300I**                     *                              *******0*
  -0.600I                       ***                            0***
  -0.900I*                      *****                          *****0
  -1.200I**                     **                             ***
  -1.500I                                                      ******0***
  -1.800I*                                                     *****
  -2.100I*                      *                              ***
  -2.400I                                                      ***
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

MEAN        0.385                0.029                        -0.577
S. DEV.     1.370                0.945                         1.069
N          82.000               23.000                        60.000
MAXIMUM     2.334                1.170                          2.046
MINIMUM    -1.770               -2.000                        -2.523

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

                                     SUM OF SQUARES   DF   MEAN SQUARE   F RATIO   PROB. F EXCEEDED
MEAN       -0.243    BETWEEN          17.0257          2    8.5129        7.8150    0.0007
S. DEV.     1.110    WITHIN          111.1089        102    1.0893
N         105.       TOTAL           128.1346        104
MAXIMUM     2.054
MINIMUM    -2.523
```

D - Phosphorus Loading Per Unit Area

```
TABULATION OF VARIABLE     4 L      WITH STRATIFICATION ON VARIABLE     9 TS
                                                                                    EXCLUDED
                                                                                    VALUES
            CL15                      NESO                      EUTR                 TS
.............................................................................................
MIDPOINTS
   2.0001
   1.8001
   1.6001                                                      *
   1.4001                                                      ****
   1.2001                                                      *****
   1.0001                       *                              *********0*
   0.8001                                                      ******0
   0.6001                                                      *******
   0.4001                       **                             ***
   0.2001**                     **                             *******0***
  -0.0001*                      **                             *****
  -0.2001*                      *****                          *****
  -0.4001*                      ****0                          ********0
  -0.6001**                     **                             *
  -0.8001**                     **
  -1.0001********               *
  -1.2001**                     **
  -1.4001**                     **
  -1.6001**
  -1.8001
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

MEAN       -0.533               -0.444                         0.482
S. DEV.     0.582                0.581                          0.591
N          22.000               23.000                         60.000
MAXIMUM     0.182                0.920                          1.525
MINIMUM    -1.523               -1.398                         -0.658

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

                                     SUM OF SQUARES   DF   MEAN SQUARE   F RATIO   PROB. F EXCEEDED
MEAN        0.004    BETWEEN          33.7347          2   16.8674       52.9497    0.0000
S. DEV.     0.798    WITHIN           32.4925        102    0.3186
N         105.       TOTAL           66.2273        104
MAXIMUM     1.525
MINIMUM    -1.523
```

Table 2.5-2 (continued)

E - Phosphorus Loading Per Unit Volume

```
TABULATION OF VARIABLE    S LZE    WITH STRATIFICATION ON VARIABLE    9 TS
```

	OLIG	MESO	EUTR	EXCLUDED VALUES TS

MIDPOINTS

(group means are denoted by M's if they coincide with *'s, M's otherwise)

	OLIG	MESO	EUTR
MEAN	-2.042	+1.491	-0.284
S. DEV.	0.813	0.900	0.900
N	22.000	23.000	60.000
MAXIMUM	-0.431	0.395	1.176
MINIMUM	-3.851	-2.701	-2.250

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

			SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	PROB. F EXCEEDED
MEAN	-0.927	BETWEEN	61.9901	2	30.9958	43.4413	0.0000
S. DEV.	1.138	WITHIN	72.7762	102	0.7135		
N	105.	TOTAL	134.7663	104			
MAXIMUM	1.176						
MINIMUM	-3.851						

F - Retention Coefficient

```
TABULATION OF VARIABLE    R R    WITH STRATIFICATION OF VARIABLE    9 TS.
```

	OLIG	MESO	EUTR	EXCLUDED VALUES TS

MIDPOINTS

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

	OLIG	MESO	EUTR
MEAN	0.541	0.463	0.384
S. DEV.	0.261	0.261	0.287
N	22.000	23.000	60.000
MAXIMUM	0.930	0.930	0.930
MINIMUM	-0.040	0.020	-0.080

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

			SUM OF SQUARES	DF	MEAN SQUARE	F RATIO	PROB. F EXCEEDED
MEAN	0.434	BETWEEN	0.4200	2	0.2100	2.7321	0.0698
S. DEV.	0.282	WITHIN	7.8409	102	0.0769		
N	105.	TOTAL	8.2609	104			
MAXIMUM	0.930						
MINIMUM	-0.080						

Table 2.5-2 (continued)

G - Average Inlet Phosphorus Concentration

```
TABULATION OF VARIABLE      6 CI      WITH STRATIFICATION ON VARIABLE      9 TS

                                                                        EXCLUDED
                                                                        VALUES
              OLIG                    MESO                   EUTR        TS
MIDPOINTS  ..........................................................................
 1.200I
 1.000I
 0.800I
 0.600I
 0.400I
 0.200I                                                      *
 0.000I                                                      *
-0.200I                                                      ***
-0.400I                                                      *******
-0.600I                               *                      *****
-0.800I                                                      *********
-1.000I*                              ***                    ***********
-1.200I*                              ***                    ***********
-1.400I***                            M*****                 ********
-1.600I*******                        **                     ***
-1.800I****                           ******
-2.000I******                         **
-2.200I*
-2.400I
-2.600I

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

MEAN       -1.704              -1.442              -0.861
S. DEV.     0.313               0.369               0.492
N          22.000              23.000              60.000
MAXIMUM    -1.005              -0.516               0.641
MINIMUM    -2.200              -2.083              -1.630

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

                                     SUM OF SQUARES   DF   MEAN SQUARE  F RATIO   PROB. F EXCEEDED

MEAN       -1.171     BETWEEN       14.0763       2    7.0382    37.8343    0.0000
S. DEV.     0.564     WITHIN        18.9722     102    0.1860
N          105.       TOTAL         33.0485     104
MAXIMUM     0.641
MINIMUM    -2.200
```

H - Average Outlet Phosphorus Concentration

```
TABULATION OF VARIABLE      7 CO      WITH STRATIFICATION ON VARIABLE      9 TS

                                                                        EXCLUDED
                                                                        VALUES
              OLIG                    MESO                   EUTR        TS
MIDPOINTS  ..........................................................................
 0.450I
 0.300I                                                      *
 0.150I
-0.000I
-0.150I                                                      ***
-0.300I                                                      *
-0.450I                                                      *
-0.600I                                                      **
-0.750I                                                      ***
-0.900I                                                      ******
-1.050I                                                      *******
-1.200I                                                      M****
-1.350I                              ***                     ***********
-1.500I                              **                      **********
-1.650I                              **                      *******
-1.800I*                             M****                   **
-1.950I****                          ****
-2.100I**********                    *******
-2.250I******
-2.400I*

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, M'S OTHERWISE

MEAN       -2.127              -1.816              -1.190
S. DEV.     0.155               0.203               0.352
N          22.000              23.000              60.000
MAXIMUM    -1.806              -1.316              -0.274
MINIMUM    -2.417              -2.161              -1.769

ALL GROUPS COMBINED (CASES EXCLUDED IF SPECIAL CODES FOR EITHER VARIABLE)

                                     SUM OF SQUARES   DF   MEAN SQUARE  F RATIO   PROB. F EXCEEDED

MEAN       -1.501     BETWEEN       18.2660       2    9.1345    66.2120    0.0000
S. DEV.     0.558     WITHIN        14.0727     102    0.1380
N          105.       TOTAL         32.3407     104
MAXIMUM     0.274
MINIMUM    -2.417
```

2.5.1 Lake Data Base Description (continued)

intercorrelations among these variables, this univariate analysis
should be viewed only as a preliminary description of the data,
and not as a basis for evaluating models.

In order to further characterize the data, the product-
moment correlation matrix of the variables is presented in
Table 2.5-3. A high degree of association among the variables
is indicated. Accordingly, a univariate or bivariate analysis
would be virtually meaningless in establishing cause-effect
relationships. Using the BMDP2R program[26], stepwise linear
regressions were performed to determine the proportion of the
variance of each variable which could be explained by those
remaining. In this analysis, the variables were grouped into
two categories: "independent" and "dependent". Trophic state,
outlet concentration, and retention coefficient were included
in the latter because they are viewed as response variables,
whose values depend upon the lake's nutrient dynamics. Only
"independent" variables were allowed to enter into the regression
analyses, subject to significance and tolerance constraints. Table
2.5-3 shows that between 55 and 87% of the variance in the inde-
pendent variables could be explained. A linear model including
T , L , and CI could explain 90% of the variance in the outlet
concentration data. Of the three linearly*-related variables Z ,

---

* On logarithmic scales, $QS = Z - T$.

Table 2.5-3

Correlations Among Variables in Analysis

Product-Moment Correlation Coefficients :

Variable [a]

	Z	T	QS	L	L/Z	CI	CO	R	TS[b]
Z	1.000								
T	.745	1.000							
QS	-.411	-.915	1.000						
L	-.529	-.788	.755	1.000					
L/Z	-.804	-.875	.707	.930	1.000				
CI	-.157	.203	-.372	.327	.293	1.000			
CO	-.512	-.242	.020	.618	.655	.847	1.000		
R	.631	.827	-.747	-.598	-.692	.230	-.289	1.000	
TS	-.412	-.361	.244	.701	.670	.640	.741	-.226	1.000

Results of Stepwise Linear Regressions :

Equation	R^2	SEE	Variables Allowed[c] to Enter
Z = .690 + .360 T - .281 CI	.654	.293	T,L,CI,QS
Z = 1.011 + .331 T	.555	.330	T,L,QS
T = .951 - 1.331 L + 1.016 CI	.858	.442	Z,L,CI,QS
T = -1.196 + 1.027 Z - .760 L	.770	.537	Z,L,QS
L = .690 - .640 T + .719 CI	.868	.293	Z,T,CI,QS
L = -.134 - .566 T	.620	.495	Z,T,QS
C1 = -1.024 + .616 T + .907 L	.666	.329	Z,T,L,QS
CO = -.371 - .310 T - .145 L + 1.029 CI	.901	.178	Z,T,L,CI,QS
1-R = .515 - .210 T	.684	.159	Z,T,L,CI,QS
$\log_{10}$[1-R] = -.401 - .292 T - .119 L	.680	.177	Z,T,L,CI,QS

a - all variables transformed to $\log_{10}$, except R and TS
b - TS = trophic state = 1,2,3, for oligotrophic, mesotrophic, and
    eutrophic lakes, respectively
c - subject to F ≥ 3.0 and tolerance ≥ .01

2.5.1 Lake Data Base Description (continued)

T , and QS , hydraulic residence time alone entered significantly
in all of the regressions.

The regressions in Table 2.5-3 serve as means of character-
izing the multivariate distribution of the data. Because of the
high degree of collinearity it will be difficult to establish
specific cause-effect relationships based upon these data.
Accordingly, application of any of the analysis below to other
lakes should be done with extreme caution. A gauge of the appli-
cability of the analysis to a given lake would be how well it
conforms to the multivariate distribution characterized in Tables
2.5-2 and 2.5-3. If a lake's depth, residence time, loading, and
average inlet concentration are within the ranges of the distribu-
tions in Table 2.5-2 and if they can be predicted reasonably well
by the relationehips in Table 2.5-4, as gauged by the respective
standard errors of estimate, then the lake could be considered a
member of the same "population" and the results discussed below
could be applied with increased confidence. Applicability is
also subject to geographical constraints.

2.5.2 Phosphorus Retention Models

Dillon's model suggests that the retention coefficient for total phosphorus is an important factor required to estimate the response of a lake's trophic state to a given phosphorus loading. In order to predict such response a priori, i.e., without taking any measurements on the lake or on its outlet, a means of estimating R from known morphometric or hydrologic factors is required. Kirchner and Dillon[42] suggested an empirical formulation which permitted estimation of R from the surface overflow rate, QS (equation (2.4-3)). The next step in the analysis is to examine some more theoretically-based models for prediction of the phosphorus retention coefficient and to compare them with the strictly empirical models in Table 2.5-3 and with Dillon and Kirchner's equation. One primary objective is to estimate the extent to which the sedimentation of phosphorus can be represented as a first-order reaction. As discussed previously, this is one of the principal assumptions made in the application of Dillon's model to predict the ultimate response of a given lake to a reduction or increase in phosphorus loading.

The qualified success of Vollenweider's simple, first-order model[15,94] for phosphorus removal suggested that it serve as a basis for the present analysis:

$$1 - R = \frac{1}{1 + KT} \qquad (2.5-1)$$

2.5.2 <u>Phosphorus Retention Models</u> (continued):

where,

K = effective first-order phosphorus
decay coefficient (1/yr)

A more general form of this model is proposed in which K is
allowed to vary with depth and residence time:

$$K = a \ [1/T]^{b} \ z^{c} \tag{2.5-2}$$

The inclusion of these effects can be considered as a test
of the simple first-order model, in which K is assumed to be
constant. Evaluation of the extent to which K , or, equi-
valently, the parameters a , b , and c , vary with concen-
tration can provide a basis for evaluating first-order kinetics.

Using the above model, the parameters a , b , and c
have been estimated for each of the five subsets of the original
105-lake data base: one for each of the three trophic states;
one containing all lakes; and one containing lakes with average
inlet concentrations less than 0.1 g/m^{3}. The data set was
stratified to examine the stability of the parameters and hence
the validity of the model across groups. The last group was
included to eliminate those lakes with extemely high phosphorus

2.5.2   Phosphorus Retention Models (continued):

concentrations, which had generally higher standard errors of

estimate and for which trophic classification would be obvious

in any case.   The BMDP3R nonlinear regression program[26] was

employed to estimate the parameters.

Results are given in Table 2.5-4.   The parameter expressing

depth-dependence, c , was not significantly different from zero

in any of the groups.   Accordingly, the model was simplified by

setting c equal to zero and new estimates for parameters a

and b were derived.   t-Tests indicated no significant differ-

ences of the parameter estimates across groups.   Thus, the first-

order model appeared to be stable, with optimal parameter estimates

independent of concentration.   The standard errors of estimate

were 0.110, 0.155, and 0.164 for oligotrophic, mesotrophic, and

eutrophic lakes, respectively.

Subsequent analyses have been performed using the parameter

estimates derived from the last data set in Table 2.5-4, i.e.

a = 0.824 , b = 0.546 , and c = 0 .   Normality in the residuals

is indicated by the probability plot in Figure 2.5-1.   Figures

2.5-2 and 2.5-3 plot observed versus estimated retention coeffi-

cients and average outlet concentrations, respectively.   The

symbols O , M , and E have been used to represent the various

trophic states.   Residuals patterns generally reflect higher

standard errors in the eutrophic groups, but no trend.   No trends

Table 2.5-4

Parameter Estimates for Phosphorus Retention Model

Model : $1 - R = \dfrac{1}{1 + TK}$

$K = a \left[\dfrac{1}{T}\right]^{b} Z^{c}$

T = hydraulic residence time (yrs)
Z = mean depth (m)
K = effective first-order sedimentation coefficient (1/yr)
R = retention coefficient
a,b,c = estimated parameters

Statistic	Oligotrophic	Mesotrophic	Data Set Eutrophic	All Lakes	CI < .1g/m³
N	22	23	60	105	71
a	1.383 ± .785*	.711 ± .301	.832 ± .249	.866 ± .188	.803 ± .227
b	.445 ± .092	.531 ± .111	.471 ± .083	.500 ± .054	.549 ± .058
c	-.204 ± .212	.040 ± .176	.114 ± .133	.023 ± .091	.010 ± .111
SEE**	.113	.159	.165	.150	.137
R^{2}***	.849	.704	.693	.723	.727
(c = 0.) a	.817 ± .105	.778 ± .118	1.050 ± .124	.912 ± .068	.824 ± .067
b	.489 ± .080	.517 ± .092	.434 ± .076	.492 ± .046	.546 ± .046
SEE	.110	.155	.164	.150	.136
R^{2}	.849	.704	.693	.723	.727

* estimate ± one standard deviation
** SEE = standard error of estimate = standard deviation of residuals, adjusted for number of parameters estimated
*** R^{2} = 1 - [residual sum of squares/ observed sum of squares]
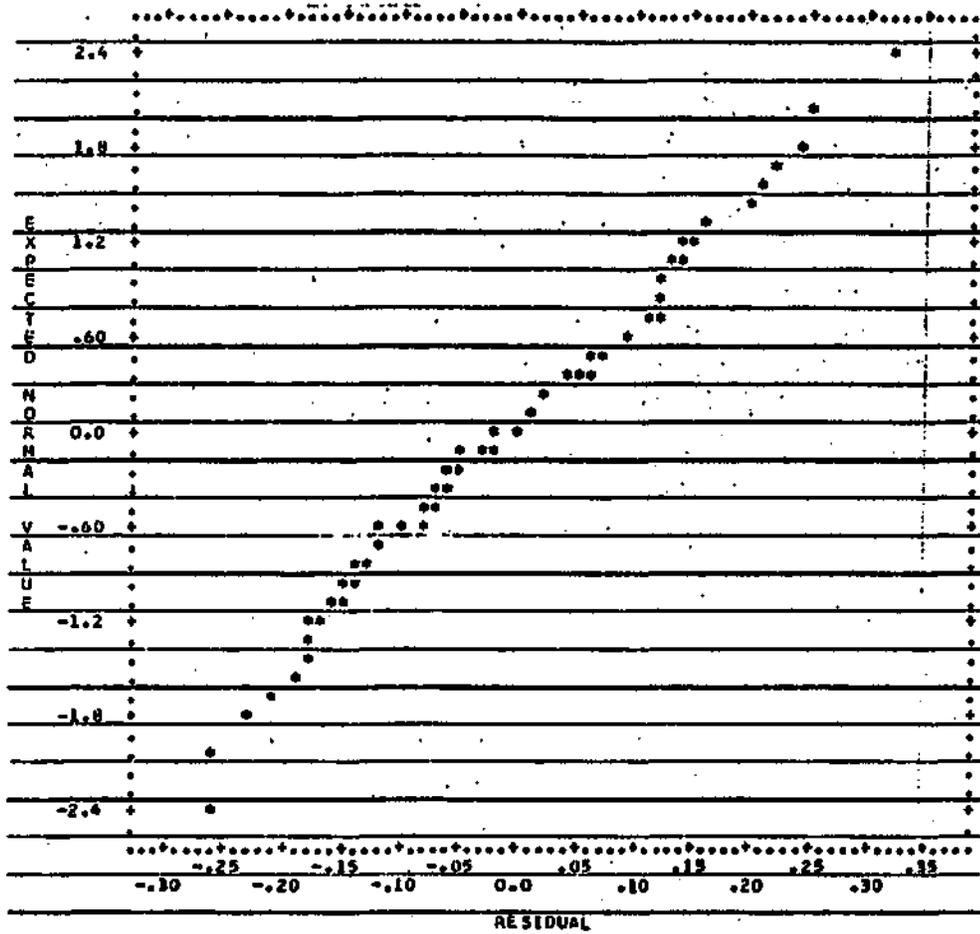
Figure 2.5-1

Normal Probability Plot of Retention Coefficient Residuals
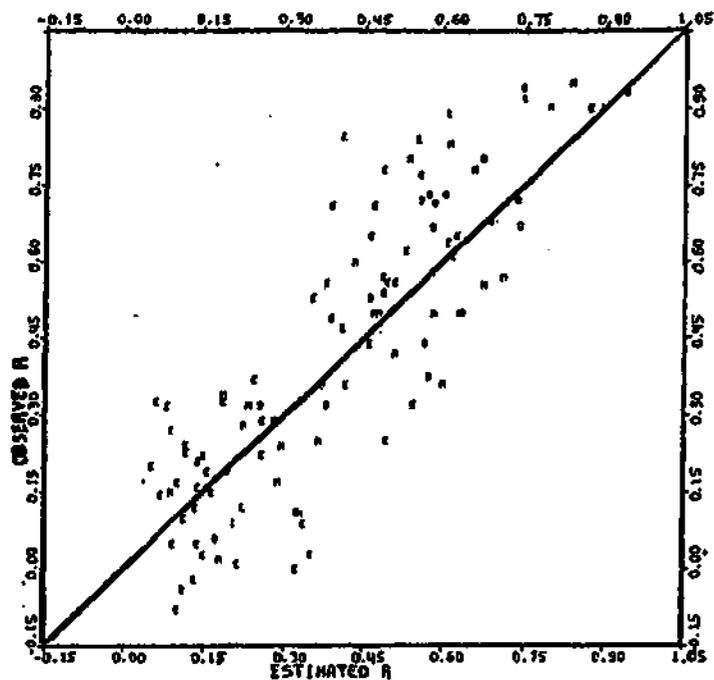
Figure 2.5-2

Observed and Estimated Phosphorus Retention
Coefficients
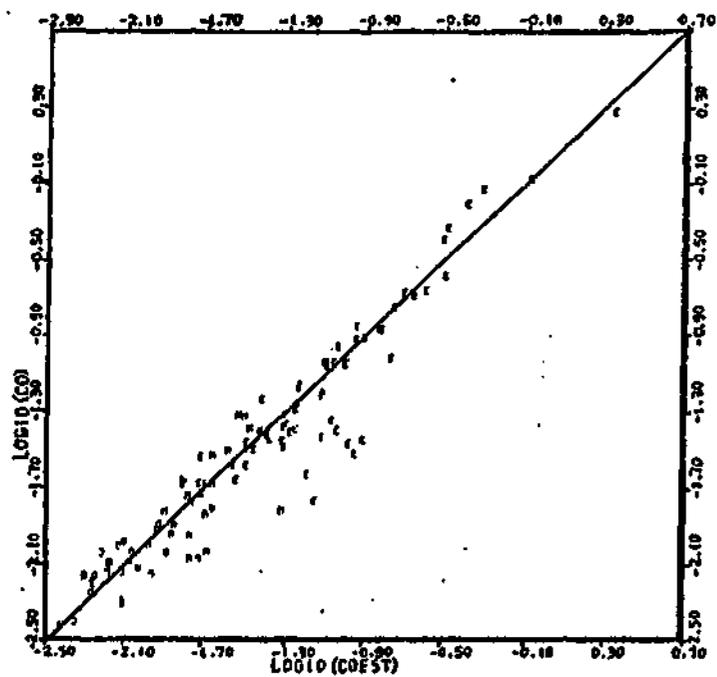
Figure 2.5-3

Observed and Estimated Phosphorus Outlet
Concentrations

2.5.2 Phosphorus Retention Models (continued)

were evident in plots of residuals versus other independent
variables, including Z, T, QS, L, lake surface area, or
drainage area.

Table 2.5-5 compares the standard errors of estimate in
predicting the retention coefficient and average outlet concen-
tration with the respective standard errors of the empirical
models in Table 2.5-3 and of Dillon and Kirchner's equation.
The derived model performs about as well as the empirical linear
regressions and generally performs better than the latter.
Advantages of the non-linear model over the linear regressions
are its partial theoretical basis, relative parsimony (2 param-
eters), and property that the computed retention coefficient is
restricted to the range of 0 to 1. Using only the eleven lakes
which were in common with Dillon and Kirchner's data set, the
standard errors of the retention coefficient estimates were
0.137 and 0.111, for the non-linear model and Dillon and Kirchner's
equation, respectively.

Parameter estimates indicate that the effective first-order
decay coefficient in these lakes is roughly inversely proportional
to the square root of the mean hydraulic residence time. This
dependence is further examined in Table 2.5-6 and Figure 2.5-4.
Table 2.5-6 presents analyses of variance on the decay coefficient
in the original form, K, and on a form corrected for residence

Table 2.5-5

Comparisons of Phosphorus Retention Models[b]

Model	Predicted Variable			
	Retention Coefficient		Average Outlet Conc.[a]	
	R^2	SEE	R^2	SEE
Dillon and Kirchner (equation 2.4-3) $R = .426 \exp(-.271\ QS) + .574 \exp(-.00949\ QS)$	.512	.197	.807	.245
Linear Regressions (Table 2.5-3) $1 - R = .515 - .210\ T$ $CO = -.371 - .310\ T - .145\ L + 1.029\ CI$	.684 -	.159 -	- .901	- .178
Nonlinear Regression (Table 2.5-4) $1 - R = 1/[\ 1 + .824\ T^{.454}]$	.720	.151	.906	.171

a - base 10 logarithm

b - 105 lakes included in analysis

Table 2.5-6

Histograms of Observed and "Corrected" First-Order Sedimentation
Coefficients for Total Phosphorus

A - Observed

B - Corrected

Figure 2.5-4

Dependence of First-Order Phosphorus Sedimentation Coefficient on
Hydraulic Residence Time

2.5.2 Phosphorus Retention Models (continued)

time effects, K_c :

$$K = \frac{R}{1 - R} \frac{1}{T}$$ (2.5-3)

$$K_c = K T^b , \qquad b = 0.546$$ (2.5-4)

The $\log_{10}$ transformation has been applied to these variables, and the four lakes in the data set with negative retention coefficients have been excluded. Table 2.5-6 indicates significant stratification of the uncorrected decay coefficient, K , with trophic state, but insignificant stratification of the coefficient after it has been corrected for the effects of hydraulic residence time. The standard deviations of the corrected decay coefficients are 0.224, 0.391, and 0.468 for oligotrophic, mesotrophic, and eutrophic lakes, respectively. This may reflect a corresponding increase in noise levels in the systems or in the data. The effect of T on K is also illustrated in Figure 2.5-4. Again, greater deviations are observed for eutrophic lakes.

The parameter estimates indicate that lake mean depth alone has little influence on phosphorus sedimentation rate, except insofar as depth may influence residence time. The final two-parameter model is similar to that proposed independently by

2.5.2 Phosphorus Retention Models (continued)

Larsen and Mercier[45], whose estimates for a and b were 1.12
and 0.49, respectively, as derived from a sample of 20 oligotrophic
and mesotrophic lakes. The apparent dependence of K on T
could be explained in part by deviations from the completely-mixed
assumption inherent in the first-order model formulation. Equa-
tions (1) and (2) can be combined to give:

$$1 - R = \frac{1}{1 + KT} = \frac{1}{1 + a\,T^{1-b}} \qquad (2.5\text{-}5)$$

If the true sedimentation coefficient were independent of T ,
then the same results would be reached if T were replaced by
an effective hydraulic residence time, given by:

$$T_{ef} = T^{1-b} \qquad (2.5\text{-}6)$$

Substitution of T_{ef} for T^{1-b} in equation (5) gives the
standard first-order equation:

$$1 - R = \frac{1}{1 + a\,T_{ef}} \qquad (2.5\text{-}7)$$

2.5.2   Phosphorus Retention Models (continued)

In terms of lake volume, V , and mean flow, Q :

$$T_{ef} = [\frac{V}{Q}]_{ef} = [\frac{V}{Q}]^{1-b} \qquad (2.5-8)$$

Conceivably, thermal stratification could reduce the effective volume of the lake and partially account for the observed dependence of K on T . Vollenweider attempted to account for this on one of his earlier modifications of the first-order model [15] He suggested that the effects of stratification could be incorporated into the model by employing the following expression for T_{ef} :

$$T_{ef} = \frac{V}{Q} + \frac{V}{V_e} - 1 \qquad (2.5-9)$$

where V_e , "mean exchange epilimnion", was defined as that portion of the lake's volume taking part in the washout process. He did not propose a general model for V_e , however.

Temperature effects may also account in part for higher mean sedimentation coefficients observed in lakes with lower residence times. The temperatures of low residence time lakes

2.5.2 Phosphorus Retention Models (continued)

would be expected to respond faster to seasons, because of their

higher advective thermal energy exchange rates. The response

time would not depend upon lake mean depth, because of the

apparent linear increase of vertical transport rates with depth,

as noted by Snodgrass[78] and Blanton[7]. This dependence would

cancel the effects of higher surface-area-to-volume ratios in

shallower lakes. Accordingly, one would expect that the peak,

volume-averaged temperature would decrease with increasing

hydraulic residence time. Assuming that the reactions involved

in phosphorus sedimentation are temperature-dependent, this could

partially account for the observed dependence of K on T .

A third explanation is that the reactions for phosphorus

deposition may be dependent upon the presence of other substances,

for example, other nutrients for biologically mediated reactions,

iron for chemically mediated reactions, or sediment, for adsorption-

sedimentation reactions. Lakes with higher flushing rates would

have greater supplies of these supplementary materials, which may,

in turn, increase the net removal rates of phosphorus from the

systems.

The above explanations should be qualified with a note that

multicolinearity in the data (Table 2.5-3) generally makes it

difficult to attribute variations in the dependent variable to

specific factors. Theoretical interpretation of the parameter

2.5.2   <u>Phosphorus Retention Models</u> (continued)

estimates is thus rather tenuous. This re-emphasizes the need
to insure that a given lake conforms to the correlation structure
of these data before applying any of the previous or forthcoming
results to that lake.

The stability of the parameter estimates across trophic
states has not provided a basis for rejecting the validity
of a first-order model in which the retention coefficient is
independent of concentration. Multicolinearity and/or errors
in the data may have effectively crippled the ability of the
analysis to detect significant parameter differences across
groups, however. The first-order assumption could best be
tested with time· series data from a number of lakes. The impli-
cations of first-order behavior are particularly important in
two respects: in predicting system response to changes in
phosphorus loading and in predicting response to changes in
hydrologic regime. The former has implications in the alloca-
tion of acceptable nutrient loadings. The latter has implica-
tions in the interpretation of data obtained from sampling under
a given set of hydrologic conditions. This involves the conversion
or extrapolation of a lake's nutrient budget, measured during a
given year, to long-term-average conditions.

The higher standard errors of estimate characteristic of the
eutrophic group may be attributed to a number of factors. From a

2.5.2 Phosphorus Retention Models (continued)

theoretical point of view, this could be due to certain character-
istics of eutrophic lakes which profoundly affect nutrient dynamics,
such as anaerobiosis or zero-order dependence of algal growth rates
on phosphorus concentrations. It may also be attributed to higher-
frequency or higher-amplitude variations of phosphorus concentrations
in eutrophic lakes. Such variations would increase the errors in
the retention coefficients estimated from a fixed number of lake
outlet samples per year (typically, 12, for the NES lakes). Lower
stability is a general characteristic of eutrophic ecosystems. In
addition, in this particular data set, the geometric mean hydraulic
residence time for the eutrophic group is 0.25 years, as compared
to 2.42 and 1.07 years for the oligotrophic and mesotrophic groups,
respectively (Table 2.5-2). A lower residence time would cause
less attenuation of the relatively high-frequency variations
characteristic of lake inflows. This, in turn, would increase
the sampling frequency required in order to quantify average lake
(or outlet) conditions to within a given standard error.

2.5.3 Error Analysis

The residuals obtained in estimating the retention coefficient according to the above model result from errors in measurement (input/output estimates), as well as from model errors. In order to disect the residual variance into these two components, some assumptions must be made about their forms. The actual and measured values of $1 - R$ are assumed to be given by:

$$Y = 1 - R = \frac{CO}{CI} \qquad (2.5-10)$$

$$Y_m = \frac{CO\,(1 + e_o)}{CI\,(1 + e_i)} \qquad (2.5-11)$$

where,

e_o, e_i = independent, normal, random deviates with mean zero and standard deviations S_{e_o} and S_{e_i}, respectively

The variables e_o and e_i represent measurement errors in outlet and inlet concentration estimates. Their standard deviations equal the coefficients of variation of the respective estimates:

$$S_{e_o} = CV_{CO_m} \qquad (2.5-12)$$

2.5.3 Error Analysis (continued)

$$S_{e_i} = CV_{ci_m} \qquad (2.5-13)$$

According to the first-order model, the estimated value of
Y is given by:

$$Y_e = \frac{1}{1 + K_e T} = \frac{1}{1 + KT(1 + d_k)} \qquad (2.5-14)$$

where,

K_e = estimated first-order decay coefficient (1/yr)

K = actual first-order decay coefficient (1/yr)

d_k = model error variable with mean zero and standard deviation s_{k_k}

A calculated residual is given by:

$$r_y = Y_m - Y_e = \frac{CO\,(1 + e_o)}{CI\,(1 + e_i)} - \frac{1}{1 + KT(1 + d_k)}$$

$$= \frac{(1 + e_o)}{(1 + e_i)} \frac{1}{1 + KT} - \frac{1}{1 + KT(1 + d_k)} \qquad (2.5-15)$$

2.5.3 <u>Error Analysis</u> (continued)

Since the error variables are assumed to be independent, the
variance of the residuals can be calculated from the following
expected value formula[5] :

$$s_{r_y}^2 = \left(\frac{\partial r_y}{\partial e_o}\right)^2 s_{e_o}^2 + \left(\frac{\partial r_y}{\partial e_i}\right)^2 s_{e_i}^2 + \left(\frac{\partial r_y}{\partial d_k}\right)^2 s_{d_k}^2 \qquad (2.5\text{-}16)$$

$$= Y^2 \left(s_{e_o}^2 + s_{e_i}^2\right) + Y^2 \left(1 - Y\right)^2 s_{d_k}^2 \qquad (2.5\text{-}17)$$

The first term in this equation essentially represents
measurement error, and the second, model error. The corres-
ponding expression for the residual variance obtained in
estimating the logarithm of the outlet concentration is given
by:

$$s_{r_c}^2 = s_{e_o}^2 + s_{e_i}^2 + (1 - Y)^2 s_{d_k}^2 \qquad (2.5\text{-}18)$$

Equation (18) indicates that the measurement error component
of the concentration residuals is constant, while the model
error component decreases to zero as Y approaches 1, or as
T approaches zero. This property is used below to quantify
the two components, via further analysis of observed residual

2.5.3 Error Analysis (continued)

patterns.

For any range of Y values, a value of $S^2_{r_c}$ can be computed directly from the residuals. Equation (18) can be viewed as a linear model for the residual variance, in which the independent variable is $(1 - Y)^2$, the slope is measurement error, $(S^2_{e_0} + S^2_{e_1})$, and the intercept is model error, $(S^2_{d_k})$. In order to estimate the slope and intercept, the lake data have been divided into five groups of approximately equal sample size, based upon residence time (or Y values). For each group, the mean $(\overline{Y})$, variance (S^2_Y), and mean squared residual have been calculated. In order to account for variation of Y within each group, the expected value of $S^2_{r_c}$ has been computed:

$$E\ (S^2_{r_c}) \approx S^2_{e_0} + S^2_{e_1} + (1 - \overline{Y})^2\ S^2_{d_k} + \frac{1}{2}\ (\frac{\partial^2 S^2_{r_c}}{\partial\ Y^2})\ S^2_Y$$

$$= S^2_{e_0} + S^2_{e_1} + [(1 - \overline{Y})^2 + S^2_Y]\ S^2_{d_k}$$

(2.5-19)

Using the values computed for the various data groups, a linear regression of mean squared residual on the variable $(1 - \overline{Y})^2 + S^2_Y$ has been done. The results (Figure 2.5-5) indicate a high correlation between these two variables

Figure 2.5-5

Outlet Concentration Residual Variance for

Various Residence Time Intervals

$$S_{r_c}^2 = S_m^2 + [(1 - \bar{Y})^2 + s_Y^2] \; S_{d_k}^2$$

$$S_m^2 = .018 \pm .010$$

$$S_{d_k}^2 = .299 \pm .038$$

$$r = .977$$

$S_{r_c}^2$ = Mean Squared Residual in Estimating ln(C0)

$(1 - \bar{Y})^2 + s_Y^2$
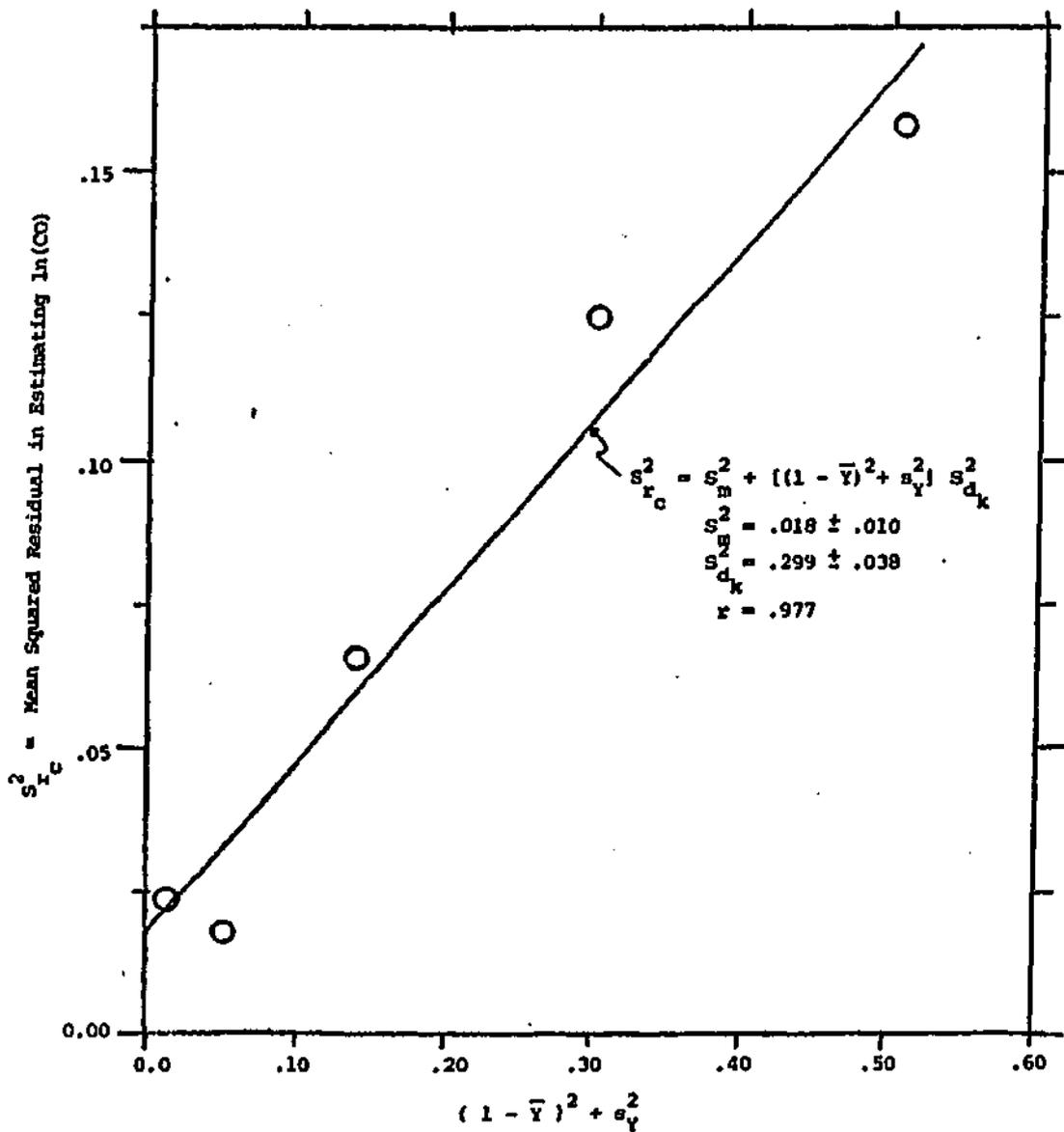
2.5.3 Error Analysis (continued)

(r = 0.977). The least squares estimates of the slope and
intercept are given by:

$$\text{slope} = s_{d_k}^2 = 0.299 \pm 0.038$$

$$\text{intercept} = s_{e_1}^2 + s_{e_0}^2 = 0.0178 \pm 0.010$$

The slope, or model error, is well-determined relative to the
intercept, which is significantly different from zero at between
the 80% and 90% confidence levels. Thus, according to this
analysis, the residual variance of Y and ln (CO) are given
by:

$$s_{r_y}^2 = 0.018 \ Y^2 + 0.30 \ Y^2 \ (1 - Y)^2 \qquad (2.5-20)$$

$$s_{r_c}^2 = 0.018 + 0.30 \ (1 - Y)^2 \qquad (2.5-21)$$

In each equation, the first term represents measurement error,
and the second, model error.

Figures 2.5-6 and 2.5-7 illustrate the abilities of these
equations to depict residual patterns in the data. The variance

Figure 2.5-6
Estimated Model and Measurement Error Components
of Outlet Concentration Residuals
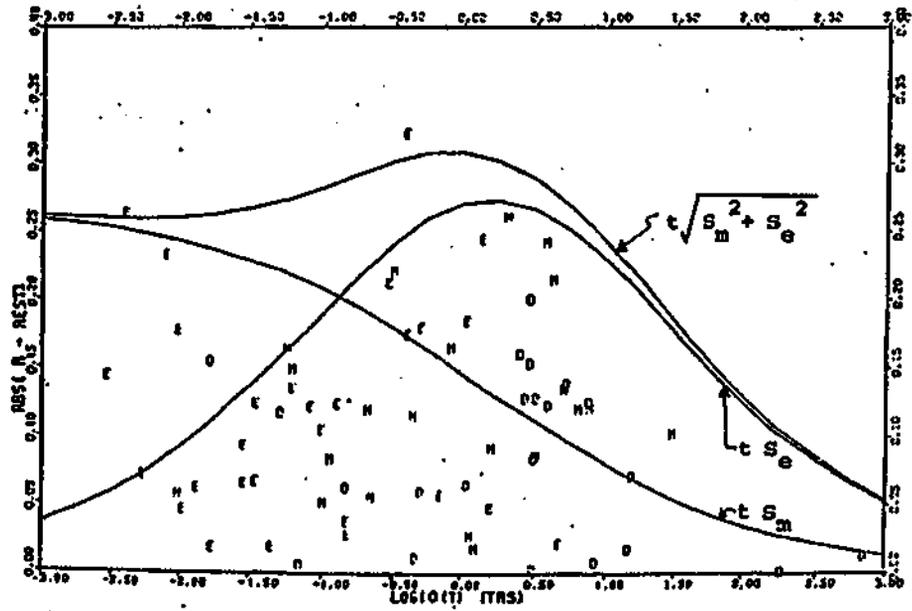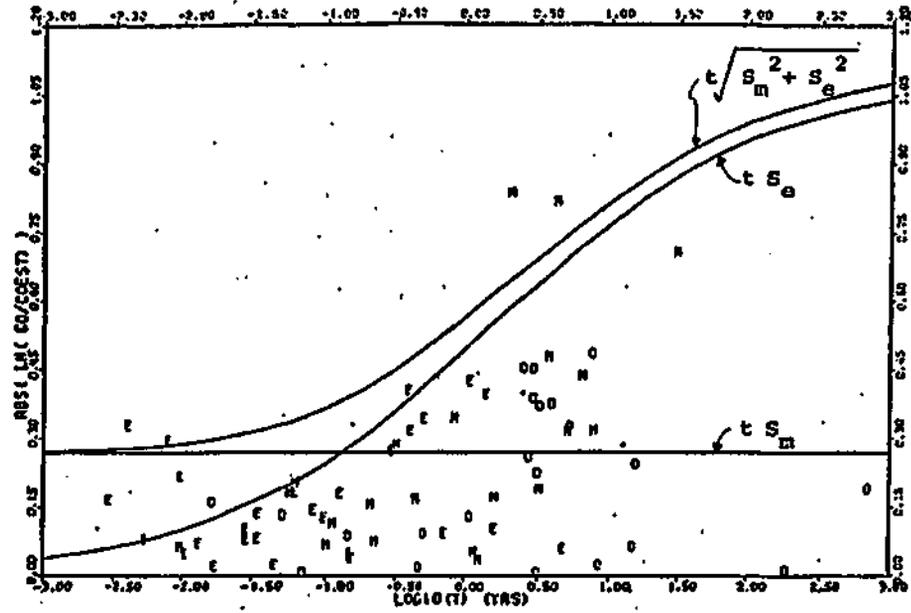
Figure 2.5-7
Estimated Model and Measurement Error Components
of Retention Coefficient Residuals

S_e = model error $\qquad$ $\sqrt{S_m^2 + S_e^2}$ = total error

S_m = measurement error $\qquad$ t = 2.0

2.5.3 <u>Error Analysis</u> (continued)

components have been multiplied by four, to represent approximate
95% confidence regions. These figures permit further interpretation
of the residuals as originating from either of the two sources of
error for various values of residence time. In both cases,
measurement error dominates at low values of T , while model error
dominates at high values. The best estimate of the measurement
error variance can be interpreted in terms of the coefficients of
variation of the inlet and outlet phosphorus concentration estimates:

$$s_{e_i}^2 + s_{e_o}^2 = 0.018 = CV_{ci_m}^2 + CV_{co_m}^2 \qquad (2.5-22)$$

for $CV_{ci_m} = CV_{co_m} = CV_m$,

$$CV_m = \sqrt{0.018/2} = 0.095 \approx 0.1 \qquad (2.5-23)$$

Thus, the estimated measurement error term corresponds roughly
to coefficients of variation on the order of 0.1. Based upon
the standard error the intercept, the 95% confidence range for
this statistic is from 0 to 0.15.

In applying the model developed above to predict the

2-115

2.5.3 Error Analysis (continued)

phosphorus retention coefficient for a given lake, three sources of error would be present -- model error, parameter error, and independent variable error:

$$Y = 1 - R = (1 + a T^{1-b})^{-1} \qquad (2.5\text{-}24)$$

$$S_Y^2 = S_e^2 + S_p^2 + S_i^2 \qquad (2.5\text{-}25)$$

Model error, S_e^2, is estimated from the above analysis:

$$S_e^2 = 0.30 \ Y^2 \ (1 - Y)^2 \qquad (2.5\text{-}26)$$

The second error term, S_p^2, is attributed to uncertainty in the parameter estimates:

$$S_p^2 = \left(\frac{\partial Y}{\partial a}\right)^2 S_a^2 + \left(\frac{\partial Y}{\partial b}\right)^2 S_b^2 + 2 \left(\frac{\partial Y}{\partial a}\right)\left(\frac{\partial Y}{\partial b}\right) S_a S_b \rho_{ab} \qquad (2.5\text{-}27)$$

2.5.3 <u>Error Analysis</u> (continued)

where,

$$a = 0.824$$

$$b = 0.546$$

$$\rho_{ab} = -0.0064$$

$$s_a = 0.067$$

$$s_b = 0.046$$

Evaluation of the expression gives:

$$s_p^2 = 0.001 \, Y^4 \, T^{.908} \, (4.49 + 1.44 \, \overline{\ln T}^2 + 0.032 \, \overline{\ln T}) \qquad (2.5-28)$$

The third term in equation (25), s_i^2, is associated with possible error in the estimate of the independent variable, T :

$$s_i^2 = \left(\frac{\partial Y}{\partial T}\right)^2 s_T^2$$

$$= 0.140 \, Y^4 \, T^{.908} \left(\frac{s_T}{T}\right)^2 \qquad (2.5-29)$$

$$= 0.140 \, Y^4 \, T^{.908} \, \overline{CV}_T^2$$

2.5.3 <u>Error Analysis</u> (continued)

These three error components are plotted against T on logarithmic scales in Figure 2.5-8 To evaluate the last term, $\overline{CV}_T$ is assumed to be 0.10, corresponding to an approximate 95% confidence region of ± 20% in the estimate of T . The plot shows clearly that the model error term dominates over the other terms by about one and two orders of magnitude, respectively, for all residence time values.

Further reduction of the parameter error term would be achieved by increasing the sample size, i.e., by adding additional lakes to the data set. However, if these lakes were drawn from the same general population, the effect on the total projection error would be small, since the parameter error term is already insignificant compared with the model term, which would not be expected to change much with sample size. Thus, for this type of data and model, sample size appears to have been adequate for parameter estimation purposes.

The dashed line in Figure 2.5-8 represents measurement error, estimated according to equation (20):

$$S_m^2 = 0.018 \ Y^2 \qquad\qquad (2.5\text{-}30)$$

As noted previously, the estimate of this variance component

Figure 2.5-8

Variance Components of Retention Coefficient Prediction

mean squared residual

model error $(S_{\bar{q}_k}^2 = .30)$

measurement error $(CV_{ci_m}^2 + CV_{co_m}^2 = .018)$

parameter error

residence time error $(CV_T = .10)$

S_r^2 = Variance of Predicted Value of $(1 - R)$

$$\text{total} = \text{model} + \text{parameter} + \text{residence time}$$
$$\text{error} \quad \text{error} \quad \text{error} \quad \text{error}$$

.01 .02 .05 .10 .20 .50 1.0 2.0 5.0 10. 20. 50. 100. 200.

Mean Hydraulic Residence Time (years)
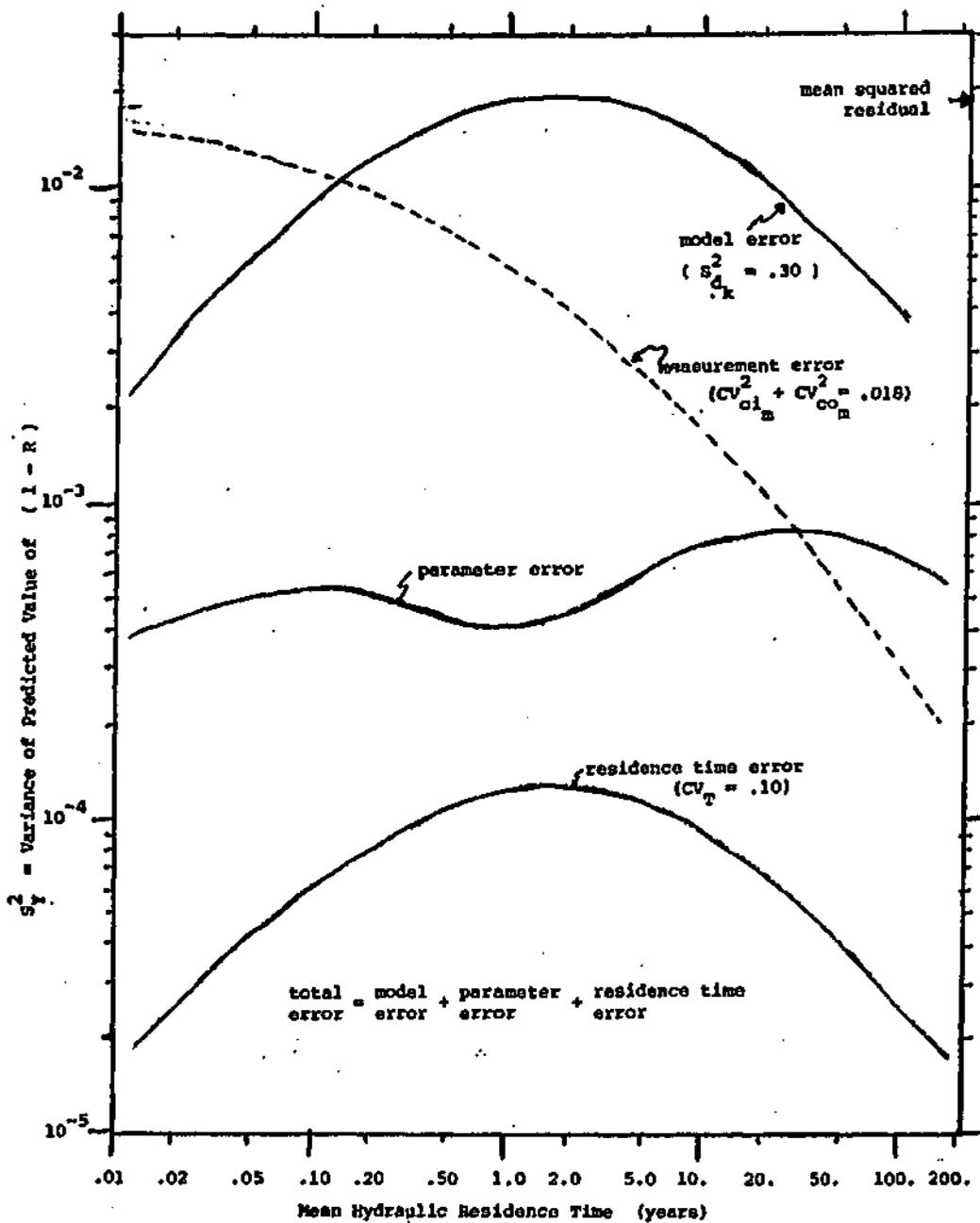
2.5.3 <u>Error Analysis</u> (continued)

is not very accurate, with a 95% confidence range of 0 to 0.045 Y^2 . In addition, it is specific to this data set, so that it would apply only in cases where the nutrient budgets are estimated using data of the same quantity and quality, and techniques similar to those employed in the National Eutrophication Survey. For lakes with residence times less than about 0.15 years, measurement errors dominate over other types. In this region, any errors in the model projection would be difficult to detect with this type of data.

If a projection of the retention coefficient were made for a given lake and subsequently compared with a measured value, model error would only be detectable if the following criterion were satisfied:

$$t = \frac{|Y_m - Y_e|}{S_m} \geq t_{\alpha,n} \qquad (2.5\text{-}31)$$

where,

Y_m = measured value of $1 - R$

Y_e = estimated value of $1 - R$

S_m = standard error of measured value

$t_{\alpha,n}$ = t statistic at significance level α and n degrees of freedom in Y_m estimate

2.5.3 Error Analysis (continued)

If this procedure were repeated for a large number of lakes at a given residence time, the first and second moments of t would be given by:

$$E(t) = 0 \qquad (2.5\text{-}32)$$

$$s_t^2 = 1 + \frac{s_e^2}{s_m^2} \qquad (2.5\text{-}33)$$

where

$$s_e^2 = \text{model error variance}$$

$$s_m^2 = \text{measurement error variance}$$

The fraction of lakes in the population for which equation (31) would be satisfied would be given by:

$$\beta = 1 - \int_{-t_{\alpha,n}}^{+t_{\alpha,n}} f_t \, dt \qquad (2.5\text{-}34)$$

2-121

2.5.3 Error Analysis (continued)

where,

β = fraction of lakes satisfying equation (31)

f_t = normal probability density function t with mean zero and standard deviation S_t

At a given residence time, the value of S_t^2 can be estimated from equations (33) and (21):

$$s_t^2 = 1 + \frac{s_e^2}{s_m^2} = 1 + \frac{(1-Y)^2 \, s_{d_k}^2}{s_{e_i}^2 + s_{e_o}^2} \approx 1 + 16.7 \, (1-Y)^2 \qquad (2.5-35)$$

Equations (34) and (35) have been evaluated as a function of hydraulic residence time, for $t_{\alpha,n} = t_{.05,\infty} \approx 2$, and for various values of the variance ratio $s_{d_k}^2 / (s_{e_i}^2 + s_{e_o}^2)$ ranging from 5 to 40. The results (Figure 2.5-9) depict the detectability of model errors as a function of residence time and model error to measurement error ratio. The nominal value of 16.7 for the latter represents the best estimate of the curve for data of the type used in estimating the model. At a given residence time, the fraction increases with variance ratio. As T approaches zero, model error vanishes and the

Figure 2.5-9

Detectability of Model Errors in Estimating the Retention Coefficient as a
Function of Residence Time and Model to Measurement Error Variance Ratio
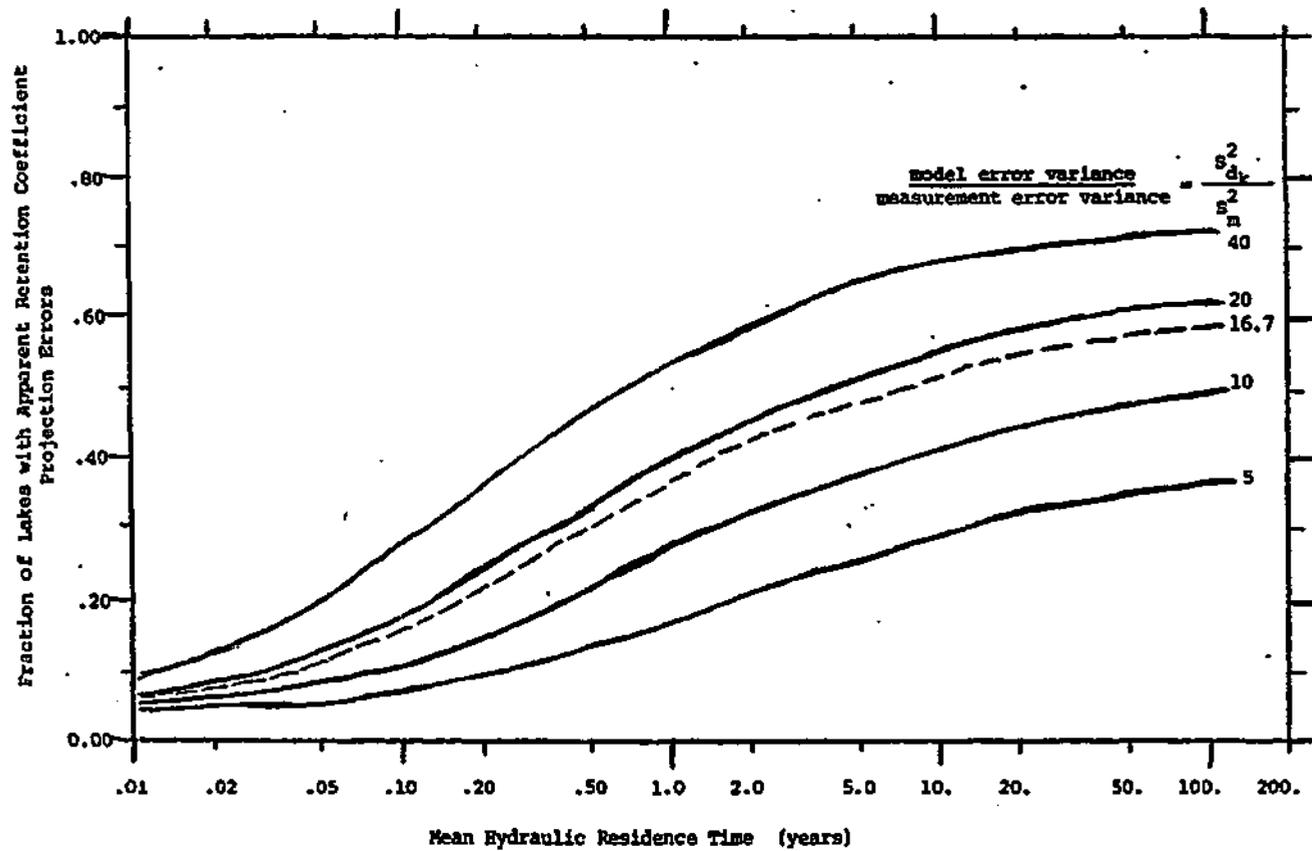
2.5.3 Error Analysis (continued)

statistic approaches 0.05, corresponding to the selected significance level of $t_{\alpha,n}$. All of the curves approach upper asymptotes as T increases, reflecting an increasing detectability of projection errors. The curve corresponding to the nominal variance ratio has been mapped onto the frequency distribution of residence times in the data set (Table 2.5-2). The result indicates that significant model errors would be detectable in approximately 31% of the lakes, if the data base were similar to that provided by the National Eutrophication Survey. In Figures 2.5-6 and 2.5-7, twenty-four residuals lie outside the estimated measurement error curves. This corresponds to an observed model error detection frequency of 34%.

In applying the model to predict outlet concentration, another error term must be added:

$$CO = CI\,Y$$

$$\frac{s_{co}^2}{CO^2} = \frac{s_{ci}^2}{CI^2} + \frac{s_Y^2}{Y^2}$$

$$CV_{co}^2 = CV_{ci_m}^2 + \frac{1}{Y^2}(s_e^2 + s_p^2 + s_i^2) \simeq CV_{ci_m}^2 + \frac{1}{Y^2}s_e^2$$

$$= CV_{ci_m}^2 + 0.30\,(1-Y)^2 \tag{2.5-36}$$

2.5.3 <u>Error Analysis</u> (continued)

Because of their relative insignificance, the parameter and
residence time error terms have been neglected. The added
term, $CV_{ci_m}^2$, accounts for uncertainty in the inlet concen-
tration, or phosphorus loading estimate.

Equation (36) has been plotted in Figure 2.5-10 for
various assumed values of CV_{ci_m} . The dashed line again
represents measurement error, assuming $CV_{co_m} = 0.10$, as
estimated above. The relative importance of the various
error components can be used as a basis for allocating
additional effort to model development or to monitoring
in the interest of reducing the variance of the projection.
Model error dominates over other various components at high
residence times. Accordingly, development and/or application
of a more sophisticated model may be justified in this region,
providing that its model and parameter errors are lower. As
residence time decreases, potential measurement errors domi-
nate, and the prediction becomes increasingly sensitive to
errors in the loading or inlet concentration estimate. At
lower residence times, this model would be considered adequate
and additional effort would be most effectively applied in
monitoring (or source modelling) in order to develop a better
loading estimate.

In a specific application, the variance of the inlet

2.5.3 Error Analysis (continued)

phosphorus concentration estimate can be calculated according
to the techniques discussed in Section 2.3.2, if the estimate
is based upon direct measurements. If such are not available,
the estimate can be based upon land use data, using the loading
factors in Table 2.3-1. The ranges of these loading factors
can be used to estimate variance. If a rectangular distribu-
is assumed, the standard deviation is given by 0.29 times the
range[77]. For the ranges and mean values for total phosphorus
loading given in Table 2.3-1, a coefficient of variation of
about 1 is indicated. Comparison with the other error components
in Figure 2.5-10 indicates that the accuracy of an outlet phos-
phorus concentration projection would be limited most by
potential errors in such indirect loading estimates. This
type of error could presumably be reduced by considering some
of the other factors influencing nutrient export discussed
in Section 2.3.1. However, the current state-of-the-art of
watershed models is such that the accuracy of the lake model
derived above would probably not limit the accuracy of an
outlet phosphorus concentration projection. Thus, the model
would be considered adequate for use with indirect loading
estimates.

A final aspect to consider is the potential errors
involved in predicting chlorophyll concentration in lakes
based upon predictions of the phosphorus retention model.

2.5.3 <u>Error Analysis</u> (continued)

Dillon and Rigler[14] have examined the relationship between measured total phosphorus concentration at spring overturn and mean summer chlorophyll-a concentration based upon data from 19 lakes in southern Ontario and from 27 other lakes in North America, as reported in the literature. All of these lakes were phosphorus-limited systems, with total nitrogen to total phosphorus ratios greater than 15 at spring overturn. They fitted the following regression line to their data (Figure 2.5-11) :

$$\ln (Chl) = 7.393 + 1.449 \ln (CS) \qquad (2.5\text{-}37)$$

$$r = 0.95$$

$$SEE = 1.01$$

where,

Chl = mean summer chlorophyll-a
concentration (mg/m^3)

CS = total phosphorus concentration
of spring overturn (g/m^3)

SEE = standard error of estimate
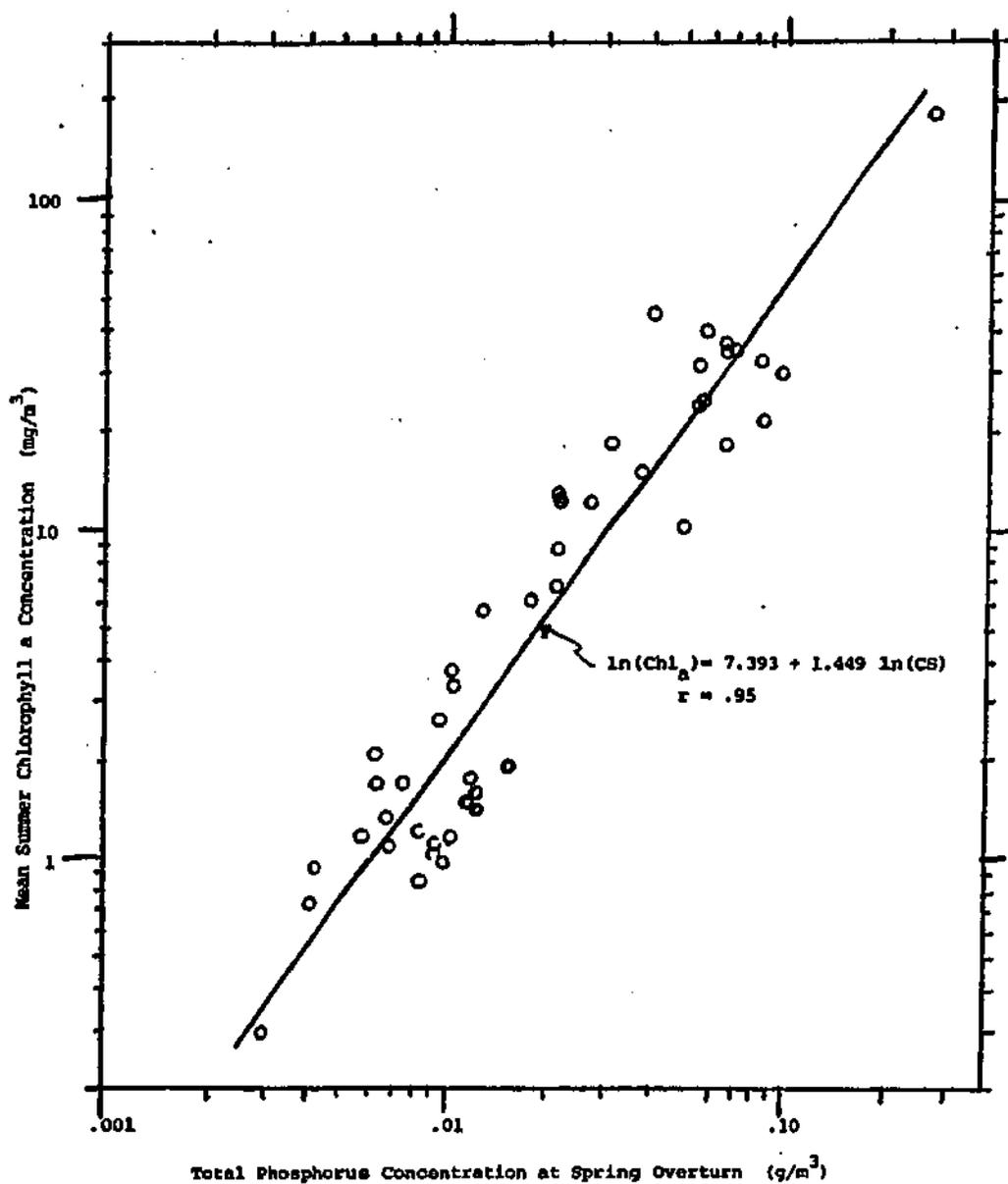
Figure 2.5-11

Relationship Between Mean Summer Chlorophyll a and
Total Phosphorus Concentration at Spring Overturn
According to Dillon and Rigler[14]

2.5.3 Error Analysis (continued)

Coupling of this model with the phosphorus retention model would permit estimation of chlorophyll concentrations from loading, hydraulic residence time, and depth estimates.

In order to do this, a means of relating the spring overturn to mean outlet phosphorus concentration would be required. It is assumed that, if the data were compiled, a model would be estimated for this purpose. Independent variables in such a model would probably include mean outlet concentration, inlet concentration, residence time, and possibly some indicator of seasonal flow variations. Alternatively, the chlorophyll equation above could be re-estimated based upon outlet phosphorus concentration. At spring overturn, lakes are completely mixed, so that $CO = CS$ would not be a bad assumption during this period. Since flows are generally highest during this period, the annual flow-weighted-average outlet concentration would be a strong function of the spring concentration, and vice-versa. For the sake of the following arguments, it is assumed that the spring overturn concentration can be equated to the average outlet concentration.

This permits evaluation of potential errors involved in predicting summer chlorophyll from loading, residence time, and depth estimates:

2.5.3 Error Analysis (continued)

$$\ln (Chl) = 7.393 + 1.449 \ln (CO) \qquad (2.5\text{-}38)$$

$$s^2_{\ln (Chl)} = s^2_i + s^2_p + s^2_r \approx \overline{cv}^2_{Chl} \qquad (2.5\text{-}39)$$

The independent variable error in this case is calculated from the total error in predicting the outlet phosphorus concentration (equation (36)):

$$s^2_i \approx \left(\frac{\partial \ln (Chl)}{\partial \ln (CO)}\right)^2 s^2_{k(CO)} \approx 1.449^2 \, cv^2_{co} = 2.10 \, cv^2_{co} \qquad (2.5\text{-}40)$$

$$= 2.10 \, (cv^2_{ci_m} + 0.30 \, (1 - Y)^2) \qquad (2.5\text{-}41)$$

This error results from uncertainty in the inlet concentration estimate and from model error. The parameter error is due to uncertainty in the slope and intercept of the regression line in Figure 2.5-11 :

$$s^2_p = \overline{SEE}^2 \left(\frac{1}{n} + \frac{[\ln CS) - \overline{\ln (CS)}]^2}{n \, s^2_{\ln (CS)}} \right) \qquad (2.5\text{-}42)$$

For values of $\overline{SEE}^2$, n , $\overline{\ln (CS)}$, and $s^2_{\ln(CS)}$ derived from

2.5.3 Error Analysis (continued)

the Dillon and Rigler article[14], and assuming that $CO \approx CS$:

$$s_p^2 = 1.02 \left(\frac{1}{46} + \frac{[\ln (CO) + 3.950]^2}{46 \ (1.016)} \right)$$

$$\qquad (2.5-43)$$

$$= 0.022 \left(1 + \frac{[\ln (CO) + 3.950]^2}{1.016} \right)$$

The residual error term, also derived from the Dillon and Rigler article is given by:

$$s_r^2 = \overline{SEE}^2 = 1.02 \qquad (2.5-44)$$

The four terms of the total chlorophyll prediction error are plotted versus hydraulic residence time in Figure 2.5-12. The independent variable term, s_i^2 , has been evaluated according to equation (41), assuming $CV_{ci_m} \approx 0.1$. The parameter error term has been evaluated for various values of CO .

Figure 2.5-12 shows clearly that the residual error term dominates over phosphorus prediction and parameter error terms for all values of hydraulic residence time. It is unlikely that the negative distribution of the error components would change

Figure 2.5-12

Variance Components of Mean Summer Chlorophyll a Predictions
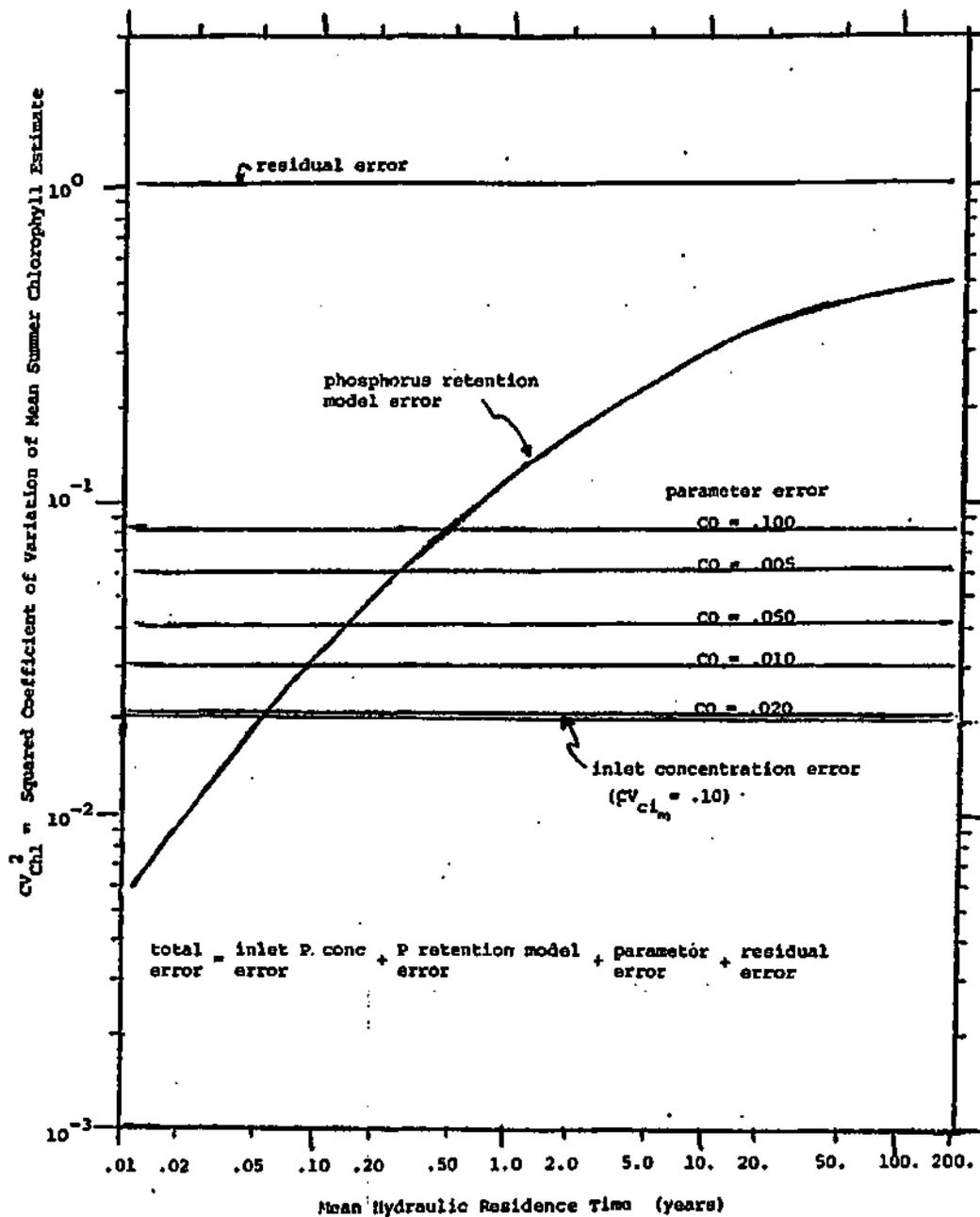
2.5.3 Error Analysis (continued)

much if equation were re-estimated in terms of CO instead of

CS , i.e. this distribution is probably not very sensitive

to the assumption that CS = CO . In addition, the Dillon and

Rigler equation was derived from a fairly narrow range of lakes.

An analysis of data from the National Eutrophication Survey lakes[87]

indicates higher standard error of estimate in the chlorophyll-

phosphorus relationship than that found by Dillon and Rigler.

These results indicate that the limiting factor in the

accuracy of a chlorophyll prediction would probably be in the

chlorophyll-lake phosphorus relationship and not in the lake

or outlet concentration prediction. Thus, the phosphorus

retention model could be considered adequate in this

application, and future modelling efforts would best be

focused upon the chlorophyll-phosphorus relationship. A

chlorophyll prediction would be useful in that it could be

related directly to criteria and to such trophic state

indicators as productivity, transparency, and hypolimnic

dissolved oxygen deficit.

2.5.4 Trophic State Prediction

The next step in the analysis is to examine the feasibility of predicting lake trophic state from the kinds of data described above. The models of Vollenweider[93,95] and Dillon[16,17] will serve as a primary basis for the analysis. Each of these models is viewed as a set of coordinate axes. Parameter estimates define the locations of the lines delimiting trophic state. The goal of this work is to evaluate both the models and their respective parameter estimates. Accordingly, the ability of each model to predict trophic state has been examined using two sets of parameter values: that originally proposed by the respective authors and an "optimal" set estimated for this data base. Optimal parameter values have been found by posing the classification problem as a linear discriminant analysis and utilizing the BMDP7M discriminant analysis program[26] to determine the parameter estimates which best separate the trophic states on the coordinate axes defined by each model.

Vollenweider's first and second models are essentially bi-variate classification functions employing L and Z , L and QS , respectively, while Dillon's model is a univariate function employing CO . In estimating optimal coefficients for these models, the corresponding variables have been forced into the classification scheme and all other variables have been excluded. Using the stepwise feature of the BMDP7M program, an optimal linear classification model has also been estimated. In the

2.5.4 Trophic State Prediction (continued)

stepwise mode, the program selects the variables which provide the most discrimination power in order of decreasing power, until some specified lower level of significance is reached. Reckhow[68] has employed stepwise discriminant analysis in developing a model for classifying oxic and anoxic lakes.

Homogeneity of variance across groups is one of the principal assumptions of the analysis, when pooled estimates of the covariance matrix of the entered variables are employed[77]. In order to eliminate the apparent skewness in the distribution of the log-transformed average outlet concentration data for the eutrophic group (Table 2.5-2) and to promote homogeneity of variance across groups, lakes with average outlet concentratione greater than 0.2 g/m^3 have been excluded from the discriminant analysis. These lakes are highly eutrophic and would not be misclassified by any of the models. The analysis have been performed on the remaining 22 oligotrophic, 23 mesotrophic, and 50 eutrophic lakes.

Table 2.5-7 summarizes the group means and standard deviations for each of the variables in the analysis. The definitions and transformations of these variables have been previously defined in Table 2.5-2. The variable "COEST" is the average outlet concentration, as estimated using the model for phosphorus retention derived previously:

Table 2.5-7

Summary of Distributions and Initial Discriminating Powers of Variables in Discriminant Analysis

MEANS				
GROUP = VARIABLE	OLIG	MESO	EUTR	ALL GPS.
3 Z	1.25843	1.04731	0.41312	0.97334
4 T	0.38777	0.07902	-0.48816	-0.26652
5 L	-0.83315	-0.44404	0.46511	-0.04512
6 CO	-2.12684	-1.81557	-1.31076	-1.62197
7 CI	-1.70923	-1.46233	-1.01617	-1.28468
8 COEST	-2.11365	-1.77567	-1.21675	-1.55977
9 QS	0.87608	1.01829	1.50128	1.23956
10 TS	1.00000	2.00000	3.00000	2.29474
COUNTS	22.	23.	50.	95.

STANDARD DEVIATIONS				
GROUP = VARIABLE	OLIG	MESO	EUTR	ALL GPS.
3 Z	0.44866	0.47320	0.44015	0.45019
4 T	1.07009	0.94547	1.08116	1.04772
5 L	0.48175	0.56076	0.62004	0.57730
6 CO	0.15507	0.26253	0.26423	0.24322
7 CI	0.31525	0.36882	0.33414	0.33854
8 COEST	0.17629	0.24098	0.26531	0.24131
9 QS	0.74836	0.72228	0.77709	0.75777
10 TS	0.0	0.0	0.0	0.0

●●●

STEP NUMBER 0							
VARIABLE	F TO FORCE REMOVE LEVEL DF= 2 93		●	VARIABLE	F TO FORCE ENTER LEVEL DF= 2 52		TOLERANCE
			●	3 Z	7.901	1	1.000000
			●	4 T	9.187	1	1.000000
			●	5 L	47.126	1	1.000000
			●	6 CO	95.614	1	1.000000
			●	7 CI	36.171	1	1.000000
			●	8 COEST	117.140	1	1.000000
			●	9 QS	6.494	1	1.000000

2.5.4 Trophic State Prediction (continued)

$$COEST = (1 - RE) \ CI \qquad (2.5\text{-}45)$$

$$1 - RE = \frac{1}{1 + 0.824 \ T^{.454}} \qquad (2.5\text{-}46)$$

The base 10 logarithmic transformation has been applied to COEST.

The second part of Table 2.5-7 summarizes the initial discrimination power of the variables. The F statistics represent univariate analysis of variance statistics, analogous to those found in Table 2.5-2. Table 2.5-7 indicates that COEST had the highest discrimination power, followed by CO , L , CI , T , Z , and QS . The F values were all significant at the 95% level or greater. As discussed previously, however, this does not indicate that there were seven independent and significant factors, because of intercorrelations among these variables (Table 2.5-3).

Tables 2.5-8 to 2.5-12 present the results of the discriminant analyses. In each case, the table contains the significance levels of the entered and remaining variables, summary statistics characterizing the degree of group separation, classification functions, classification matrix, and a jackknifed classification

2.5.4 Trophic State Prediction (continued)

matrix. Table 2.5-13 summarizes and compares the statistics

obtained for the various models.

To estimate optimal parameters for Vollenweider's first

model, Z and L have been forced into the classification

scheme (Table 2.5-8). The low F level for Z (0.28) indicates

that it did not add significantly to the classification power

in L alone (F = 33.6). Appreciable power remained, however,

in the excluded variables, which had F levels ranging from

33 to 48, after both L and Z had entered. Thus, this model

apparently did not make use of all of the available discriminating

power in the data. Wilk's lambda and the overall F are multi-

variate analysis of variance statistics[12], the interpretation of

which will be discussed later. An F matrix characterizes the

separation of each of the three possible pairs of groups and

indicates that the distinction between oligotrophic lakes and

eutrophic lakes was relatively high (F = 39.6), while that

between mesotrophic and eutrophic lakes was fairly low (F = 2.8).

The classification functions correspond to the parameter

estimates. Each function consists of a constant plus a coefficient

for each variable. In classifying a given observation, a function

value is computed for each group, and the observation is estimated

to have the highest probability of belonging to the group with

the highest corresponding function value. In order to derive

2.5.4 Trophic State Prediction (continued)

the lines delimiting the mesotrophic from the eutrophic states
on the L versus Z axes, for example, the corresponding
classification functions would be subtracted from one another
and solved for L as a function of Z .

The classification matrix summarizes the ability of the
model to classify the lakes correctly. Observed states cor-
respond to rows, and estimated states, to columns. If the model
worked perfectly, all of the lakes would be on the upper left
to lower right diagonal. In Table 2.5-8, 72.7%, 52.2%, and
78.0% of the oligotrophic, mesotrophic, and eutrophic lakes,
respectively, were classified correctly. The "jackknifed"
classification matrix provides a somewhat more robust estimate
of the probability of correctly classifying a lake. To derive
this matrix, each case (lake) was excluded from the analysis one
lake at a time, the classification functions were re-calculated,
and the excluded lake was classified based upon the derived
functions. Thus, the overall percent correct for the jackknifed
classifications, 66.3%, is a measure of the probability of
correctly classifying a new lake which has not been used to
estimate the parameters of the discriminant model.

Vollenweider's second model (Table 2.5-9) compared favorably
with his first. Both L and QS entered significantly, although
some discrimination power still remained in the excluded variables.

Table 2.5-8

Table 2.5-9

Discriminant Analysis - Vollenweider's First Model

Discriminant Analysis - Vollenweider's Second Model

Table 2.5-8 (left)

STEP NUMBER 2
VARIABLE ENTERED 3 2

VARIABLE	F TO FORCE REMOVE LEVEL		VARIABLE	F TO FORCE ENTER LEVEL		TOLERANCE
	DF= 2 91			DF= 2 90		
3 2	0.280	1	4 T	41.240	0	0.084709
5 L	33.595	1	6 CO	37.516	0	0.043777
			7 CI	41.246	0	0.307022
			8 COAST	49.432	0	9.374414
			9 OS	41.248	0	0.161172

U-STATISTIC OR WILKS LAMBDA 0.4909301   DEGREES OF FREEDOM 2 2 92
APPROXIMATE F-STATISTIC 19.438   DEGREES OF FREEDOM 4.00 182.00

F - MATRIX   DEGREES OF FREEDOM = 2 91

	OLIG	MESO
MESO	2.82	
EUTR	39.55	20.29

CLASSIFICATION FUNCTIONS

VARIABLE	GROUP = OLIG	MESO	EUTR
3 2	2.84094	3.27044	5.54940
5 L	-0.75362	0.25045	3.11757
CONSTANT	-3.08075	-3.81341	-6.10934

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP -		
		OLIG	MESO	EUTR
OLIG	72.7	16	4	2
MESO	52.2	8	12	3
EUTR	78.0	1	10	39
TOTAL	70.5	25	26	44

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP -		
		OLIG	MESO	EUTR
OLIG	72.7	16	4	2
MESO	39.1	10	9	4
EUTR	76.0	1	11	38
TOTAL	66.3	27	24	44

Table 2.5-9 (right)

STEP NUMBER 2
VARIABLE ENTERED 9 OS

VARIABLE	F TO FORCE REMOVE LEVEL		VARIABLE	F TO FORCE ENTER LEVEL		TOLERANCE
	DF= 2 91			DF= 2 90		
5 L	101.010	1	3 2	2.800	0	0.768844
9 OS	37.090	1	4 T	2.808	0	0.161958
			6 CO	3.427	0	0.579624
			7 CI	0.697	0	0.003000
			8 COAST	7.305	0	0.211177

U-STATISTIC OR WILKS LAMBDA 0.2721270   DEGREES OF FREEDOM 2 2 92
APPROXIMATE F-STATISTIC 41.722   DEGREES OF FREEDOM 4.00 182.00

F - MATRIX   DEGREES OF FREEDOM = 2 91

	OLIG	MESO
MESO	6.11	
EUTR	105.00	49.65

CLASSIFICATION FUNCTIONS

VARIABLE	GROUP = OLIG	MESO	EUTR
5 L	-24.18817	-19.26869	-9.23694
9 OS	18.19509	15.06960	9.00196
CONSTANT	-19.13276	-13.64922	-5.61044

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP -		
		OLIG	MESO	EUTR
OLIG	86.4	19	3	0
MESO	60.9	6	14	3
EUTR	84.0	0	8	42
TOTAL	78.0	25	25	45

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP -		
		OLIG	MESO	EUTR
OLIG	81.8	18	4	0
MESO	56.5	7	13	3
EUTR	84.0	0	8	42
TOTAL	76.0	25	25	45

2.5.4 Trophic State Prediction (continued)

The F statistics and Wilk's lambda indicate that the trophic states were separated considerably better on the L versus QS axes, as compared with the L versus Z axes. Overall, the model classified 78.9% of the lakes correctly, an increase of 8.4% over the first model. For the oligotrophic group, the percent correct increased from 72.7% to 86.4%, corresponding to an approximate halving in the percent misclassifications.

Table 2.5-10 indicates that Dillon's model generally performed somewhat less well than Vollenweider's second model, but better than his first. Significant discrimination power remained in COEST (F = 9.05) and in L (F = 6.25), after CO had entered. The mesotrophic lakes were classified poorly relative to either of the two previous models. The overall percent correct was 73.7%. As in the case of Vollenweider's second model, but not of his first, no eutrophic lakes were misclassified as oligotrophic or vice-versa.

In the stepwise analysis (Table 2.5-11), both COEST and L entered at F levels of 49.0 and 8.4, respectively, leaving no significant discrimination power in the remaining variables. Thus, these two variables alone constitute an optimal linear classification model for these lakes and data. As expected, this model performed better than any of those examined previously. 84.2% of the lakes were classified correctly, with most of the

Table 2.5-10

Discriminant Analysis – Dillon's Model

| STEP NUMBER 1 | | | | | | |
| VARIABLE ENTERED 6 CO | | | | | | |

VARIABLE	F TO FORCE REMOVE LEVEL DF= 1 92	*	VARIABLE	F TO FORCE ENTER LEVEL DF= 2 91		TOLERANCE
6 CO	95.614 1	*	3 Z	0.086	0	0.930412
		*	4 T	1.184	0	0.986442
		*	5 L	6.250	0	0.922725
		*	7 CI	0.916	0	0.784617
		*	8 COEST	9.065	0	0.444114
		*	9 QS	2.204	0	0.999982

U-STATISTIC OR WILKS' LAMBDA 0.3249276 DEGREES OF FREEDOM 1 2 92
APPROXIMATE F-STATISTIC 95.613 DEGREES OF FREEDOM 2.00 92.00

F - MATRIX DEGREES OF FREEDOM = 1 92

	OLIG	MESO
MESO	18.42	
EUTR	172.00	67.46

CLASSIFICATION FUNCTIONS

GROUP = VARIABLE	OLIG	MESO	EUTR
6 CO	-35.95340	-30.69153	-22.15797
CONSTANT	-36.33212	-28.95992	-15.62055

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - OLIG	MESO	EUTR
OLIG	86.4	19	3	0
MESO	43.5	9	10	3
EUTR	82.0	0	9	41
TOTAL	73.7	27	22	44

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - OLIG	MESO	EUTR
OLIG	86.4	19	3	0
MESO	43.5	9	10	3
EUTR	82.0	0	9	41
TOTAL	73.7	27	22	44

Table 2.5-11

Discriminant Analysis – Stepwise

| STEP NUMBER 2 | | | | | | |
| VARIABLE ENTERED 5 L | | | | | | |

VARIABLE	F TO FORCE REMOVE LEVEL DF= 2 91	*	VARIABLE	F TO FORCE ENTER LEVEL DF= 2 90		TOLERANCE
5 L	8.613 1	*	3 Z	0.299	1	0.860772
8 COEST	48.974 1	*	4 T	0.371	1	0.247567
		*	6 CO	0.371	1	0.412862
		*	7 CI	0.633	1	0.192632
		*	9 QS	0.636	1	0.030441

U-STATISTIC OR WILKS' LAMBDA 0.2378966 DEGREES OF FREEDOM 2 2 92
APPROXIMATE F-STATISTIC 67.747 DEGREES OF FREEDOM 4.00 182.00

F - MATRIX DEGREES OF FREEDOM = 2 91

	OLIG	MESO
MESO	12.18	
EUTR	128.26	55.87

CLASSIFICATION FUNCTIONS

GROUP = VARIABLE	OLIG	MESO	EUTR
5 L	-0.61977	0.63086	2.11229
8 COEST	-36.01045	-30.51033	-21.70268
CONSTANT	-39.33014	-28.09106	-14.95984

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - OLIG	MESO	EUTR
OLIG	90.9	20	2	0
MESO	73.9	4	17	2
EUTR	86.0	0	7	43
TOTAL	84.2	24	26	45

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - OLIG	MESO	EUTR
OLIG	90.9	20	2	0
MESO	73.9	4	17	2
EUTR	86.0	0	7	43
TOTAL	84.2	24	26	45

2.5.4 Trophic State Prediction (continued)

improvement in the mesotrophic group.

As a final step in the discriminant analysis, the BMDP7M program performs a principal component analysis of the classification functions. The principal components are linear functions of the entered variables which are independent of each other. The components are determined so as to maximize the variance (or discrimination power) of the first component, or canonical variable. Canonical correlation coefficients computed for all principal components characterize their relative discriminating powers.

The stepwise analysis (Table 2.5-11) indicated that the rate of phosphorus input, L , in addition to the estimated mean outlet concentration, COEST, were the best variables to use in a linear model to classify the lakes. These two variables are obviously correlated, since the latter was calculated from:

$$COEST = (1 - RE) \ CI = (1 - RE) \ \frac{LT}{2} \qquad (2.5\text{-}47)$$

The two principal components of these variables had canonical correlation coefficients of 0.87 and 0.07 , respectively. This indicated that essentially all of the classification ability was located in the first canonical variable. This result

2.5.4 Trophic State Prediction (continued)

effectively reduced the bivariate classification model to a univariate one. The analysis suggested that the lakes could best be classified along the axis of the following canonical variable:

$$\log X = 0.1853 \log L + 0.8147 \log COEST \qquad (2.5-48)$$

$$= 0.1853 \log L + 0.8147 \log [L(1-RE)/QS] \qquad (2.5-49)$$

$$= \log L - 0.8147 \log [QS/(1-RE)] \qquad (2.5-50)$$

The BMDP7D program[26] was again employed to examine the stratification of the trophic states along this axis. The analysis of variance indicated that the logarithm of X had more discrimination power ($F = 119$) than any of the variables previously examined (Table 2.5-2). However, the standard deviations of $\log_{10} X$ in the three groups were 0.16, 0.23, and 0.37, respectively. Bartlett's test[77] indicated that the hypothesis of homogeneity of variance could be rejected at the 99.5% level. In order to stabilize the variance across groups, the following transformation was employed:

$$XT = -1/X^{.25} \qquad (2.5-51)$$

$$X = L [QS/(1-RE)]^{-.8147} \qquad (2.5-52)$$

2.5.4 Trophic State Prediction (continued)

The transformed variable had standard deviations of 0.28, 0.34,
and 0.33, respectively. Bartlett's test indicated that the
hypothesis of homogeneity of variance across groups could not
be rejected at the 75% level. When the 10 lakes with outlet
concentrations greater than 0.2 g/m^3 were excluded from the
analysis, the transformation in equation (51) was still found
to be necessary to stabilize the variance across groups.

The discriminant analysis program was re-run, allowing XT
alone to enter. Table 2.5-12 shows that no significant discrimi-
nation power remained in the excluded variables. 86.3% of the
lakes were classified correctly along this axis. The analysis
of variance in the second part of Table 2.5-12 illustrates the
distribution of this variable across groups, using the entire
data set. The overall F level of 162.3 oan be compared with
the levels for the original variables, which ranged from 3.3 to
66.2 (Table 2.5-2). It seemed apparent that this canonical
variable incorporated essentially all of the linear discriminating
power of the variables in the analysis.

Table 2.5-13 summarizes and compares the statistics obtained
for the various discriminant runs. An important statistic
characterizing the overall separation of the groups is Wilk's
lambda, the multivariate analysis of variance statistic, which
has a possible range of zero to one[12]. The analog of lambds in

Table 2.5-12

Discriminant Analysis – Canonical Variable Derived from Stepwise Analysis[a]

$$a - XT = - X^{-.25}$$

$$X = L [QS/(1-RE)]^{-.815}$$

Table 2.5-13
Summary of Statistics Characterizing Discriminant Analyses

Model	Variables	Wilk's Lambda, Λ	Degrees of Freedom	$1-\Lambda$ [a]	Overall F	Degrees of Freedom	F Matrix O/M	M/E	O/E	Degrees of Freedom	Total Percent Misclassified	Average Percent Misclassified
Vollenweider I	L,Z	0.491	2,2,92	0.509	19.4	4,182	2.8	20.3	39.6	2,91	29.5	32.4
Vollenweider II	L,QS	0.272	2,2,92	0.728	41.7	4,182	8.1	49.7	105.0	2,91	21.1	22.9
Dillon	CO	0.325	1,2,92	0.675	95.6	2,92	18.4	67.9	172.0	1,92	26.5	29.4
Stepwise Model	L,COEST	0.238	2,2,92	0.762	47.8	4,182	12.2	54.4	128.2	2,91	15.8	16.4
Stepwise, First Canonical Variable	X [b]	0.225	1,2,92	0.775	158.4	2,92	37.3	102.1	291.2	1,92	14.7	15.1

[a] Analog of R^2 in a multiple regression analysis.

[b] $X = L[QS/(1-RE)]^{.815}$

2.5.4   Trophic State Prediction (continued)

the univariate case is the within-group sum of squares over
the total sum of squares. The smaller the value of lambda,
the greater the separation of the groups. One minus lambda,
also listed in Table 2.5-13 is an analog of R^2 in a multiple
regression analysis, in the sense that it is the multivariate
version of the explained sum of squares over the total sum of
squares. The "Total Percent Misclassified" has been computed
from the total number of misclassifications and the total number
of lakes, while, the "Average Percent Misclassified" represents
the average of the percents misclassified for each group. The
latter effectively normalizes the statistic for differences in
the numbers of samples across groups. The statistics in Table
2.5-13 provide a means of ranking the various models. They
generally indicate that the L versus QS axes (Vollenweider's
second model) provide the most discrimination power of the three
established models, while the functions derived from the stepwise
analysis are superior overall.

Using the entire sample of lakes, the functions derived
above and Vollenweider's and Dillon's models, with original and
optimal parameters, are compared in Table 2.5-14 and displayed
in Figures 2.5-13 through 2.5-18. For each model and parameter
set, Table 2.5-14 gives the loading equation used to classify
the groups, the classification matrix, and percentage misclassi-
fications by group and overall. Two forms of Dillon's model have

Table 2.5-14

Comparisons of Model Performance in Classifying 105 Lakes

Model / Loading Equations	Observed Trophic State	Estimated Trophic State O	M	E	Percent Error	Total Percent Error	Average Percent Error
Vollenweider I[a] $L_{OM} = 0.025\ Z^{0.602}$ $L_{ME} = 0.050\ Z^{0.602}$	O M E	12 3 0	5 7 1	5 13 59	45.5 69.6 1.7	25.7	38.9
Vollenweider I[b] $L_{OM} = 0.054\ Z^{0.548}$ $L_{ME} = 1.268\ Z^{-0.089}$	O M E	16 8 1	4 12 12	2 3 47	27.3 47.8 21.7	28.6	32.3
Vollenweider II[a] $L_{OM} = 0.100\ QS^{0.500}$ $L_{ME} = 0.200\ QS^{0.500}$	O M E	21 8 0	1 11 4	0 4 57	4.5 52.2 6.7	16.2	21.1
Vollenweider II[b] $L_{OM} = 0.057\ QS^{0.639}$ $L_{ME} = 0.181\ QS^{0.606}$	O M E	19 6 0	3 14 8	0 3 52	15.6 39.1 13.3	19.0	22.0
Dillon (R Observed)[a] $L_{OM} = 0.010\ QS/(1-R)$ $L_{ME} = 0.020\ QS/(1-R)$	O M E	18 7 0	4 9 2	0 7 58	18.2 60.3 3.3	18.9	27.3
Dillon (R Observed)[b] $L_{OM} = 0.0107\ QS/(1-R)$ $L_{ME} = 0.0274\ QS/(1-R)$	O M E	19 8 0	3 10 9	0 5 51	13.6 56.5 15.0	23.8	28.3
Dillon (R Estimated)[a] $L_{OM} = 0.010\ QS/(1-RD)$ $L_{ME} = 0.020\ QS/(1-RD)$	O M E	20 8 0	2 8 3	0 7 57	9.1 65.2 5.0	19.0	26.4
Dillon (R Estimated)[b] $L_{OM} = 0.0107\ QS/(1-RD)$ $L_{ME} = 0.0274\ QS/(1-RD)$	O M E	22 9 0	0 11 6	0 3 54	0.0 52.2 10.0	17.1	20.7
Stepwise Discriminant Analysis $L_{OM} = 0.0170\ [QS/(1-RE)]^{0.866}$ $L_{ME} = 0.0655\ [QS/(1-RE)]^{0.794}$	O M E	20 4 0	2 17 7	0 2 53	9.1 26.1 11.7	14.3	15.6
Stepwise Discriminant Analysis[c] $L_{OM} = 0.0190\ [QS/(1-RE)]^{0.815}$ $L_{ME} = 0.0549\ [QS/(1-RE)]^{0.815}$	O M E	20 4 0	2 17 5	0 2 55	9.1 26.1 8.5	12.4	14.5

[a]Original parameters.

[b]Optimal parameters.

[c]First canonical variable.
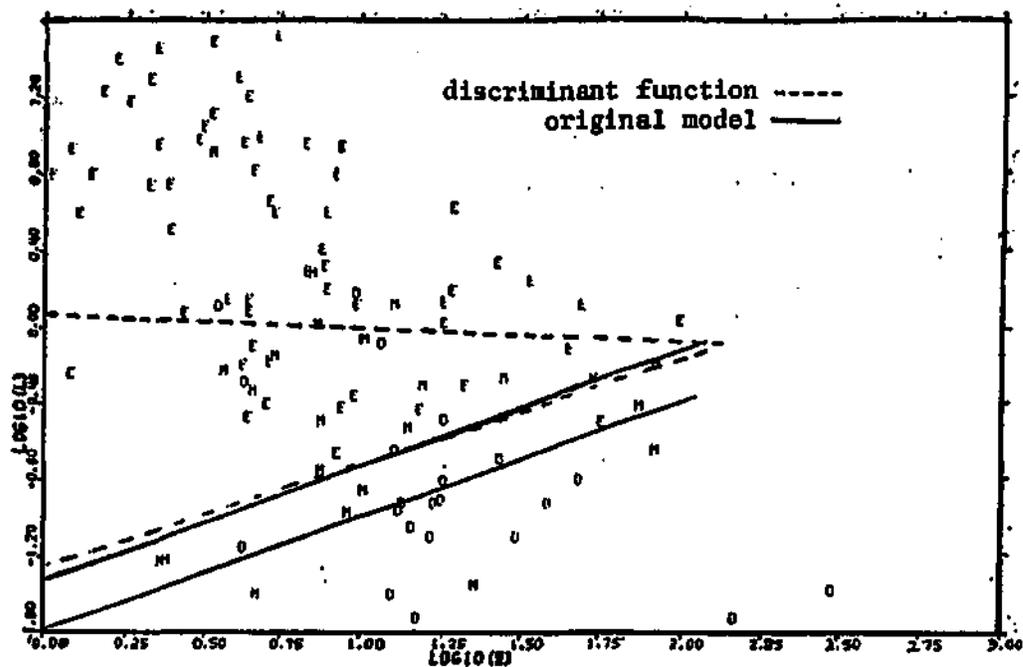
Figure 2.5-13

Lake Classification According to Vollenweider's First Model
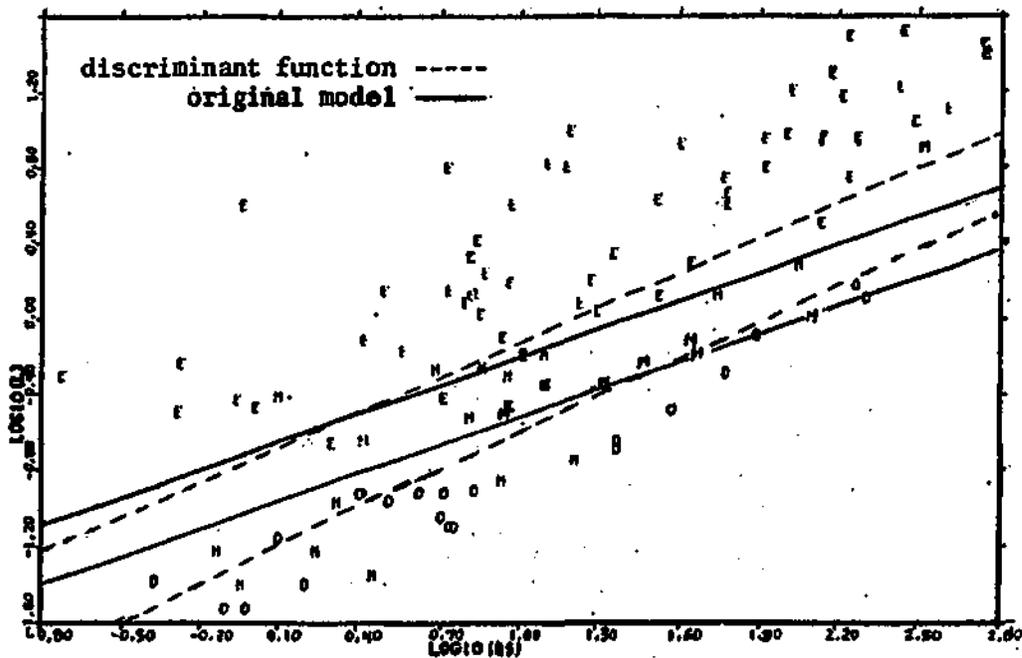
Figure 2.5-14

Lake Classification According to Vollenweider's Second Model
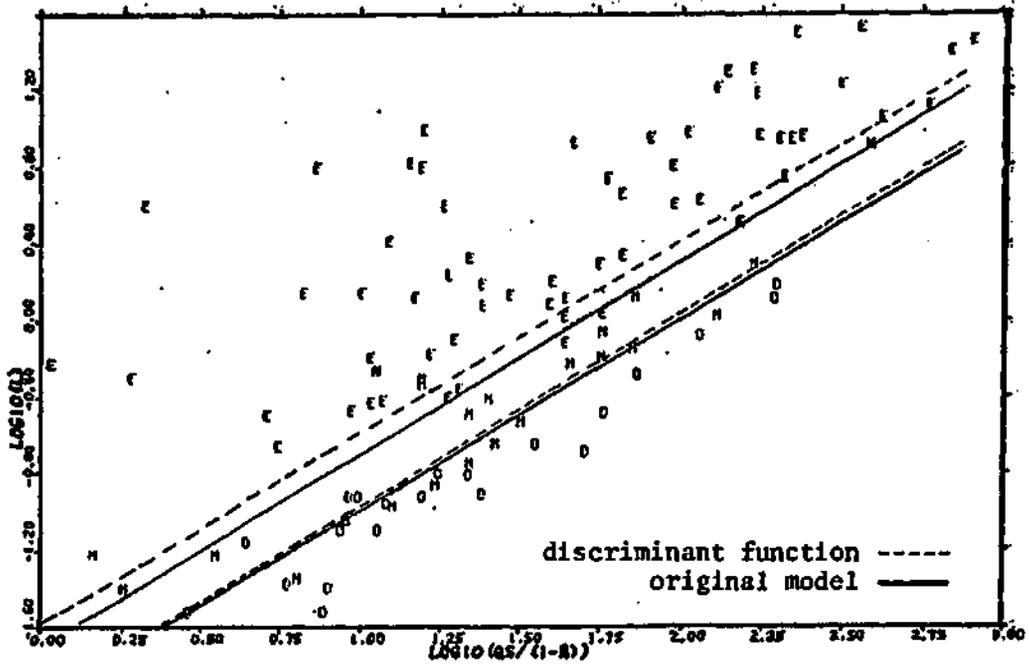
Figure 2.5-15

Lake Classification According to Dillon's Model with
Observed Retention Coefficients
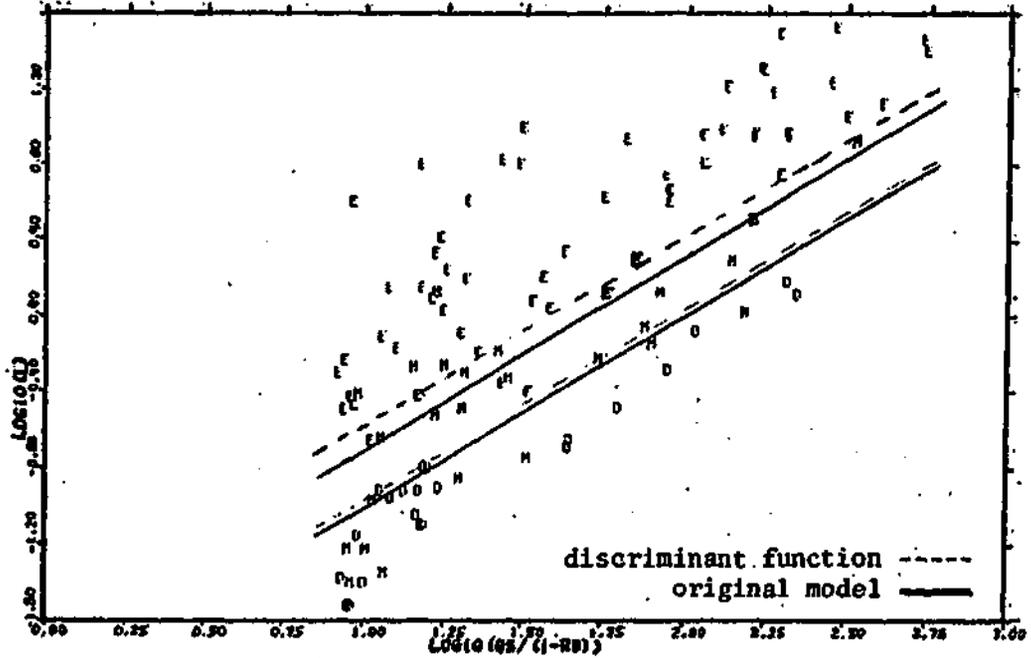
Figure 2.5-16

Lake Classification According to Dillon's Model with
Estimated Retention Coefficients

2.5.4   <u>Trophic State Prediction</u> (continued)

been evaluated: one employing the observed retention coefficient
and one employing the retention coefficient estimated according
to Dillon and Kirchner's equation for R as a function of QS
(equation (2.4-3)).

Comparing the percents misclassified by group shows that the
major effect of estimating optimal parameters for each model has
been to reduce the errors in the mesotrophic group and to distri-
bute misclassifications more evenly among the three trophic states.
Generally, little effect on the total or average percents correct
is observed as a result of estimating optimal parameters for each
model, except in the case of Vollenweider's first model, for which
the shift in line distinguishing eutrophic from mesotrophic lakes
was relatively dramatic (Figure 2.5-13). Optimization of the
parameters for the established models generally had more effect
on the "dangerous" than on the "permissible" loading lines. The
effect has been to raise the former. For instance, the mean
outlet concentration separating mesotrophic from eutrophic lakes
according to Dillon's model was raised from 0.020 g/m^3 to 0.027
g/m^3. The latter figure is closer to the value of 0.025 g/m^3
recommended by the Environmental Protection Agency as a water
quality criterion "to maintain conditions free of nuisance algal
blooms"[85].

The average percents misclassified ranged from 14.5% for

2.5.4 Trophic State Prediction (continued)

the model based upon the first principal component derived from the stepwise analysis to 38.9% for Vollenweider's first model with its original parameter values. Misclassification by more than one trophic state occurred only in the case of the latter. The analysis suggests that the L versus Z axes are clearly inferior to the other classification schemes examined. It is more difficult to distinguish among the remaining models, however.

Using the estimated, as opposed to the observed retention coefficient, seems to have improved the performance of Dillon's model somewhat. Likewise, the stepwise discriminant analysis selected COEST instead of CO in deriving the optimal linear classification model for the data. If average outlet concentration, as an indicator of average lake concentration, is an important factor in determining lake trophic state, then the above results suggest that the estimated outlet concentrations (according to Dillon's equation or the model derived in the previous section) may be superior indicators of actual lake conditions than the reported values. As discussed previously, it is likely that significant errors exist in the NES estimates of the retention coefficients, because of the assumptions employed in converting the phosphorus balance observed for the year of sampling to an "average hydrologic year". The phosphorus retention models may act as filters by removing errors in the

2.5.4 Trophic State Prediction (continued)

reported retention coefficient (or outlet concentration) data.
This suggests that the evaluation of Dillon's model based upon
observed retention coefficients may have been hampered somewhat
by the quality of the data.

The results of the stepwise analysis are presented on two
sets of coordinate axes: COEST versus L (Figure 2.5-17) and
L versus QS/(1-RE) (Figure 2.5-18). The fact that L entered
significantly into the classification scheme after COEST is
reflected by the slight downward slope of the classification lines
in Figure 2.5-17. If these lines were horizontal, concentration
alone would be significant. The apparent importance of both
concentration and loading suggests that both the amount of nutrient
available and the rate at which it is applied determine trophic
response, although the latter is of relatively minor importance.[*]
The axes have been transformed in Figure 2.5-18 to isolate loading
on the y-axis and to display the classification scheme in a manner
similar to Figures 2.5-13 through 2.5-16, representing the other
models evaluated. The fact that the discriminant lines in
Figure 2.5-18 are nearly parallel explains the success of the
first principal component in capturing most of the discrimination
power of these varaibles. The horizontal axis in Figure 2.5-18
can be interpretated as the areal removal rate of phosphorus per
unit of concentration:

---

* Another interpretation is that the canonical variable is more closely
  related to spring overturn phosphorus concentration than is outlet
  concentration alone; the former has been shown to correlate with
  mid-summer chlorophyll concentrations (se Figure 2.5-11)
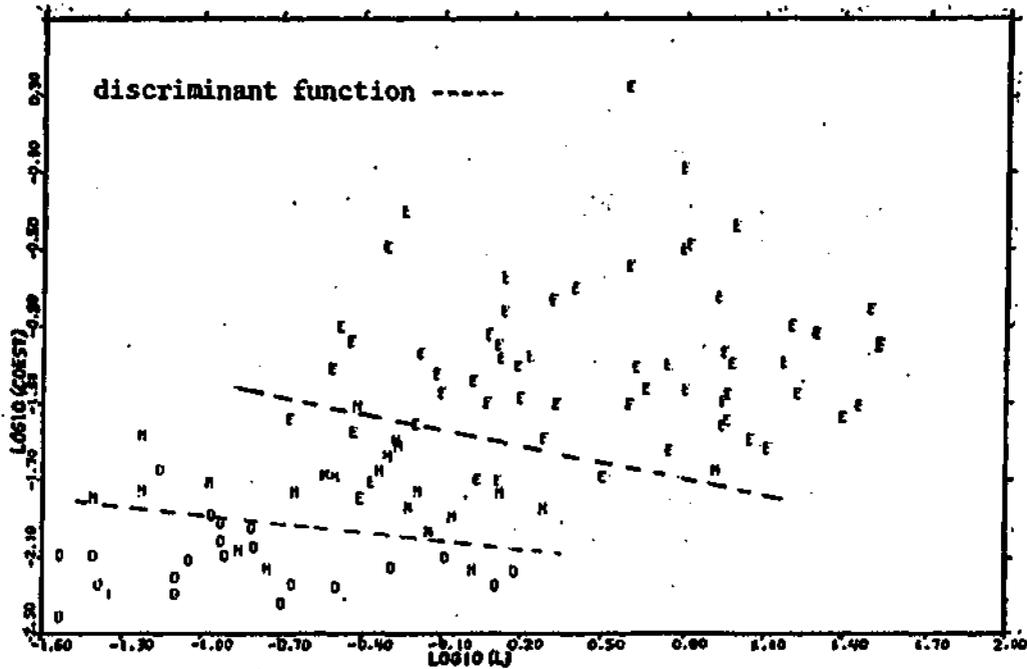
Figure 2.5-17

Lake Classification According to Model Derived from Stepwise
Discriminant Analysis on COEST vs. L Axes
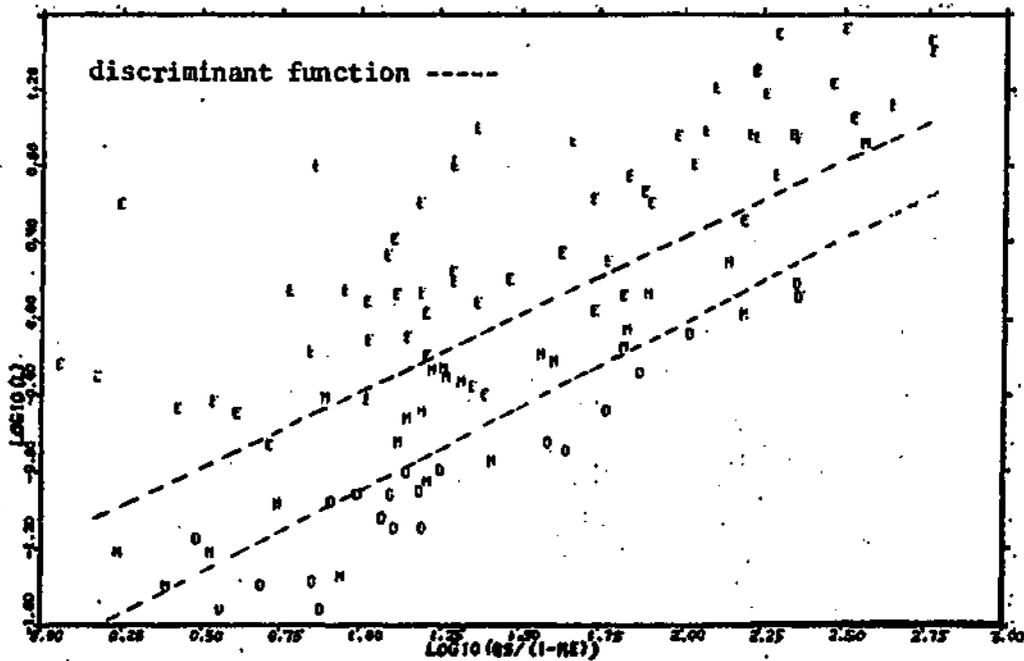
Figure 2.5-18

Lake Classification According to Model Derived from Stepwise
Discriminant Analysis on L vs. QS/(1-RE) Axes

2.5.4 <u>Trophic State Prediction</u> (continued)

$$\frac{QS}{1 - RE} = QS\,(1 + KT) = QS + KZ = QS + U \quad . \qquad (2.5-53)$$

where

K   =   effective first-order decay
coefficient (1/yr)

$= 0.824\ T^{-.546}$

U   =   effective phosphorus settling
velocity (m/yr)

$= KZ$

The settling velocity is a term first introduced by
Vollenweider[94]. The total removal rate is comprised of a
flushing term, QS , and a sedimentation term, U .

Subsequent analysis will deal with the discriminant model
based upon the first canonical variable derived from the
stepwise analysis. This model represents the best summary
of these particular data. No notion of the general superiority
of this model over the others investigated is implied, with the
possible exception of Vollenweider's first model. The techniques
employed below to interpret and apply the discriminant model
could be applied equally as well to the other models or other
data bases.

2.5.5 Discussion of Misclassified Lakes

One of the advantages of posing the lake trophic state estimation problem as a formal discriminant analysis is that the computed classification functions can be used to estimate the posterior probability that a given lake belongs to a given class. The following formula is useful is this regard[26]:

$$P_{ij} = \frac{e^{f_{ij}}}{\sum\limits_{k=1}^{n_k} e^{f_{ik}}}$$ (2.5-54)

where,

n_k = number of groups

f_{ik} = classification function value for lake i and group k

P_{ij} = posterior probably that lake i belongs to group j

Table 2.5-15 lists the jackknifed probabilities for the misclassified lakes in the analysis based upon the first canonical variable derived from the stepwise analysis (Table 2.5-12). The probabilities provide means of assessing the severity of a given misclassification. For instance, Cayuga, reportedly mesotrophic, was misclassified as eutrophic by the discriminant model, but the computed

Table 2.5-15

Misclassified Lakes

Lake	Obs. T.S.	Est T.S.	Classification Probabilities[a]		
			O	M	E
Winnepesaukee	O	M	.456	.542	.002
Crystal	O	M	.144	.830	.026
Sacandaga	M	O	.890	.110	.000
Keuka	M	O	.996	.004	.000
Cayuga	M	E	.010	.493	.497
Clyde	M	E	.003	.256	.741
Leech	M	O	.504	.494	.002
Long	M	O	.797	.203	.000
Okanagan	M	E	.002	.205	.794
Pine	M	O	.618	.382	.001
Lower St Regis	E	M	.135	.840	.026
Conesus	E	M	.009	.501	.490
Bemidgi	E	M	.073	.854	.074
Powder Mill	E	M	.001	.547	.442
Bodensee-Obersee	E	M	.003	.771	.197
Zurichsee	E	M	.031	.756	.213

a - Jackknifed

2.5.5   <u>Discussion of Misclassified Lakes</u> (continued)

probability levels for the mesotrophic and eutrophic states
were 0.493 and 0.497, respectively. Thus, this misclassifica-
tion was not a severe one. A more serious error occurred in
the case of Sacandaga, a reportedly mesotrophic lake which
was estimated to have probabilities of 0.89 and 0.11 for the
oligotrophic and mesotrophic states, respectively.

Patterns in the lake misclassifications have been compared
with other residual patterns in the analysis, including those
derived from the independent variable regressions in Table
2.5-3 and from the phosphorus retention models. No consistent
patterns were evident. Possible sources of misclassification
errors include erroneous data, model errors, and non-steady-
state conditions.

Errors could occur in both the dependent and the independent
variables. Mistakes in the originally-reported classifications
represent dependent variable errors. The most severe misclassi-
fication occurred in the case of Keuka, classified by the EPA
as mesotrophic, but which was assigned to the oligotrophic class
with a posterior probability of 0.996. Examination of the NES
working paper on this lake indicated no basis for the mesotrophic
classification. The lake had relatively high transparency, low
chlorophyll levels, and no evidence of hypolimnetic dissolved
oxygen depression. It is not surprising that most of the

2.5.5 Discussion of Misclassified Lakes (continued)

classification errors involved mesotrophic lakes, since a lake
in this class could be misclassified in either of two directions.
The misclassifications arising from errors in the originally-
reported trophic states might be traced to errors in data or
in its interpretation. Errors of the latter type could be
reduced by developing and applying a more objective classifica-
tion scheme, similar to Shannon and Brezonik's Trophic State
Index[73]. Another source of misclassifications is errors in the
independent variables, i.e. in the estimates of phosphorus
loading, hydraulic residence time, or mean depth.

Model errors would occur in systems with particular char-
acteristics which would tend to alter their behavior relative
to other lakes. One such case is Powder Mill, a eutrophic pond
with a residence time of 6.6 days and with extensive growths
of aquatic macrophytes.[86]The misclassification of the pond as
mesotrophic may be due to the relatively rapid flushing of
nutrients having less of a regulating effect upon the producti-
vity of rooted vegetation than upon the productivity of suspended
phytoplankton. As noted in Section 2.2, lakea with aquatic weed
problems generally do not conform to classification schemes
developed for lakes dominated by algal productivity. Possible
effects of nitrogen limitation in some systems may be another
source of model error.

2.5.5 <u>Discussion of Misclassified Lakes</u> (continued)

A third type of error would arise from non-steady-state conditions existing in systems whose nutrient budgets and/or biota had not fully responded to changes in trends in phosphorus loadings at the time of sampling. For instance, the under-classification of New York State lakes Sacandaga, Keuka, and Conesus may have been due in part to the effects of detergent legislation which was being implemented in parts of New York at about the time of the NES sampling (1972). This might have influenced the phosphorus loadings, but the trophic states would not have had time to respond fully to any such changes. The under-classification of Bodensee-Obersee may have been due to a similar effect. Loading values of 4 and 1.1 grams/m^2-yr are given for this lake in Vollenweider's 1968 [93] and 1973 [95] papers, respectively. If the difference in loading is not due to an error, the lack of response time may explain this misclassification, since the latter loading value was used.

According to Uttormark and Wall's [92] dynamic lake classification scheme (Figure 2.2-1), the water quality of the under-classified lakes would be improving and that of over-classified lakes would be degrading. Accordingly, the misclassifications in Table 2.5-15 would serve as a basis for selecting likely candidates for lake protection and renewal programs. Such decisions would be made, however, only after other possible sources of classification error had been eliminated.

2.5.6 Application Strategies

The following section demonstrates the application of the discriminant model in a management context. The particular problem addressed is that of rationally selecting design value for an "acceptable" phosphorus loading for a given lake under conditions of uncertainty in the information used to formulate the problem. The particular management objective selected for the analysis is that the lake of concern be "non-eutrophic", i.e. mesotrophic or oligotrophic. The approach developed below could be applied equally as well to other possible objectives, such as achieving or preserving oligotrophic status. When there is uncertainty in the information or models employed in this type of design, probabilistic considerations become important. Under such conditions, the rational design basis is to satisfy a given probability level of achieving the management objective.

The property that the discriminant model can be used to generate classification probabilities is particularly useful in this application. The principal component analysis of the stepwise discriminant model has reduced the classification problem to a single dimension. Using the results in Table 2.5-12 and equations (51), (52), and (54), a probability can be assigned to any value of the canonical variable, X, according to:

2.5.6 <u>Application Strategies</u> (continued)

$$p(e : X) = \frac{\exp(f_e)}{\exp(f_o) + \exp(f_m) + \exp(f_e)} \qquad (2.5\text{-}55)$$

$$f_e = -18.508 - 20.487 \; XT$$

$$f_m = -36.773 - 29.327 \; XT$$

$$f_o = -53.804 - 35.646 \; XT$$

$$XT = -X^{-.25}$$

$$X = L \; [QS/(1 - RE)]^{-.8147}$$

$$1 - RE = 1./[1. + .824 \; T^{.454}]$$

Substitution of the corresponding exponential into the
numerator of equation (55) would generate a probability for
either of the other two trophic states from estimates of L ,
Z , and T . A plot of these probabilities against the base
10 logarithm of X is given in Figure 2.5-19. This plot
clearly shows the separation of the trophic states according
to the discriminant model. Figures 2.5-20 and 2.5-21 contain
the same information plotted on log L versus log (QS/[1-RE])
axes. The lines in Figures 2.5-20 and 2.5-21 correspond to lines
of constant probability of eutrophy and oligotrophy, respectively.
In a management context, the probabilistic interpretation of
these lines renders them more useful than the "permissible" and

Figure 2.5-19

Trophic State Probabilities Versus Canonical Variable X
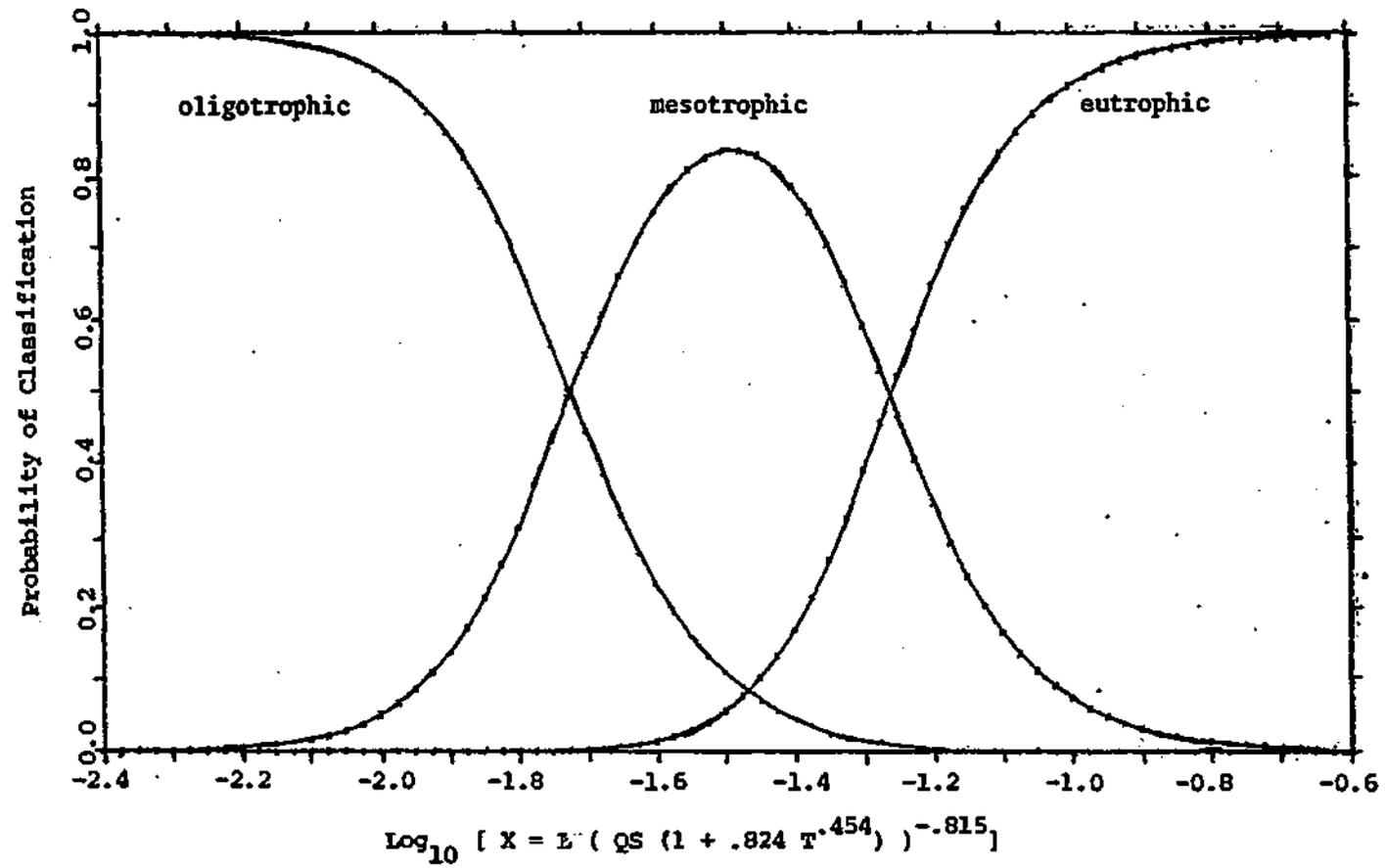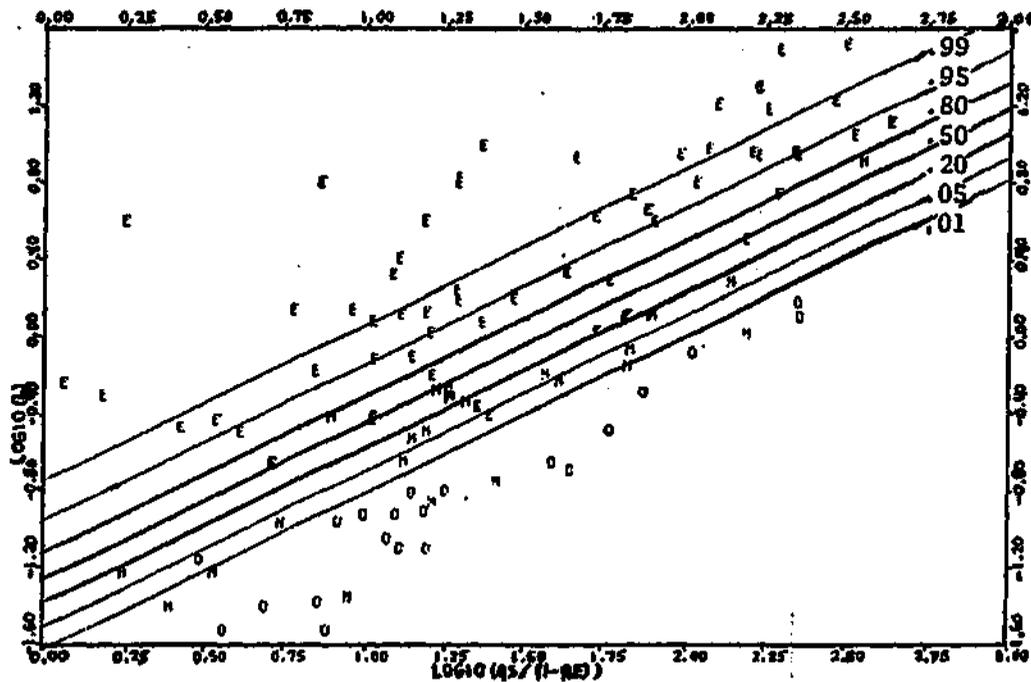
Figure 2.5-20

Lines of Constant Eutrophic Probability on L vs. QS/(1-RE) Axes
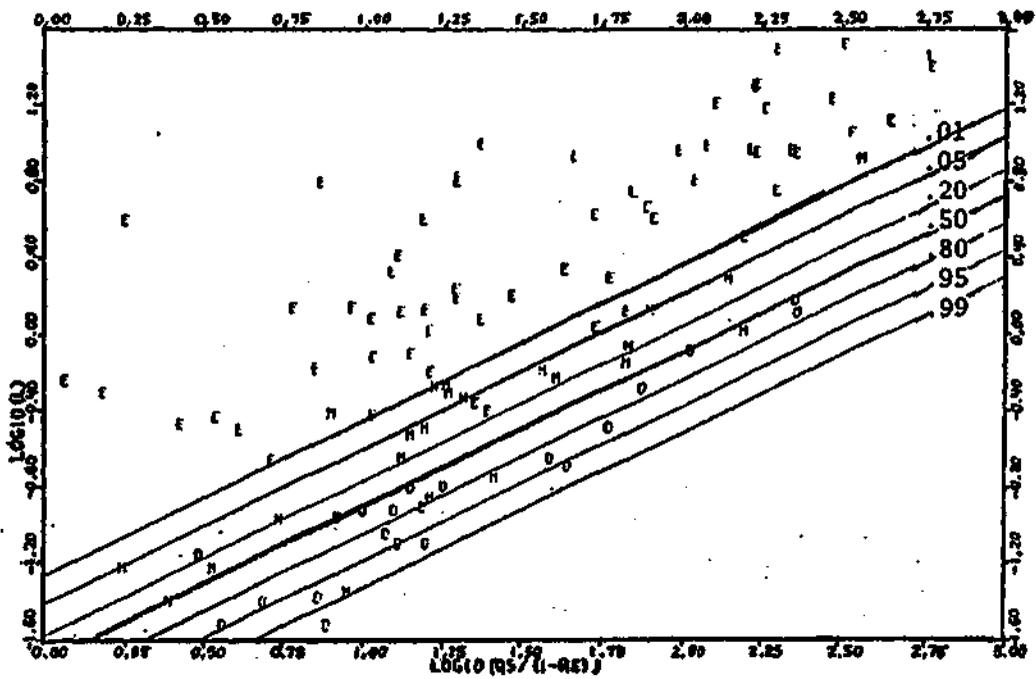
Figure 2.5-21

Lines of Constant Oligotrophic Probability on L vs. QS/(1-RE) Axes

2.5.6   Application Strategies (continued)

"dangerous" lines employed in the Vollenweider plots.

Suppose that the design basis were to be 95% sure that a given lake would be non-eutrophic. If there were no uncertainty in the variables needed to compute the value of X as a function of management strategy, i.e. in L , Z , or T , the rational loading allocation would correspond to the intersection of the 0.05 line in Figure 2.5-20 and a vertical line at the value of QS/(1-RE) appropriate for the particular lake. An equivalent method would be to find the value of log X in Figure 2.5-19 corresponding to a 0.05 eutrophic probability and to solve for the equivalent value of L . This has been done for eutrophic probabilities of 0.01, 0.05, and 0.10:

$$L_{.01} = 0.0237 \ [QS \ (1 + 0.824 \ T^{.454})]^{.815} \qquad (2.5\text{-}56)$$

$$L_{.05} = 0.0310 \ [QS \ (1 + 0.824 \ T^{.454})]^{.815} \qquad (2.5\text{-}57)$$

$$L_{.10} = 0.0354 \ [QS \ (1 + 0.824 \ T^{.454})]^{.815} \qquad (2.5\text{-}58)$$

These equations could be considered design equations for cases in which there is no uncertainty in the loading, residence time, or mean depth estimates.

2.5.6 Application Strategies (continued)

When X cannot be specified exactly, however, the problem
is somewhat more complicated. Uncertainty in X arises from
sampling, measurement, and/or estimation errors in the estimates
of L , T , and Z , all of which refer to mean quantities.
This uncertainty is more precisely defined as the inability to
exactly specify the X value which would result from the imple-
mentation of a particular management strategy. For instance,
it would not be possible to exactly estimate the phosphorus
removal efficiency of a future tertiary treatment plant.
Existing or future non-point sources of phosphorus may be poorly-
defined and thereby contribute substantial uncertainty to the
estimate of total phosphorus loading. The effects of changes
in land use on non-point sources would likewise be difficult
to predict with much accuracy. Uncertainty in X would also
rise from errors in the estimates or measurements of hydrology
(groundwater and surfacewater flows) and lake morphology. The
total uncertainty in X measures the relative ability of the
planner or designer to match the target value of X specified
by the discriminant model, given the available information and
control technologies.

A first-order error analysis (Table 2.5-16) can be used
to roughly estimate the error distribution of X from the error
distributions of L , T , and Z . Clean estimates of the
covariance matrix of L , T , and Z are difficult to obtain.

Table 2.5-16

First-Order Error Analysis for Estimates of Canonical Variable X

Original definition :

$$X = L \ [\ QS \ / \ (\ 1 - RE \ ) \ ]^{-.815}$$

In terms of CI, QS, and Z :

$$X = CI \ QS^{.185} \ [\ 1 + .824 \ Z^{.454} QS^{-.454}]^{-.815}$$

From Expected Value Theory :

$$Var(X) = \sum_{i=1}^{3} \sum_{j=1}^{3} \ (\frac{\partial X}{\partial y_i}) \ (\frac{\partial X}{\partial y_j}) \ Cov(y_i, y_j)$$

where:

$$y_1 = CI$$
$$y_2 = QS$$
$$y_3 = Z$$

Assuming $Cov(y_i, y_j) = 0$ for $i \neq j$:

$$\frac{Var(X)}{X^2} = \frac{Var(CI)}{CI^2} + \alpha^2 \frac{Var(Z)}{Z^2} + (\alpha + .185)^2 \frac{Var(QS)}{QS^2}$$

$$\boxed{\overline{CV}_X^2 = \overline{CV}_{CI}^2 + \alpha^2 \overline{CV}_Z^2 + (\alpha + .185)^2 \overline{CV}_{QS}^2}$$

where :

$$\alpha = \frac{.305 \ Z^{.454} \ QS^{-.454}}{1 + .824 \ Z^{.454} QS^{-.454}}$$

2.5.6 Application Strategies (continued)

The problem can be simplified considerably in cases in which the off-diagonal elements of the matrix can be assumed to be negligible. In order to reduce the importance of the off-diagonal elements, the algebra in Table 2.5-16 has been done with X expressed in terms of CI, QS, and Z. This roughly partitions the total error into a concentration, a hydrologic, and a morphometric component. Since each of these variables represents a mean value, the standard error of the population mean can be computed directly if the estimate is based directly upon independent observations. Consideration of measurement errors and subjective assessments of uncertainty by people familiar with the respective sampling and measurment problems can also be incorporated into the error estimates. Since QS and Z represent relatively easily-observable quantities, errors in these estimates would be expected to contribute relatively little to the total uncertainty in X. Accordingly, most of the uncertainty would generally be attributed to the average inlet concentration estimates. Because the second deviation of X with respect to CI vanishes, higher-order terms in the expected value of X are insignificant and have been neglected.

Techniques for estimating the error distribution of a loading or average inlet concentration estimate from continuous flow and grab-sample concentration measurements have been

2.5.6 Application Strategies (continued)

discussed previously (Section 2.3.2). If loading has been
estimated based upon tributary land use or population data,
the distributions of unit loading factors (e.g. $g-P/m^2$ Urban
Area-year) found in the literature could serve as a basis to
approximate the variance of the average inlet concentration
estimate.

Derived from the equations in Table 2.5-16, an estimate
of the coefficient of variation of X can provide a means of
assessing the effects of this type of uncertainty upon the
trophic state projection and upon the rational loading alloca-
tion. Two forms are suggested for the error distribution of X :
normal and log-normal. The former seems appropriate in the sense
that X represents an estimate of a mean quality, the distribution
of which would tend toward normality as sample size increases. The
latter has the desirable and realistic property of postivity and
probably more adequately reflects the distributions of the measure-
ments employed in estimating L and QS . The calculations outlined
below have been done for each of these assumed distributions.

Assume that the error distribution of X can be described by
completely by the two parameters $\hat{X}$ and $S_{\hat{X}}$. The following
multiplicative rule can be employed to combine the trophic
state and error distributions:

$$p(e : \dot{X}, S_{\hat{X}}) = p(e : X) \cdot p(X : \hat{X}, S_{\hat{X}})$$
(2.5-59)

2.5.6 Application Strategies (continued)

The probability of a eutrophic state classification, given $\hat{X}$ and $S_{\hat{X}}$, equals the probability of the classification, given X (equation (55)), times the probability of X, given $\hat{X}$ and $S_{\hat{X}}$. This equation can be evaluated by substituting the expression for the probability density function of X according to the assumed distribution and integrating over all possible X :

For normal $\hat{X}$:

$$p(e : \hat{X}, S_{\hat{X}}) = \int_0^\infty p(e : X) \frac{1}{S_{\hat{X}}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{X-\hat{X}}{S_{\hat{X}}})^2} dX \qquad (2.5\text{-}60)$$

For lognormal $\hat{X}$:

$$p(e : \hat{X}, S_{\hat{X}}) = \int_{-\infty}^{\infty} p(e : X) \frac{1}{S_{\widehat{\log X}}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\log X - \widehat{\log X}}{S_{\widehat{\log X}}})^2} d\log X \qquad (2.5\text{-}61)$$

From the theory of the lognormal distribution[1] :

$$S_{\widehat{\log X}} \simeq \sqrt{\ln(1 + \overline{CV}_X^2)}/2.303 \qquad (2.5\text{-}62)$$

$$\widehat{\log X} \simeq \log\hat{X} - \frac{1}{2} S_{\widehat{\log X}}^2 \qquad (2.5\text{-}63)$$

$$\overline{CV}_X = S_{\hat{X}}/\hat{X} \qquad (2.5\text{-}64)$$

2.5.6 Application Strategies (continued)

Equations (60) and (61) have been integrated numerically using Simpson's rule and the results are presented in Tables 2.5-17 and 2.5-18. Tables 2.5-19 and 2.5-20 present corresponding results for the oligotrophic state. Mesotrophic probabilities can be determined by difference. Essential aspects of these tables are summarized in the contour plots of Figure 2.5-22. The probabilities have been evalued for $\hat{X}$ values ranging from 0.006 to 0.25 and $CV_{\hat{X}}$ values from 0. to 0.8. At $CV_X = 0.$, the results correspond to the curves in Figure 2.5-19. As the coefficient of variation of the X estimate increases, these curves spread out and the trophic state distinctions become less clear, as is evident in Figure 2.5-22.

At low probabilities of eutrophy, the effect of uncertainty in X is more pronounced for the lognormal distribution as compared with the normal distribution. This reflects the skewness of the former toward high X values. Because of this aspect, the lognormal tables and contour lines are more conservative than the normal ones. The normal tables and lines are probably more appropriate for use when the estimate is based upon a large number of direct measurements and is relatively well-defined.

The above results can be used to estimate the effects of uncertainty in X on the design value which satisfies a given

Table 2.5-17

Eutrophic Probabilities for Normal Error Distribution of X
Coefficient of Variation of X

$\log_{10}X$	X	0.0	.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
-2.2000	0.00631	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
-2.1800	0.00661	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
-2.1600	0.00692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
-2.1400	0.00724	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002
-2.1200	0.00759	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002
-2.1000	0.00794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
-2.0800	0.00832	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0005
-2.0600	0.00871	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007
-2.0400	0.00912	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0005	0.0009
-2.0200	0.00955	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012
-2.0000	0.01000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0006	0.0010	0.0017
-1.9800	0.01047	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0008	0.0014	0.0022
-1.9600	0.01096	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0006	0.0011	0.0018	0.0029
-1.9400	0.01148	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0008	0.0015	0.0025	0.0039
-1.9200	0.01202	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0006	0.0011	0.0020	0.0033	0.0050
-1.9000	0.01259	0.0000	0.0000	0.0001	0.0001	0.0002	0.0004	0.0008	0.0015	0.0027	0.0043	0.0065
-1.8800	0.01318	0.0001	0.0001	0.0001	0.0001	0.0002	0.0005	0.0011	0.0021	0.0035	0.0056	0.0083
-1.8600	0.01380	0.0001	0.0001	0.0002	0.0002	0.0003	0.0008	0.0015	0.0028	0.0047	0.0072	0.0106
-1.8400	0.01445	0.0002	0.0002	0.0002	0.0003	0.0005	0.0011	0.0021	0.0037	0.0061	0.0093	0.0133
-1.8200	0.01514	0.0003	0.0003	0.0004	0.0005	0.0007	0.0015	0.0028	0.0049	0.0079	0.0118	0.0166
-1.8000	0.01585	0.0004	0.0004	0.0006	0.0008	0.0011	0.0021	0.0038	0.0065	0.0101	0.0149	0.0206
-1.7800	0.01660	0.0006	0.0007	0.0008	0.0011	0.0015	0.0029	0.0051	0.0084	0.0129	0.0186	0.0254
-1.7600	0.01738	0.0009	0.0010	0.0012	0.0016	0.0021	0.0039	0.0068	0.0109	0.0163	0.0231	0.0310
-1.7400	0.01820	0.0014	0.0015	0.0018	0.0023	0.0030	0.0053	0.0089	0.0139	0.0205	0.0284	0.0375
-1.7200	0.01906	0.0020	0.0021	0.0025	0.0032	0.0041	0.0070	0.0115	0.0177	0.0255	0.0347	0.0453
-1.7000	0.01995	0.0029	0.0030	0.0035	0.0044	0.0056	0.0093	0.0149	0.0223	0.0315	0.0420	0.0536
-1.6800	0.02089	0.0041	0.0043	0.0049	0.0060	0.0076	0.0123	0.0190	0.0279	0.0385	0.0505	0.0634
-1.6600	0.02188	0.0057	0.0060	0.0068	0.0082	0.0102	0.0159	0.0241	0.0346	0.0467	0.0601	0.0743
-1.6400	0.02291	0.0079	0.0083	0.0093	0.0110	0.0135	0.0206	0.0304	0.0425	0.0562	0.0711	0.0865
-1.6200	0.02399	0.0108	0.0112	0.0125	0.0147	0.0177	0.0263	0.0379	0.0517	0.0671	0.0834	0.0998
-1.6000	0.02512	0.0146	0.0151	0.0167	0.0194	0.0231	0.0334	0.0468	0.0625	0.0795	0.0970	0.1144
-1.5800	0.02630	0.0195	0.0202	0.0221	0.0254	0.0298	0.0419	0.0573	0.0748	0.0933	0.1119	0.1302
-1.5600	0.02754	0.0258	0.0266	0.0290	0.0329	0.0382	0.0522	0.0696	0.0880	0.1086	0.1282	0.1471
-1.5400	0.02884	0.0338	0.0347	0.0376	0.0422	0.0484	0.0645	0.0838	0.1046	0.1255	0.1458	0.1651
-1.5200	0.03020	0.0438	0.0449	0.0483	0.0537	0.0608	0.0789	0.1000	0.1221	0.1439	0.1646	0.1843
-1.5000	0.03162	0.0563	0.0576	0.0615	0.0676	0.0757	0.0957	0.1183	0.1414	0.1637	0.1846	0.2039
-1.4800	0.03311	0.0717	0.0732	0.0776	0.0844	0.0933	0.1149	0.1387	0.1624	0.1849	0.2056	0.2245
-1.4600	0.03468	0.0905	0.0921	0.0969	0.1044	0.1140	0.1367	0.1612	0.1850	0.2072	0.2275	0.2458
-1.4400	0.03631	0.1130	0.1148	0.1199	0.1278	0.1378	0.1611	0.1857	0.2092	0.2308	0.2502	0.2675
-1.4200	0.03802	0.1397	0.1415	0.1468	0.1549	0.1649	0.1881	0.2121	0.2347	0.2552	0.2735	0.2897
-1.4000	0.03981	0.1709	0.1727	0.1778	0.1857	0.1954	0.2176	0.2403	0.2614	0.2805	0.2973	0.3122
-1.3800	0.04169	0.2067	0.2084	0.2131	0.2202	0.2291	0.2494	0.2700	0.2892	0.3063	0.3214	0.3347
-1.3600	0.04365	0.2471	0.2485	0.2525	0.2583	0.2658	0.2832	0.3010	0.3177	0.3326	0.3458	0.3573
-1.3400	0.04571	0.2918	0.2926	0.2953	0.2996	0.3052	0.3187	0.3331	0.3468	0.3592	0.3701	0.3798
-1.3200	0.04787	0.3403	0.3403	0.3414	0.3435	0.3467	0.3556	0.3660	0.3763	0.3858	0.3944	0.4021
-1.3000	0.05012	0.3911	0.3907	0.3900	0.3896	0.3900	0.3935	0.3993	0.4059	0.4124	0.4185	0.4240
-1.2800	0.05248	0.4439	0.4428	0.4401	0.4370	0.4343	0.4319	0.4327	0.4354	0.4387	0.4422	0.4456
-1.2600	0.05495	0.4973	0.4955	0.4908	0.4849	0.4790	0.4704	0.4661	0.4646	0.4647	0.4655	0.4667
-1.2400	0.05755	0.5500	0.5476	0.5411	0.5325	0.5235	0.5087	0.4991	0.4934	0.4901	0.4883	0.4873
-1.2200	0.06026	0.6009	0.5980	0.5901	0.5791	0.5672	0.5463	0.5314	0.5215	0.5149	0.5104	0.5073
-1.2000	0.06310	0.6492	0.6454	0.6369	0.6239	0.6094	0.5828	0.5629	0.5489	0.5390	0.5319	0.5266
-1.1800	0.06607	0.6940	0.6901	0.6808	0.6664	0.6498	0.6181	0.5934	0.5754	0.5624	0.5527	0.5453
-1.1600	0.06919	0.7355	0.7315	0.7214	0.7061	0.6879	0.6510	0.6227	0.6019	0.5848	0.5727	0.5633
-1.1400	0.07245	0.7713	0.7684	0.7584	0.7427	0.7234	0.6838	0.6506	0.6254	0.6064	0.5919	0.5806
-1.1200	0.07586	0.8045	0.8013	0.7916	0.7760	0.7562	0.7138	0.6772	0.6488	0.6270	0.6103	0.5971
-1.1000	0.07944	0.8332	0.8331	0.8211	0.8060	0.7862	0.7419	0.7023	0.6710	0.6467	0.6279	0.6130
-1.0800	0.08318	0.8581	0.8553	0.8471	0.8328	0.8134	0.7680	0.7260	0.6920	0.6654	0.6446	0.6281
-1.0600	0.08710	0.8796	0.8772	0.8697	0.8565	0.8378	0.7923	0.7481	0.7119	0.6832	0.6606	0.6425
-1.0400	0.09121	0.8980	0.8959	0.8892	0.8772	0.8595	0.8141	0.7687	0.7306	0.7000	0.6757	0.6562
-1.0200	0.09551	0.9137	0.9119	0.9060	0.8952	0.8787	0.8342	0.7879	0.7481	0.7159	0.6901	0.6692
-1.0000	0.10000	0.9271	0.9254	0.9204	0.9108	0.8956	0.8524	0.8056	0.7646	0.7309	0.7037	0.6816
-0.9800	0.10472	0.9384	0.9370	0.9326	0.9242	0.9103	0.8689	0.8220	0.7799	0.7449	0.7165	0.6933
-0.9600	0.10965	0.9479	0.9467	0.9429	0.9356	0.9232	0.8838	0.8371	0.7942	0.7582	0.7286	0.7044
-0.9400	0.11482	0.9559	0.9547	0.9517	0.9454	0.9343	0.8971	0.8510	0.8076	0.7706	0.7400	0.7149
-0.9200	0.12023	0.9627	0.9617	0.9591	0.9537	0.9439	0.9090	0.8637	0.8200	0.7822	0.7508	0.7248
-0.9000	0.12590	0.9684	0.9677	0.9654	0.9607	0.9521	0.9196	0.8753	0.8315	0.7931	0.7609	0.7342
-0.8800	0.13183	0.9732	0.9726	0.9707	0.9667	0.9591	0.9290	0.8859	0.8421	0.8033	0.7705	0.7433
-0.8600	0.13805	0.9773	0.9768	0.9751	0.9717	0.9652	0.9373	0.8955	0.8520	0.8128	0.7794	0.7514
-0.8400	0.14454	0.9807	0.9802	0.9789	0.9760	0.9703	0.9447	0.9043	0.8611	0.8217	0.7879	0.7592
-0.8200	0.15137	0.9836	0.9832	0.9820	0.9796	0.9747	0.9513	0.9123	0.8695	0.8300	0.7958	0.7667
-0.8000	0.15850	0.9860	0.9856	0.9847	0.9826	0.9784	0.9570	0.9195	0.8773	0.8378	0.8032	0.7737
-0.7800	0.16597	0.9882	0.9877	0.9869	0.9852	0.9816	0.9621	0.9261	0.8845	0.8450	0.8101	0.7804
-0.7600	0.17379	0.9898	0.9895	0.9888	0.9874	0.9843	0.9665	0.9321	0.8912	0.8517	0.8166	0.7866
-0.7400	0.18198	0.9912	0.9910	0.9904	0.9892	0.9866	0.9705	0.9375	0.8973	0.8580	0.8228	0.7925
-0.7200	0.19055	0.9925	0.9923	0.9918	0.9908	0.9885	0.9739	0.9424	0.9033	0.8638	0.8285	0.7981
-0.7000	0.19954	0.9936	0.9934	0.9930	0.9921	0.9902	0.9769	0.9468	0.9082	0.8693	0.8340	0.8033
-0.6800	0.20894	0.9945	0.9943	0.9939	0.9932	0.9916	0.9796	0.9509	0.9130	0.8744	0.8391	0.8082
-0.6600	0.21879	0.9952	0.9951	0.9948	0.9942	0.9928	0.9819	0.9545	0.9174	0.8792	0.8439	0.8128
-0.6400	0.22910	0.9959	0.9951	0.9955	0.9950	0.9938	0.9840	0.9578	0.9216	0.8836	0.8485	0.8171
-0.6200	0.23990	0.9964	0.9962	0.9961	0.9957	0.9947	0.9858	0.9608	0.9253	0.8878	0.8527	0.8211
-0.6000	0.25121	0.9969	0.9968	0.9966	0.9962	0.9954	0.9873	0.9635	0.9288	0.8917	0.8567	0.8248

Table 2.5-18

Eutrophic Probabilities for Lognormal Error Distribution of X

Coefficient of Variation of X

$\log_{10}$ X	X	0.0	.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
-2.2000	0.00631	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0019
-2.1800	0.00661	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0011	0.0023
-2.1600	0.00692	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0014	0.0027
-2.1400	0.00724	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0017	0.0033
-2.1200	0.00759	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0021	0.0039
-2.1000	0.00794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0026	0.0046
-2.0800	0.00832	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0015	0.0031	0.0055
-2.0600	0.00871	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0018	0.0038	0.0065
-2.0400	0.00912	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0023	0.0046	0.0076
-2.0200	0.00955	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0028	0.0055	0.0090
-2.0000	0.01000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0015	0.0035	0.0066	0.0105
-1.9800	0.01047	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0007	0.0019	0.0043	0.0078	0.0122
-1.9600	0.01096	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0025	0.0053	0.0093	0.0142
-1.9400	0.01148	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0031	0.0064	0.0113	0.0164
-1.9200	0.01202	0.0000	0.0000	0.0000	0.0001	0.0001	0.0005	0.0016	0.0039	0.0078	0.0130	0.0190
-1.9000	0.01259	0.0000	0.0000	0.0001	0.0001	0.0002	0.0007	0.0021	0.0049	0.0094	0.0152	0.0218
-1.8800	0.01318	0.0001	0.0001	0.0001	0.0002	0.0003	0.0010	0.0027	0.0061	0.0113	0.0178	0.0251
-1.8600	0.01380	0.0001	0.0001	0.0002	0.0003	0.0004	0.0013	0.0035	0.0076	0.0136	0.0208	0.0287
-1.8400	0.01445	0.0002	0.0002	0.0003	0.0004	0.0006	0.0018	0.0045	0.0094	0.0162	0.0242	0.0327
-1.8200	0.01514	0.0003	0.0003	0.0004	0.0006	0.0009	0.0024	0.0058	0.0115	0.0192	0.0280	0.0372
-1.8000	0.01585	0.0004	0.0004	0.0006	0.0008	0.0013	0.0033	0.0074	0.0141	0.0227	0.0324	0.0422
-1.7800	0.01660	0.0006	0.0007	0.0009	0.0012	0.0018	0.0043	0.0094	0.0171	0.0267	0.0372	0.0477
-1.7600	0.01738	0.0009	0.0010	0.0013	0.0017	0.0025	0.0057	0.0117	0.0206	0.0313	0.0427	0.0538
-1.7400	0.01820	0.0014	0.0015	0.0018	0.0024	0.0035	0.0075	0.0147	0.0248	0.0366	0.0487	0.0604
-1.7200	0.01936	0.0020	0.0021	0.0026	0.0034	0.0048	0.0097	0.0182	0.0296	0.0425	0.0555	0.0677
-1.7000	0.01995	0.0029	0.0031	0.0036	0.0047	0.0064	0.0125	0.0224	0.0352	0.0492	0.0630	0.0757
-1.6800	0.02089	0.0041	0.0043	0.0051	0.0064	0.0086	0.0160	0.0275	0.0417	0.0567	0.0712	0.0843
-1.6600	0.02188	0.0057	0.0060	0.0070	0.0087	0.0114	0.0203	0.0334	0.0491	0.0651	0.0802	0.0937
-1.6400	0.02291	0.0079	0.0083	0.0095	0.0117	0.0150	0.0255	0.0404	0.0575	0.0744	0.0900	0.1038
-1.6200	0.02399	0.0108	0.0113	0.0128	0.0155	0.0195	0.0319	0.0486	0.0669	0.0847	0.1038	0.1147
-1.6000	0.02512	0.0146	0.0152	0.0171	0.0204	0.0252	0.0395	0.0580	0.0776	0.0960	0.1124	0.1264
-1.5800	0.02630	0.0195	0.0202	0.0225	0.0265	0.0323	0.0487	0.0689	0.0895	0.1085	0.1249	0.1388
-1.5600	0.02754	0.1258	0.1267	0.1295	0.1343	0.1413	0.1595	0.1812	0.1327	0.1223	0.1384	0.1521
-1.5400	0.02884	0.0338	0.0349	0.0382	0.0438	0.0516	0.0721	0.0952	0.1173	0.1367	0.1529	0.1662
-1.5200	0.03020	0.0438	0.0451	0.0491	0.0556	0.0644	0.0868	0.1109	0.1333	0.1525	0.1683	0.1811
-1.5000	0.03162	0.0563	0.0571	0.0624	0.0698	0.0797	0.1036	0.1284	0.1507	0.1695	0.1847	0.1968
-1.4800	0.03311	0.0717	0.0735	0.0786	0.0869	0.0976	0.1227	0.1477	0.1696	0.1876	0.2020	0.2133
-1.4600	0.03468	0.0905	0.0924	0.0982	0.1072	0.1185	0.1442	0.1689	0.1903	0.2069	0.2203	0.2306
-1.4400	0.03631	0.1130	0.1151	0.1213	0.1308	0.1425	0.1681	0.1920	0.2116	0.2274	0.2394	0.2486
-1.4200	0.03802	0.1367	0.1420	0.1484	0.1581	0.1698	0.1945	0.2168	0.2349	0.2489	0.2594	0.2674
-1.4000	0.03981	0.1739	0.1732	0.1796	0.1891	0.2003	0.2233	0.2435	0.2594	0.2714	0.2803	0.2868
-1.3800	0.04169	0.2067	0.2069	0.2130	0.2238	0.2340	0.2544	0.2718	0.2851	0.2949	0.3019	0.3068
-1.3600	0.04365	0.2471	0.2490	0.2544	0.2621	0.2707	0.2876	0.3016	0.3119	0.3192	0.3242	0.3275
-1.3400	0.04571	0.2918	0.2932	0.2975	0.3035	0.3102	0.3228	0.3327	0.3398	0.3444	0.3472	0.3486
-1.3200	0.04787	0.3400	0.3410	0.3438	0.3478	0.3518	0.3595	0.3651	0.3685	0.3702	0.3707	0.3703
-1.3000	0.05012	0.3911	0.3914	0.3925	0.3939	0.3953	0.3974	0.3983	0.3979	0.3966	0.3947	0.3923
-1.2800	0.05248	0.4439	0.4435	0.4427	0.4414	0.4400	0.4363	0.4322	0.4279	0.4234	0.4190	0.4147
-1.2600	0.05496	0.4973	0.4962	0.4934	0.4896	0.4851	0.4756	0.4666	0.4502	0.4506	0.4437	0.4373
-1.2400	0.05755	0.5500	0.5483	0.5438	0.5374	0.5302	0.5155	0.5011	0.4888	0.4780	0.4685	0.4601
-1.2200	0.06026	0.6009	0.5987	0.5927	0.5843	0.5745	0.5541	0.5355	0.5193	0.5054	0.4935	0.4831
-1.2000	0.06310	0.6492	0.6465	0.6395	0.6293	0.6174	0.5924	0.5696	0.5496	0.5328	0.5184	0.5061
-1.1800	0.06607	0.6940	0.6911	0.6833	0.6719	0.6584	0.6296	0.6020	0.5796	0.5600	0.5433	0.5290
-1.1600	0.06919	0.7350	0.7320	0.7238	0.7117	0.6971	0.6653	0.6353	0.6091	0.5868	0.5679	0.5518
-1.1400	0.07245	0.7718	0.7686	0.7617	0.7483	0.7332	0.6993	0.6667	0.6378	0.6132	0.5923	0.5745
-1.1200	0.07586	0.8045	0.8016	0.7937	0.7816	0.7663	0.7314	0.6967	0.6657	0.6390	0.6162	0.5968
-1.1000	0.07944	0.8332	0.8305	0.8231	0.8114	0.7965	0.7613	0.7256	0.6926	0.6641	0.6397	0.6188
-1.0800	0.08318	0.8581	0.8556	0.8488	0.8381	0.8237	0.7890	0.7525	0.7185	0.6885	0.6627	0.6405
-1.0600	0.08710	0.8796	0.8774	0.8712	0.8613	0.8480	0.8144	0.7779	0.7431	0.7120	0.6850	0.6616
-1.0400	0.09121	0.8980	0.8960	0.8906	0.8816	0.8694	0.8375	0.8015	0.7665	0.7346	0.7066	0.6823
-1.0200	0.09551	0.9137	0.9120	0.9072	0.8993	0.8882	0.8583	0.8234	0.7885	0.7562	0.7275	0.7023
-1.0000	0.10000	0.9271	0.9256	0.9214	0.9144	0.9045	0.8770	0.8436	0.8091	0.7767	0.7475	0.7218
-0.9800	0.10472	0.9384	0.9371	0.9335	0.9274	0.9187	0.8936	0.8623	0.8284	0.7962	0.7668	0.7415
-0.9600	0.10965	0.9479	0.9468	0.9438	0.9385	0.9309	0.9083	0.8787	0.8463	0.8145	0.7851	0.7586
-0.9400	0.11482	0.9559	0.9550	0.9524	0.9479	0.9413	0.9212	0.8937	0.8628	0.8317	0.8025	0.7759
-0.9200	0.12023	0.9627	0.9619	0.9597	0.9559	0.9502	0.9324	0.9072	0.8779	0.8478	0.8190	0.7925
-0.9000	0.12590	0.9684	0.9677	0.9659	0.9626	0.9577	0.9422	0.9193	0.8918	0.8628	0.8346	0.8083
-0.8800	0.13183	0.9732	0.9726	0.9711	0.9683	0.9642	0.9506	0.9300	0.9044	0.8767	0.8492	0.8233
-0.8600	0.13805	0.9773	0.9766	0.9755	0.9731	0.9696	0.9579	0.9394	0.9158	0.8895	0.8629	0.8375
-0.8400	0.14455	0.9807	0.9802	0.9791	0.9772	0.9742	0.9641	0.9478	0.9260	0.9012	0.8757	0.8509
-0.8200	0.15137	0.9836	0.9832	0.9823	0.9816	0.9781	0.9695	0.9550	0.9352	0.9121	0.8876	0.8635
-0.8000	0.15850	0.9860	0.9856	0.9849	0.9835	0.9814	0.9740	0.9613	0.9434	0.9218	0.8986	0.8754
-0.7800	0.16597	0.9880	0.9877	0.9871	0.9859	0.9841	0.9779	0.9668	0.9507	0.9307	0.9088	0.8865
-0.7600	0.17379	0.9898	0.9895	0.9991	0.9880	0.9865	0.9812	0.9716	0.9571	0.9388	0.9182	0.8968
-0.7400	0.18198	0.9912	0.9910	0.9906	0.9897	0.9885	0.9840	0.9757	0.9628	0.9460	0.9268	0.9065
-0.7200	0.19056	0.9925	0.9923	0.9919	0.9912	0.9901	0.9863	0.9792	0.9678	0.9526	0.9348	0.9154
-0.7000	0.19954	0.9936	0.9933	0.9930	0.9925	0.9916	0.9883	0.9822	0.9722	0.9584	0.9418	0.9236
-0.6800	0.20894	0.9945	0.9943	0.9940	0.9935	0.9928	0.9900	0.9848	0.9760	0.9636	0.9482	0.9312
-0.6600	0.21879	0.9952	0.9951	0.9948	0.9944	0.9938	0.9915	0.9870	0.9794	0.9682	0.9541	0.9382
-0.6400	0.22910	0.9959	0.9957	0.9955	0.9952	0.9946	0.9927	0.9889	0.9823	0.9723	0.9594	0.9445
-0.6200	0.23990	0.9964	0.9963	0.9961	0.9958	0.9954	0.9937	0.9905	0.9848	0.9759	0.9642	0.9504
-0.6000	0.25121	0.9969	0.9968	0.9966	0.9964	0.9960	0.9946	0.9919	0.9869	0.9791	0.9684	0.9557

Table 2.5-19

Oligotrophic Probabilities for Normal Error Distribution of X

Coefficient of Variation of X

$\log_{10}X$	X	0.0	.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
-2.2000	0.00631	0.9955	0.9953	0.9948	0.9942	0.9931	0.9899	0.9800	0.9610	0.9287	0.8849	0.8511
-2.1800	0.00661	0.9941	0.9939	0.9934	0.9925	0.9913	0.9873	0.9765	0.9485	0.9233	0.8783	0.8431
-2.1600	0.00692	0.9925	0.9922	0.9916	0.9905	0.9889	0.9842	0.9723	0.9511	0.9163	0.8700	0.8340
-2.1400	0.00724	0.9903	0.9901	0.9892	0.9879	0.9860	0.9803	0.9673	0.9447	0.9086	0.8610	0.8239
-2.1200	0.00759	0.9877	0.9873	0.9863	0.9847	0.9824	0.9756	0.9612	0.9372	0.8997	0.8509	0.8126
-2.1000	0.00794	0.9843	0.9839	0.9827	0.9807	0.9779	0.9699	0.9540	0.9285	0.8896	0.8395	0.8002
-2.0000	0.00832	0.9801	0.9795	0.9781	0.9757	0.9724	0.9630	0.9456	0.9185	0.8782	0.8269	0.7867
-2.0600	0.00871	0.9748	0.9742	0.9724	0.9696	0.9657	0.9548	0.9357	0.9071	0.8654	0.8131	0.7720
-2.0400	0.00912	0.9683	0.9675	0.9654	0.9621	0.9575	0.9448	0.9243	0.8941	0.8512	0.7979	0.7562
-2.0200	0.00955	0.9602	0.9593	0.9569	0.9530	0.9477	0.9336	0.9112	0.8796	0.8356	0.7816	0.7394
-2.0000	0.01000	0.9504	0.9494	0.9466	0.9421	0.9360	0.9202	0.8963	0.8635	0.8186	0.7640	0.7216
-1.9800	0.01047	0.9384	0.9373	0.9341	0.9289	0.9222	0.9048	0.8795	0.8457	0.8002	0.7454	0.7029
-1.9600	0.01096	0.9243	0.9227	0.9191	0.9134	0.9061	0.8873	0.8618	0.8264	0.7835	0.7257	0.6833
-1.9400	0.01148	0.9067	0.9053	0.9014	0.8952	0.8872	0.8675	0.8402	0.8054	0.7596	0.7050	0.6631
-1.9200	0.01202	0.8863	0.8848	0.8806	0.8740	0.8656	0.8452	0.8177	0.7830	0.7376	0.6835	0.6422
-1.9000	0.01259	0.8623	0.8608	0.8564	0.8497	0.8411	0.8206	0.7934	0.7592	0.7145	0.6613	0.6209
-1.8800	0.01318	0.8346	0.8331	0.8288	0.8221	0.8137	0.7937	0.7674	0.7342	0.6906	0.6384	0.5992
-1.8600	0.01380	0.8030	0.8015	0.7975	0.7913	0.7834	0.7647	0.7398	0.7081	0.6659	0.6152	0.5772
-1.8400	0.01445	0.7674	0.7662	0.7627	0.7572	0.7503	0.7336	0.7108	0.6731	0.6437	0.5916	0.5551
-1.8200	0.01514	0.7281	0.7271	0.7245	0.7202	0.7147	0.7006	0.6809	0.6553	0.6077	0.5678	0.5330
-1.8000	0.01585	0.6853	0.6848	0.6832	0.6805	0.6767	0.6665	0.6496	0.6250	0.5818	0.5440	0.5109
-1.7800	0.01660	0.6397	0.6396	0.6394	0.6386	0.6373	0.6313	0.6178	0.5964	0.5632	0.5232	0.4890
-1.7600	0.01738	0.5919	0.5924	0.5937	0.5951	0.5958	0.5947	0.5856	0.5597	0.5373	0.4966	0.4674
-1.7400	0.01820	0.5429	0.5439	0.5468	0.5504	0.5539	0.5579	0.5533	0.5390	0.5115	0.4734	0.4462
-1.7200	0.01906	0.4935	0.4951	0.4996	0.5055	0.5116	0.5210	0.5211	0.5105	0.4862	0.4505	0.4254
-1.7000	0.01995	0.4448	0.4469	0.4527	0.4608	0.4695	0.4844	0.4892	0.4825	0.4612	0.4281	0.4051
-1.6800	0.02089	0.3975	0.4000	0.4073	0.4171	0.4282	0.4483	0.4578	0.4553	0.4369	0.4063	0.3853
-1.6600	0.02188	0.3524	0.3552	0.3632	0.3748	0.3881	0.4131	0.4272	0.4282	0.4132	0.3850	0.3662
-1.6400	0.02291	0.3102	0.3132	0.3217	0.3346	0.3496	0.3791	0.3976	0.4022	0.3902	0.3645	0.3476
-1.6200	0.02399	0.2712	0.2742	0.2831	0.2967	0.3131	0.3465	0.3690	0.3771	0.3680	0.3447	0.3298
-1.6000	0.02512	0.2357	0.2386	0.2475	0.2614	0.2788	0.3154	0.3415	0.3530	0.3466	0.3256	0.3126
-1.5800	0.02630	0.2036	0.2065	0.2151	0.2290	0.2469	0.2860	0.3154	0.3299	0.3261	0.3073	0.2961
-1.5600	0.02754	0.1750	0.1777	0.1859	0.1995	0.2175	0.2584	0.2906	0.3079	0.3066	0.2898	0.2803
-1.5400	0.02884	0.1497	0.1522	0.1598	0.1728	0.1906	0.2326	0.2671	0.2870	0.2879	0.2731	0.2652
-1.5200	0.03020	0.1275	0.1297	0.1367	0.1489	0.1662	0.2087	0.2451	0.2672	0.2702	0.2572	0.2508
-1.5000	0.03162	0.1081	0.1101	0.1164	0.1277	0.1442	0.1866	0.2245	0.2485	0.2534	0.2420	0.2371
-1.4800	0.03311	0.0912	0.0930	0.0987	0.1090	0.1245	0.1663	0.2053	0.2309	0.2375	0.2277	0.2241
-1.4600	0.03468	0.0766	0.0782	0.0832	0.0926	0.1071	0.1479	0.1874	0.2144	0.2225	0.2141	0.2116
-1.4400	0.03631	0.0640	0.0654	0.0698	0.0782	0.0915	0.1310	0.1708	0.1910	0.2084	0.2012	0.2000
-1.4200	0.03802	0.0532	0.0544	0.0583	0.0657	0.0780	0.1156	0.1556	0.1846	0.1951	0.1891	0.1893
-1.4000	0.03981	0.0439	0.0450	0.0484	0.0550	0.0661	0.1021	0.1415	0.1711	0.1826	0.1777	0.1785
-1.3800	0.04169	0.0360	0.0369	0.0399	0.0457	0.0557	0.0898	0.1286	0.1587	0.1710	0.1669	0.1686
-1.3600	0.04365	0.0293	0.0301	0.0327	0.0378	0.0468	0.0788	0.1168	0.1471	0.1631	0.1568	0.1592
-1.3400	0.04571	0.0237	0.0244	0.0266	0.0311	0.0390	0.0690	0.1061	0.1363	0.1498	0.1472	0.1504
-1.3200	0.04787	0.0190	0.0196	0.0215	0.0254	0.0324	0.0603	0.0962	0.1264	0.1403	0.1383	0.1421
-1.3000	0.05012	0.0151	0.0156	0.0173	0.0206	0.0268	0.0526	0.0873	0.1172	0.1314	0.1299	0.1343
-1.2800	0.05248	0.0118	0.0123	0.0137	0.0166	0.0220	0.0450	0.0792	0.1088	0.1231	0.1220	0.1269
-1.2600	0.05496	0.0092	0.0096	0.0108	0.0133	0.0180	0.0399	0.0719	0.1009	0.1154	0.1146	0.1199
-1.2400	0.05755	0.0071	0.0074	0.0085	0.0106	0.0147	0.0346	0.0652	0.0937	0.1082	0.1077	0.1133
-1.2200	0.06026	0.0055	0.0057	0.0066	0.0083	0.0119	0.0301	0.0593	0.0871	0.1015	0.1013	0.1071
-1.2000	0.06310	0.0042	0.0044	0.0051	0.0065	0.0096	0.0261	0.0538	0.0810	0.0953	0.0952	0.1013
-1.1800	0.06607	0.0031	0.0033	0.0039	0.0051	0.0077	0.0227	0.0490	0.0754	0.0895	0.0895	0.0958
-1.1600	0.06919	0.0024	0.0025	0.0030	0.0039	0.0061	0.0197	0.0445	0.0702	0.0841	0.0841	0.0907
-1.1400	0.07245	0.0018	0.0019	0.0022	0.0030	0.0049	0.0171	0.0406	0.0655	0.0791	0.0791	0.0859
-1.1200	0.07586	0.0013	0.0014	0.0017	0.0023	0.0038	0.0149	0.0370	0.0611	0.0744	0.0743	0.0814
-1.1000	0.07944	0.0010	0.0010	0.0013	0.0018	0.0031	0.0129	0.0338	0.0571	0.0733	0.0699	0.0772
-1.0800	0.08318	0.0007	0.0008	0.0009	0.0013	0.0024	0.0113	0.0309	0.0534	0.0659	0.0657	0.0732
-1.0600	0.08710	0.0005	0.0006	0.0007	0.0010	0.0019	0.0098	0.0283	0.0500	0.0621	0.0618	0.0696
-1.0400	0.09121	0.0004	0.0004	0.0005	0.0008	0.0014	0.0085	0.0259	0.0469	0.0586	0.0582	0.0661
-1.0200	0.09551	0.0003	0.0003	0.0004	0.0006	0.0011	0.0075	0.0238	0.0440	0.0553	0.0548	0.0629
-1.0000	0.10001	0.0002	0.0002	0.0003	0.0004	0.0009	0.0066	0.0219	0.0413	0.0522	0.0516	0.0599
-0.9800	0.11472	0.0002	0.0002	0.0002	0.0003	0.0007	0.0058	0.0201	0.0389	0.0493	0.0486	0.0571
-0.9600	0.10965	0.0001	0.0001	0.0002	0.0002	0.0005	0.0051	0.0186	0.0366	0.0467	0.0459	0.0544
-0.9400	0.11482	0.0001	0.0001	0.0001	0.0002	0.0004	0.0046	0.0171	0.0345	0.0442	0.0434	0.0518
-0.9200	0.12023	0.0001	0.0001	0.0001	0.0001	0.0003	0.0040	0.0159	0.0326	0.0420	0.0410	0.0493
-0.9000	0.12590	0.0001	0.0001	0.0001	0.0001	0.0002	0.0036	0.0147	0.0308	0.0399	0.0388	0.0469
-0.8800	0.13183	0.0000	0.0000	0.0000	0.0001	0.0002	0.0032	0.0136	0.0291	0.0336	0.0368	0.0445
-0.8600	0.13805	0.0000	0.0000	0.0000	0.0001	0.0001	0.0027	0.0126	0.0275	0.0361	0.0349	0.0422
-0.8400	0.14455	0.0000	0.0000	0.0000	0.0000	0.0001	0.0024	0.0117	0.0261	0.0345	0.0332	0.0399
-0.8200	0.15137	0.0000	0.0000	0.0000	0.0000	0.0001	0.0023	0.0109	0.0247	0.0329	0.0316	0.0377
-0.8000	0.15850	0.0000	0.0000	0.0000	0.0000	0.0001	0.0019	0.0102	0.0234	0.0315	0.0301	0.0355
-0.7800	0.16597	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0095	0.0222	0.0302	0.0286	0.0334
-0.7600	0.17379	0.0000	0.0000	0.0000	0.0000	0.0000	0.0015	0.0089	0.0210	0.0289	0.0272	0.0314
-0.7400	0.18198	0.0000	0.0000	0.0000	0.0000	0.0000	0.0016	0.0083	0.0200	0.0277	0.0259	0.0294
-0.7200	0.19056	0.0000	0.0000	0.0000	0.0000	0.0000	0.0015	0.0077	0.0190	0.0265	0.0246	0.0276
-0.7000	0.19954	0.0000	0.0000	0.0000	0.0000	0.0000	0.0011	0.0073	0.0181	0.0254	0.0233	0.0259
-0.6800	0.20894	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0068	0.0172	0.0242	0.0222	0.0244
-0.6600	0.21879	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0064	0.0164	0.0231	0.0206	0.0230
-0.6400	0.22910	0.0000	0.0000	0.0000	0.0000	0.0000	0.0011	0.0060	0.0157	0.0220	0.0193	0.0218
-0.6200	0.23990	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0057	0.0151	0.0209	0.0179	0.0206
-0.6000	0.25121	0.0000	0.0000	0.0000	0.0000	0.0000	0.0009	0.0054	0.0144	0.0198	0.0165	0.0197

Table 2.5-20

Oligotrophic Probabilities for Lognormal Error Distribution of X
Coefficient of Variation of X

log₁₀X	X	0.0	.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
-2.2000	0.00631	0.9955	0.9951	0.9946	0.9937	0.9925	0.9877	0.9788	0.9693	0.9574	0.9453	0.9338
-2.1800	0.00661	0.9941	0.9937	0.9931	0.9920	0.9904	0.9847	0.9756	0.9639	0.9510	0.9382	0.9261
-2.1630	0.00692	0.9925	0.9920	0.9912	0.9899	0.9879	0.9811	0.9706	0.9577	0.9438	0.9303	0.9178
-2.1400	0.00724	0.9903	0.9898	0.9889	0.9872	0.9848	0.9767	0.9648	0.9506	0.9358	0.9217	0.9089
-2.1200	0.00759	0.9877	0.9871	0.9859	0.9839	0.9810	0.9714	0.9581	0.9426	0.9269	0.9123	0.8992
-2.1000	0.00794	0.9843	0.9836	0.9822	0.9798	0.9762	0.9651	0.9502	0.9336	0.9171	0.9021	0.8888
-2.0800	0.00832	0.9801	0.9793	0.9776	0.9746	0.9704	0.9577	0.9413	0.9235	0.9064	0.8913	0.8776
-2.0600	0.00871	0.9748	0.9739	0.9718	0.9683	0.9634	0.9490	0.9310	0.9122	0.8946	0.8790	0.8657
-2.0400	0.00912	0.9683	0.9673	0.9648	0.9607	0.9549	0.9387	0.9194	0.8998	0.8818	0.8662	0.8530
-2.0230	0.00955	0.9602	0.9591	0.9562	0.9514	0.9447	0.9269	0.9063	0.8860	0.8679	0.8524	0.8395
-2.0000	0.01000	0.9504	0.9491	0.9457	0.9402	0.9327	0.9132	0.8916	0.8710	0.8529	0.8377	0.8252
-1.9800	0.01047	0.9384	0.9369	0.9331	0.9268	0.9185	0.8976	0.8753	0.8546	0.8368	0.8221	0.8101
-1.9633	0.01096	0.9240	0.9223	0.9180	0.9110	0.9020	0.8800	0.8574	0.8369	0.8196	0.8056	0.7943
-1.9400	0.01148	0.9067	0.9049	0.9001	0.8925	0.8829	0.8602	0.8377	0.8177	0.8013	0.7881	0.7777
-1.9200	0.01202	0.8863	0.8843	0.8791	0.8711	0.8611	0.8383	0.8162	0.7972	0.7819	0.7698	0.7603
-1.9000	0.01259	0.8623	0.8603	0.8548	0.8466	0.8365	0.8140	0.7931	0.7754	0.7614	0.7506	0.7423
-1.8800	0.01318	0.8346	0.8325	0.8270	0.8188	0.8089	0.7876	0.7682	0.7523	0.7400	0.7306	0.7235
-1.8600	0.01380	0.8033	0.8009	0.7956	0.7877	0.7785	0.7593	0.7418	0.7283	0.7179	0.7098	0.7042
-1.8400	0.01445	0.7674	0.7655	0.7606	0.7535	0.7453	0.7284	0.7139	0.7026	0.6942	0.6883	0.6842
-1.8200	0.01514	0.7281	0.7264	0.7222	0.7163	0.7095	0.6959	0.6846	0.6761	0.6701	0.6661	0.6636
-1.8000	0.01585	0.6853	0.6840	0.6809	0.6765	0.6715	0.6618	0.6541	0.6487	0.6453	0.6434	0.6426
-1.7800	0.01660	0.6397	0.6389	0.6370	0.6344	0.6316	0.6263	0.6225	0.6205	0.6198	0.6201	0.6211
-1.7600	0.01738	0.5919	0.5916	0.5912	0.5906	0.5902	0.5897	0.5902	0.5916	0.5938	0.5964	0.5993
-1.7400	0.01820	0.5429	0.5432	0.5442	0.5458	0.5478	0.5523	0.5573	0.5623	0.5674	0.5724	0.5771
-1.7200	0.01906	0.4935	0.4943	0.4969	0.5007	0.5051	0.5146	0.5240	0.5327	0.5407	0.5480	0.5548
-1.7000	0.01995	0.4448	0.4461	0.4503	0.4558	0.4626	0.4769	0.4936	0.5029	0.5138	0.5235	0.5322
-1.6800	0.02089	0.3975	0.3992	0.4043	0.4118	0.4207	0.4395	0.4573	0.4731	0.4869	0.4990	0.5095
-1.6600	0.02188	0.3524	0.3545	0.3605	0.3694	0.3801	0.4028	0.4244	0.4435	0.4601	0.4744	0.4868
-1.6400	0.02291	0.3102	0.3125	0.3191	0.3291	0.3411	0.3671	0.3922	0.4143	0.4335	0.4500	0.4642
-1.6200	0.02399	0.2712	0.2736	0.2806	0.2912	0.3041	0.3323	0.3607	0.3856	0.4072	0.4257	0.4417
-1.6000	0.02512	0.2357	0.2381	0.2451	0.2559	0.2694	0.2999	0.3332	0.3576	0.3813	0.4018	0.4194
-1.5800	0.02630	0.2036	0.2060	0.2129	0.2236	0.2373	0.2689	0.3010	0.3303	0.3560	0.3782	0.3973
-1.5600	0.02754	0.1750	0.1772	0.1838	0.1943	0.2077	0.2397	0.2730	0.3040	0.3314	0.3551	0.3756
-1.5400	0.02884	0.1497	0.1518	0.1579	0.1678	0.1808	0.2125	0.2465	0.2787	0.3075	0.3326	0.3543
-1.5200	0.03020	0.1275	0.1294	0.1350	0.1442	0.1566	0.1875	0.2216	0.2545	0.2844	0.3106	0.3335
-1.5000	0.03162	0.1081	0.1098	0.1149	0.1233	0.1348	0.1645	0.1983	0.2316	0.2622	0.2893	0.3131
-1.4800	0.03311	0.0912	0.0927	0.0973	0.1050	0.1155	0.1436	0.1766	0.2099	0.2409	0.2688	0.2934
-1.4600	0.03468	0.0766	0.0779	0.0820	0.0889	0.0985	0.1247	0.1566	0.1895	0.2207	0.2490	0.2742
-1.4400	0.03631	0.0640	0.0652	0.0688	0.0749	0.0836	0.1078	0.1382	0.1704	0.2015	0.2301	0.2557
-1.4200	0.03802	0.0532	0.0542	0.0574	0.0628	0.0705	0.0927	0.1214	0.1526	0.1833	0.2120	0.2379
-1.4000	0.03981	0.0439	0.0448	0.0476	0.0524	0.0592	0.0794	0.1062	0.1362	0.1663	0.1947	0.2207
-1.3800	0.04169	0.0360	0.0368	0.0392	0.0434	0.0493	0.0676	0.0924	0.1210	0.1503	0.1784	0.2044
-1.3600	0.04365	0.0293	0.0300	0.0321	0.0358	0.0411	0.0573	0.0801	0.1071	0.1354	0.1630	0.1888
-1.3400	0.04571	0.0237	0.0243	0.0261	0.0293	0.0340	0.0483	0.0692	0.0945	0.1216	0.1485	0.1739
-1.3200	0.04787	0.0190	0.0195	0.0211	0.0239	0.0279	0.0406	0.0594	0.0830	0.1088	0.1349	0.1599
-1.3000	0.05012	0.0151	0.0155	0.0169	0.0193	0.0228	0.0339	0.0508	0.0726	0.0971	0.1222	0.1466
-1.2800	0.05248	0.0118	0.0122	0.0134	0.0155	0.0185	0.0282	0.0433	0.0633	0.0863	0.1104	0.1340
-1.2600	0.05496	0.0092	0.0096	0.0106	0.0123	0.0149	0.0233	0.0367	0.0550	0.0765	0.0994	0.1223
-1.2400	0.05755	0.0071	0.0074	0.0083	0.0097	0.0120	0.0192	0.0310	0.0476	0.0675	0.0893	0.1113
-1.2200	0.06026	0.0055	0.0057	0.0064	0.0076	0.0095	0.0157	0.0261	0.0413	0.0594	0.0799	0.1010
-1.2000	0.06310	0.0042	0.0043	0.0049	0.0060	0.0075	0.0128	0.0218	0.0352	0.0521	0.0714	0.0915
-1.1800	0.06607	0.0031	0.0033	0.0038	0.0046	0.0059	0.0103	0.0182	0.0301	0.0456	0.0635	0.0826
-1.1600	0.06919	0.0024	0.0025	0.0029	0.0035	0.0046	0.0083	0.0151	0.0256	0.0397	0.0564	0.0744
-1.1400	0.07245	0.0018	0.0019	0.0022	0.0027	0.0036	0.0067	0.0125	0.0217	0.0345	0.0499	0.0669
-1.1200	0.07586	0.0013	0.0014	0.0016	0.0021	0.0028	0.0053	0.0102	0.0184	0.0299	0.0440	0.0599
-1.1000	0.07944	0.0010	0.0010	0.0012	0.0016	0.0021	0.0042	0.0084	0.0155	0.0258	0.0388	0.0536
-1.0800	0.08318	0.0007	0.0008	0.0009	0.0012	0.0016	0.0033	0.0068	0.0130	0.0221	0.0340	0.0478
-1.0600	0.08710	0.0005	0.0006	0.0007	0.0009	0.0012	0.0026	0.0055	0.0108	0.0190	0.0297	0.0425
-1.0400	0.09121	0.0004	0.0004	0.0005	0.0007	0.0009	0.0020	0.0045	0.0090	0.0162	0.0259	0.0377
-1.0200	0.09551	0.0003	0.0003	0.0004	0.0005	0.0007	0.0016	0.0036	0.0075	0.0138	0.0226	0.0334
-1.0000	0.10001	0.0002	0.0002	0.0003	0.0004	0.0005	0.0012	0.0029	0.0062	0.0117	0.0196	0.0295
-0.9800	0.10472	0.0002	0.0002	0.0002	0.0003	0.0004	0.0009	0.0023	0.0051	0.0099	0.0169	0.0259
-0.9600	0.10965	0.0001	0.0001	0.0002	0.0002	0.0003	0.0007	0.0018	0.0042	0.0083	0.0146	0.0228
-0.9400	0.11482	0.0001	0.0001	0.0001	0.0001	0.0002	0.0005	0.0014	0.0034	0.0070	0.0125	0.0199
-0.9200	0.12023	0.0001	0.0001	0.0001	0.0001	0.0002	0.0004	0.0011	0.0028	0.0058	0.0107	0.0174
-0.9000	0.12590	0.0001	0.0001	0.0001	0.0001	0.0001	0.0003	0.0009	0.0022	0.0049	0.0092	0.0152
-0.8800	0.13183	0.0000	0.0000	0.0000	0.0001	0.0001	0.0002	0.0007	0.0018	0.0041	0.0078	0.0132
-0.8600	0.13805	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0014	0.0034	0.0066	0.0114
-0.8400	0.14455	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0012	0.0028	0.0056	0.0099
-0.8200	0.15137	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0023	0.0047	0.0085
-0.8000	0.15850	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0019	0.0040	0.0073
-0.7800	0.16597	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0005	0.0015	0.0033	0.0063
-0.7600	0.17379	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0012	0.0028	0.0054
-0.7400	0.18198	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0010	0.0023	0.0046
-0.7200	0.19056	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0019	0.0039
-0.7000	0.19954	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0016	0.0033
-0.6800	0.20894	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005	0.0013	0.0028
-0.6600	0.21879	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0011	0.0024
-0.6400	0.22910	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009	0.0020
-0.6200	0.23990	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0007	0.0017
-0.6000	0.25121	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0006	0.0014

Figure 2.5-.22

Contour Plots of Trophic State Probabilities vs. X and Coefficient
of Variation of X for Normal and Lognormal Error Distributions

2.5.6 Application Strategies (continued)

probability of eutrophic classification. Tables 2.5-21 and
2.5-22 have been derived by interpolating Tables 2.5-17 and
2.5-18, for the normal and log-normal distributions, res-
pectively. Rows in each table represent constant probability
levels, ranging from 0.01 to 0.2, as given in column one.
The second column indicates the values of X corresponding
to the various probabilities, assuming that X can be specified
exactly, i.e. $CV_X = 0.$. The remaining columns indicate the
percent reduction in the design values of X which would be
necessary to account for the increasing uncertainty levels.
The tables indicate that the cost of this uncertainty
generally increases with decreasing design probabilities.
Uncertainty costs are higher for the lognormal distribution.

Table 2.5-21

Effect of Uncertainty in X on Rational Design Values for Lognormally Distributed Errors

p(e)[a]	x_o[b]	Coefficient of Variation of X Estimate									
		.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
.01	.0237	-0.7[c]	-2.7	-5.7	-9.7	-19.2	-29.1	-38.2	-46.1	-52.8	-58.4
.02	.0264	-0.6	-2.4	-5.2	-8.8	-17.4	-26.3	-34.7	-42.1	-48.4	-53.7
.05	.0309	-0.5	-2.0	-4.4	-7.4	-14.5	-21.9	-28.9	-35.2	-40.7	-45.4
.10	.0354	-0.4	-1.7	-3.5	-5.9	-11.5	-17.3	-22.9	-28.0	-32.5	-36.4
.20	.0414	-0.3	-1.0	-2.3	-3.8	-7.2	-10.8	-14.3	-17.5	-20.3	-22.8

Table 2.5-22

Effect of Uncertainty in X on Rational Design Values for Normally Distributed Errors

p(e)[a]	x_o[b]	Coefficient of Variation of X Estimate									
		.05	.10	.15	.20	.30	.40	.50	.60	.70	.80
.01	.0237	-0.6[c]	-2.3	-4.9	-8.0	-14.9	-21.7	-27.8	-33.3	-38.2	-42.4
.02	.0264	-0.5	-2.1	-4.4	-7.3	-13.7	-20.1	-26.1	-31.5	-36.2	-40.4
.05	.0309	-0.5	-1.8	-3.7	-6.2	-11.8	-17.6	-23.1	-28.1	-32.6	-36.7
.10	.0354	-0.4	-1.4	-3.0	-5.0	-9.7	-14.7	-19.6	-24.1	-28.3	-32.2
.20	.0414	-0.2	-0.8	-1.8	-3.1	-6.3	-9.9	-13.7	-17.4	-20.9	-24.2

a - p(e) = probability of eutrophic classification

b - X_o = design value of X, assuming it can be specified without error; $X = L [QS (1 + .824T^{-.454})]^{-.815}$

c - percent change in X_o required to account for given level of uncertainty

2.5.7 Implications for Monitoring Program Design

Quantifying the effects of uncertainty in the decision
variable upon the rational design provides a means of rational-
izing the need for additional data acquisition. These effects
can be expressed in economic terms, i.e. in terms of the cost
associated with the additional nutrient source abatement
required to account for the uncertainty. Typically, nutrient
abatement entails substantial economic investment (e.g. pipelines
and treatment plants) or losses (e.g. restrictions in land develop-
ment). The marginal costs attributed to the uncertainty can be
compared with the monitoring and analytical investments required
in order to reduce that uncertainty, i.e. to obtain a better
loading, residence time, or mean depth estimate. Because the
abatement costs are generally on a much higher scale than moni-
toring costs, an overall optimization would probably justify
additional source monitoring or other measurements, until the
uncertainty in the decision variable, X , is low enough so
that its effect on the design is low, or until the flat portions
of the countour lines in Figure 2.5-22 are reached. In this
region, the coefficient of variation of $\hat{X}$ is about 0.15 or
less, and the effect of the assumed error distribution is small.
This approach toward rationalizing the need for monitoring by
expressing it in terms of the design costs of uncertainty has
been explored by Meta Systems, Inc.[61] in a study for the Depart-
ment of Transportation, which involved an examination of the

2.5.7 Implications for Monitoring Program Design (continued)

potential effects of additional streamflow gauging sites upon
the rational design of highway culverts.

Tables 2.5-21 and 2.5-22 indicate that if $\hat{X}$ can be defined
such that its coefficient of variation is less than about 0.15,
uncertainty costs would generally be less than 5% in the appli-
cation of the discriminant model for loading allocations at the
given design probability levels. Usually, most of the uncertainty
in X would be attributed to the loading or average inlet concen-
tration estimate. It would be of interest to consider the
practical implications of this variability objective upon actual
stream monitoring requirements.

To provide a partial basis for assessing the potential impacts
of the Cross-Florida Barge Canal, Meta Systems[62] developed nutrient
budgets on various reaches of two major drainage basins in north-
central Florida: the Oklawaha and the Withlacoochee. For each
sampling station an attempt was made to quantify both the mean
and the variance of the nutrient flux estimate on yearly and
long-term-average time scales. The techniques, assumptions, and
limitations of the estimates have been discussed previously
(Section 2.3.2).

Figure 2.5-23 is a log-log plot of the coefficients of
variation of the total phosphorus flux estimates against the
numbers of concentration samples employed in deriving the estimates.

Figure 2.5-23

Coefficient of Variation of Mean Phosphorus Flux Estimate
Against Sample Size for Various Stations in the Cross-
Florida Barge Canal Study[62]
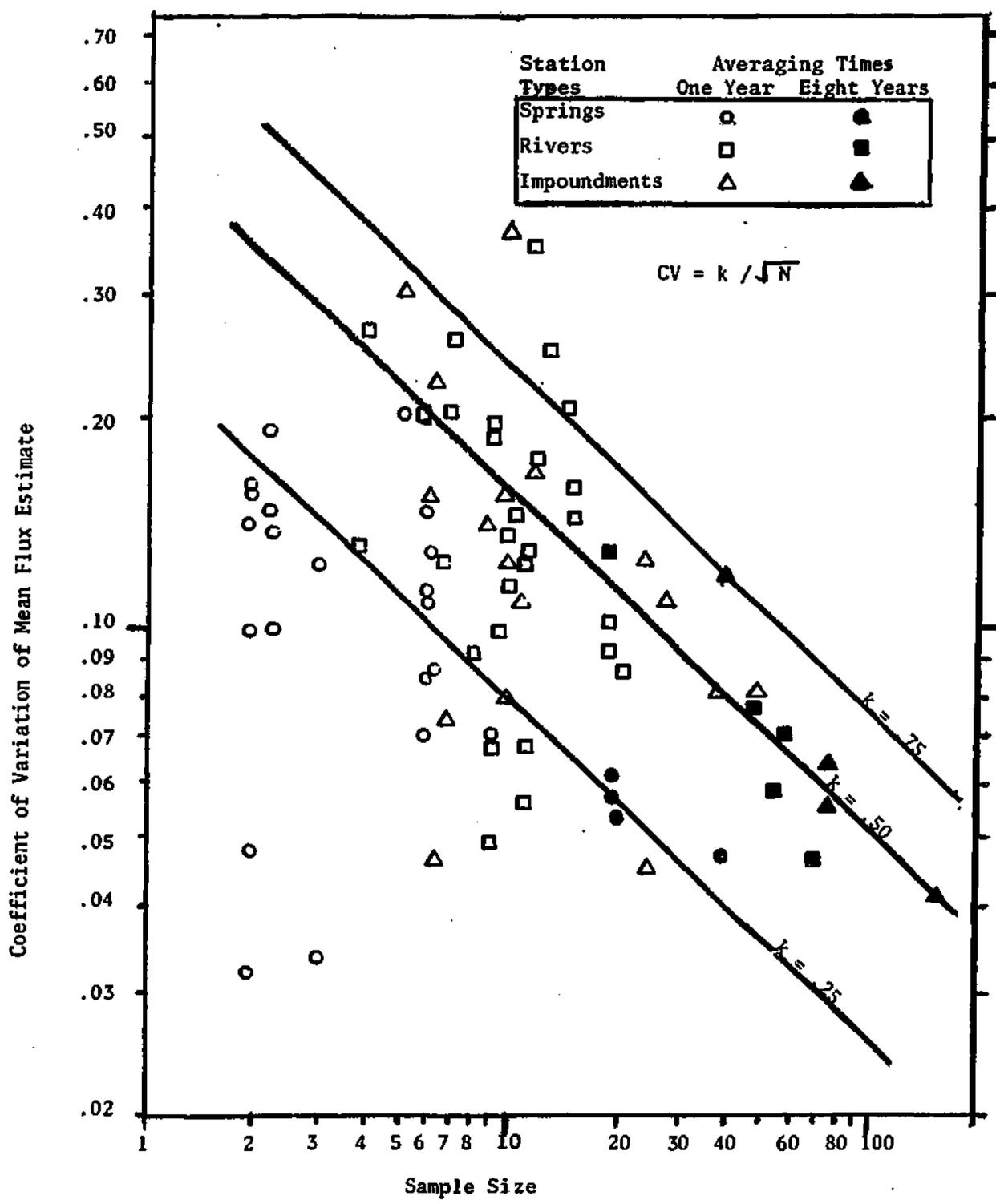
2.5.7 Implications for Monitoring Program Design (continued)

Regardless of the particular technique employed to estimate the average flux from the continuous flow and grab-sample concentration data, the standard error of the mean varies inversely as the square root of the effective concentration sample size (Table 2.3-2). Because sampling frequencies were generally less than one per month, the effect of autocorrelation upon the effective sample size was assumed to be negligible (Figure 2.3-1). The lines in Figure 2.5-23 have slopes of -1/2 and, accordingly, are lines of constant variability in the sampled stations. The relative variability is defined as k , the product of the coefficient of variation of the mean flux estimate and the square root of the sample size. In the figure, three types of stations have been distinguished: springs, river stations, and impoundment stations. Their respective locations on the plot reflect generally increasing variability. In order to define the mean flux to within a given coefficient of variation, fewer samples are required for the spring stations than for the impoundment stations. Solid symbols were derived from long-term-average flux estimates and open symbols, from yearly estimates. Because of sampling error in the variance, as well as the mean, the yearly estimates generally exhibit more scatter.

A k value of about 0.5 seems to characterize most of the open river and impoundment stations. At this level of variability, about 11 samples would yield a coefficient of variation of 0.15 in

2.5.7 Implications for Monitoring Program Design (continued)

the mean. A similar analysis of NES tributary or other stream
data would bs required in order to determine whether the variabili-
ties and resulting sampling requirements based upon the Florida
data are generally applicable. Of particular interest would be
the relationships between flux variability and such factors as
flow, flow variability, and drainage basin characteristics.

The coefficient of variation of a total loading estimate
derived from independent flux estimates can be calculated from:

$$\overline{CV}_T^2 = \sum_{i=1}^{M} \frac{\overline{CV}_i^2 \, \ell_i^2}{(\sum_{i=1}^{M} \ell_i)^2} \qquad (2.5\text{-}64)$$

If the individual fluxes are estimated with the same accuracy,
$\overline{CV}_*$:

$$\overline{CV}_T^2 = \overline{CV}_*^2 \frac{\Sigma \ell_i^2}{(\Sigma \ell_i)^2} \qquad (2.5\text{-}65)$$

Accordingly, the following bounds can be set upon the ratio of
$\overline{CV}_T$ to $\overline{CV}_*$:

2.5.7 <u>Implications for Monitoring Program Design</u> (continued)

$$\frac{1}{\sqrt{M}} \leq \frac{\overline{CV}_T}{\overline{CV}_*} \leq 1 \qquad\qquad (2.5-66)$$

The lower limit corresponds to the case in which all ℓ_i are equal. This analysis indicates that if the estimates of the individual fluxes all satisfy the maximum variability requirement, then the total loading estimate will at least satisfy that requirement.

Assuming that an average variability of 0.5 is typical, a monthly sampling frequency appears to be adequate to define a loading or inlet concentration estimate for application of this particular model. This applies only to tributaries not dominated by point sources. This result may reflect the fact that the model itself was derived primarily from nutrient balances which had been estimated based upon monthly sampling. This re-emphasizes the need to appreciate that these results are both model- and data-specific.

In view of these results, the interpretation of the probabilities generated directly from the discriminant model (i.e. $p(e:X)$ in equation 2.5-54) warrants additional discussion. To some extent, the effects of uncertainty in X have already been included in this function, since errors were present in the X estimates of

2.5.7 Implications for Monitoring Program Design (continued)

the data set employed to estimate the parameters of the

classification model. A more accurate definition of $p(e : X)$

would be: " the probability of a eutrophic classification,

given an X estimate which had been generated using the same

types and amounts of data and the same techniques employed in

developing the phosphorus budgets in the original data set" .

This may account for the apparently small effect of additional

uncertainty in X on the classification probabilities for

coefficients of variation less than 0.15. In this range, the

added uncertainty may be small compared to that already present.

The results of the error analysis conducted on the phosphorus

retention model (Section 2.5.3) indicated that coefficients of

variation in the range of 0.10 may be typical of the inlet and

outlet phosphorus concentration estimates in the data set.

The major question that remains is: what would happen

to the shape of $p(e : X)$ if better estimates of X were

available? The separation of the trophic states along the

X axis in Figure 2.5-19 may become more distinct and the

design levels of X to achieve a given probability of

eutrophic classification may become higher. Hence, uncer-

tainty costs may already be present in the X_o values of

Tables 2.5-21 and 2.5-22. These costs could only be quantified

by refining the data or using other lakes with more accurate

nutrient budget estimates. The net result is that the actual

2.5.7 Implications for Monitoring Program Design (continued)

probability of eutrophic classification for a given X value
may be lower than that calculated according to the above
scheme, if the model were applied to a lake with a loading
estimate which is more accurate than those of the original
data set. Hence, in such a case, the scheme would yield a
conservative design value for phosphorus loading. Because
of the potential economic costs associated with this conser-
vatism, a thorough re-examination and refinement of the data
and techniques employed in estimating nutrient budgets may be
justified. This would reduce the errors in the independent
variables employed in the discriminant analysis, and any
resultant uncertainty costs.

2.5.8 Summary

The availability of phosphorus balance and trophic state data on a variety of lakes has permitted analysis and comparison of existing empirical models for predicting phosphorus retention coefficients and lake trophic states. A preliminary analysis of the data revealed significant stratification of all morphologic, hydrologic, and phosphorus-related variables considered across trophic states. A high degree of multicollinearity was apparent, indicating that cause-effect relationships would be difficult to establish and that application of the results could only be properly done to lakes which conformed to the multivariate distribution of the data analyzed.

A theoretically based model for phosphorus retention was proposed and estimated. The effective first-order decay coefficient of phosphorus was found to be roughly inversely proportional to the square root of lake hydraulic residence time and independent of depth. Effects of incomplete mixing, temperature variations, and the coupling of phosphorus dynamics with various, flushing-dependent, physical, chemical, and biological processes were suggested as possible explanations for this behavior. The two parameters of the model were found to be stable across trophic states. The standard errors of estimate were substantially higher in eutrophic lakes, however. This model compared favorably with strictly empirical ones derived from linear regression and with the equation suggested by Kirchner and Dillon[42].

2.5.8 Summary (continued)

An error analysis indicated that residual errors* dominated over potential parameter and independent variable errors in predictions of the retention coefficient. On this basis, it was concluded that, for this type of model, the size of the data set was adequate for parameter estimation purposes. An attempt was made to disect the residual error into model and measurement errors. The former was shown to increase with hydraulic residence time; the latter, to decrease. The results indicated that coefficients of variations of about 0.1 may characterize measurement errors in the average inlet and outlet phosphorus concentration estimates.   It was also shown that, as hydraulic residence time decreased, the accuracy of outlet concentration predictions became increasingly dependent upon the accuracy of the inlet concentration (or loading) estimate, and therefore, that refinement of the model would have little impact upon the total prediction error at low residence times. At high residence times, application of a more sophisticated model would be justified, providing that its model and parameter errors were lower. Evidence was presented which suggested that the factor limiting the accuracy of a chlorophyll prediction from phosphorus loading, residence time, and depth would be in the empirical chlorophyll-lake phosphorus relationship, and not in the lake phosphorus prediction. Thus, it was concluded that the accuracy of the phosphorus retention model was sufficient for application with state-of-the-art chlorophyll models.

---

* model and measurement errors

2.5.8 Summary (continued)

Discriminant analysis was employed to compare existing
models for trophic state prediction and to derive an optimal
linear classification model for this data set. With the
parameters of the models optimized, the average percentages
of misclassified lakes were 32.3%, 28.3%, and 22.0% for
Vollenweider's first model (L versus Z)[93], Dillon's model
(CO)[18], and Vollenweider's second model (L versus QS)[95],
respectively. Optimization of the parameters for this data
set had little effect on the classification errors obtained
when the originally-reported parameter values of the respec-
tive models were used. Evidence suggested that the performance
of Dillon's model may have been hampered by systematic errors
in the retention coefficient estimates of the National Eutro-
phication Survey Lakes.

Stepwise discriminant analysis selected the outlet con-
centration estimate based upon the phosphorus retention model
discussed above and the total phosphorus loading as components
of an optimal linear classification model for these lakes.
After these variables had entered, no significant discrimination
powers remained in the other variables. This model misclassified
an average of 15.9% of the lakes, with most of the errors
centered in the mesotrophic group. A principal component
analysis was employed to reduce the classification model to
a univariate one, which further reduced average classification

2.5.8 <u>Summary</u> (continued)

error to 14.5%. Misclassifications were discussed as possibly
relating to errors in the dependent and independent variables
and to unsteady-state conditions in the lakes.

In applications, it was shown that the discriminant model
could be used to generate probabilities of trophic state
classification as a function of a single variable, which could
be computed from phosphorus loading, residence time, and mean
depth estimates. Means of incorporating the effects of un-
certainty in this "decision" variable on the classification
probabilities and upon the rational loading allocation for a
given lake were derived. In real applications, uncertainty in
the loading estimates would generally contribute most to the
uncertainty in the decision variable. The effect of this
uncertainty upon the rational loading allocation which satisfies
a given probability of eutrophic classification was demonstrated
as a means of justifying additional monitoring effort. These
effects were found to be small, provided that the coefficient
of variation of the decision variable estimate was less than
about 0.15. This, in turn, was shown to be possibly typical
of a flux estimate derived from monthly tributary sampling,
as employed by the NES. Because of possible effects of errors
in the original data set, it was suggested that application of
the classification model to lakes with nutrient budgets which
had been more accurately determined than those in the original

2.5.8 Summary (continued)

data set would yield conservative design values for phosphorus
loadings.

The empirical modelling approach taken in this section has been
based upon five principle assumptions : (1) mass balance; (2) complete-
ly mixed conditions; (3) phosphorus limitation of lake ecosystems;
(4) first-order kinetics for phosphorus sedimentation; (5) the possible
influence of lake morphometric and hydrologic factors upon nutrient
dynamics and trophic state response. The accuracy and value of the
models developed have been shown to be limited by various features of
the data base, including measurement or estimation errors in the inde-
pendent and dependent variables, multicollinearity, nonsteady-state
conditions existing in the lakes during sampling, and the subjectivity
involved in the original lake classifications. In this context, it is
difficult to identify model deficiencies due to effects of other con-
trolling factors or aspects of system behavior which have been ignored,
such as vertical stratification, sediment releases, or limitation
by light of other nutrients. In short, data of this sort have not
provided a very good basis for model discrimination. Because of these
aspects, the general superiority of the developed models cannot be
claimed. The primary emphases are upon the approaches taken and tech-
niques employed in their development and evaluation and upon the strate-
gies proposed for their application. In these regards, the underlying
theme is in the development and use of error estimates.

2.5.9 Suggestions for Future Work

This section has illustrated some potentially useful approaches to empirical modelling. One general concept to consider is that these models are at most as good as the data used to derive them. Improvement in the quality of the data could improve model accuracy and, in applications, enhance design efficiencies. "Data errors" do not result from measurement errors alone. The selection of techniques for data reduction can also be critical. Use of an improved set of techniques could conceivably permit improvement in data quality without having to resort to additional measurements.

One potential area for technique improvement is in the development of lake nutrient budgets. The assumptions employed in estimating mass fluxes from grab-sample concentration data have been discussed in Section 2.3.2. The likely introduction of errors due to the methods employed by the National Eutrophication Survey in converting these budgets to normal hydrologic years has also been considered in Section 2.3.4. Essentially three independent and indirect means have been used to estimate the extent of data errors inherent in the lake nutrient budgets. Analysis of the phosphorus retention coefficient residuals in Section 2.5.3 indicated that coefficients of variation of about 0.10 may be typical of errors in the mean inlet and outlet phosphorus

2.5.9 <u>Suggestions for Future Work</u> (continued)

concentration estimates. The results of the Cross-Florida
Barge Canal Study[62] indicated that monthly tributary sampling
would yield annual flux estimates with coefficients of
variation of about 0.15, for streams of the same general
characteristics as those included in that study. It was
also shown that the coefficient of variation of a total
loading estimate derived from the sum of individual stream
fluxes would generally be less than that of the individual
flux estimates. Finally, trophic state probabilities generated
by the discriminant model were found to be relatively insensi-
tive to errors in the phosphorus loading estimates for coeffi-
cients of variation of about 0.15 or less. Direct analysis
of concentration and flow data on a stream-by-stream basis
would be required in order to develop prior, direct estimates
of these errors. Such an analysis, combined with an improved
strategy for hydrologic year corrections, could conceivably
improve the quality of the nutrient budget data without
resorting to additional measurements.

Another area for improvement is in the lake trophic
state classification. In the above analysis, trophic state
classifications had been originally given by somewhat subjective
interpretations of lake response data. Errors at this stage
of the analysis could conceivably be reduced by developing and
employing a more objective classification scheme. The general

2.5.9 <u>Suggestions for Future Work</u> (continued)

approach would be to develop a continuous response index, similar to Shannon and Brezonik's "Trophic State Index"[73]. The index would include variables which could be related directly to established beneficial-use criteria. One approach would divide the lakes into discrete states corresponding to various beneficial uses, based upon the index and the established criteria. Discriminant analysis could then be used to separate the beneficial-use states based upon stimulus variables. Alternatively, the continuous index could be preserved and directly related to the stimulus variables using regression techniques, as employed by Shannon and Brezonik[74]. Either approach would provide a model which could be used as a basis for probabilistic designs, analogous to those discussed in Section 2.5.6. This work would provide a more objective model basis and reduce dependent variable (classification) errors. Again, these improvements would be possible without additional data acquisition.

2.6 General Comments and Conclusions

Policy decisions in lake water quality management are reached through a complex process involving a variety of environmental, economic, political, and technologic factors. This chapter has focused upon some aspects of the technologic functions which can be employed to provide water quality impact assessments. In making such projections, the engineer or planner must make numerous decisions, often ranging from data acquisition to model selection and parameter estimation. The variety of methods discussed indicates that there is considerable latitude in this area and that there is no uniquely correct path from the definition of the analytical problem to its solution. One hopes that the policy recommendations reached by the planner are independent of his analytic decisions, i.e. of the particular set of data and methods employed. This is subject to the constraint that the monitoring and modelling are properly implemented in each case. The robustness of a particular analysis can be tested by comparing results derived from alternative, but apparently equally-valid approaches.

Some of the factors discussed in this chapter can be used as a basis for making analytic decisions. A proper balance between monitoring and modelling effort is a key objective in this regard. In an ideal situation, the costs of the total analytic effort might be balanced by the design costs associated with uncertainty in the analysis, as discussed in Section 2.5.7.

2.6 <u>General Comments and Conclusions</u> (continued)

More often, total analytic (monitoring and modelling) effort
is constrained. In this case, a strategy would be to try to
allocate monitoring and modelling resources so as to minimize
projection uncertainty or to minimize expected regret associated
with projection errors. Still more often, the analyst does not
have control over monitoring, i.e., the data are given. In
this case, the selection and/or development of models would
again be guided by data availability.

Two general types of models have been discussed in this
chapter as relating to analysis of lake water quality problems:
source models and lake models. Source models can be used to
relate various watershed or point-source characteristics to
nutrient loadings. Some of these characteristics may be
uncontrollable; others may be decision variables. If the
various types of errors associated with the projections of
these models were sufficiently small, they could be used as a
substitute for direct monitoring data. As discussed in Section
2.3.2, the state-of-the-art of these models is such that direct
estimation of nutrient fluxes via monitoring is generally
preferable. The development of these models is essential to
understanding and eventual control of non-point sources and
their effects on lake water quality.

Lake models predict the response of lake water quality

2.6 General Comments and Conclusions (continued)

to a given set of boundary conditions. They have been roughly classified as "empirical" and "theoretical" in this chapter. Some bases for model selection are discussed below.

The applicability of a given model may be governed in part by the management criteria of concern. The more empirical models tend to predict long-term responses and, as such, apply to long-term criteria. Some of the theoretical models can be used to predict system response under short-term, critical conditions, which are often of management concern. An example would be the response of hypolimnic dissolved oxygen levels to a critical set of meteorologic conditions. The short- and long-term responses are generally correlated, in that oligotrophic and eutrophic lakes each behave in characteristic ways. Thus, the more empirical models can be used to predict short-term responses indirectly, while theoretical system models can predict such directly.

Another basis for model selection is accuracy. The quantification of various error components involved in an empirical model projection has been demonstrated in Section 2.5.3. Such an analysis can be used straightforwardly to balance monitoring and modelling efforts. The relative complexities of the "theoretical" models would render a corresponding error analysis difficult, though not impossible.

2.6 General Comments and Conclusions (continued)

One of the factors limiting such an analysis would probably be the accuracy of the estimates of the parameter covariance matrix. This matrix would be used along with the sensitivity coefficient matrix to estimate the effects of parametric uncertainty upon projection errors, according to the methods and equations discussed in Chapter 1. The covariance matrix could be estimated straightforwardly if the parameter estimates were derived directly from system response data, as in the case of the phosphorus retention model developed and analyzed in Sections 2.5.2 and 2.5.3. In more complex models, parameter estimates have generally been derived from independent laboratory studies. Estimation of parameter covariances in these cases would be more subjective. Even if the experiments provided parameter distributions, there would be uncertainty as to whether the experimental conditions adequately represented "field" conditions. Nevertheless, it would be possible to approximate the various error components of a projection derived from a theoretical model through a systematic assessment of parameter distributions and comparisons of estimated and observed system responses. This would permit straightforward comparisons of models with regard to accuracy.

A key result obtained from the analysis of the phosphorus retention coefficient residuals in Section 2.5.3 is that model errors would be detectable in only about one-third of the lakes,

2.6 General Comments and Conclusions (continued)

if measurement strategies and data reduction techniques similar
to those employed by the National Eutrophication Survey were
used. Thus, for about two-thirds of the lakes in the analysis,
there would be little benefit from employing a more accurate
model than the one developed, providing that the objective were
to predict outlet phosphorus concentration. Another key result
of this analysis is that potential model errors increase with
hydraulic residence time. This would indicate a corresponding
increase in the importance of accuracy as a criterion for model
selection. A similar error analysis would indicate whether
this were true in the cases of other water quality components.

Another basis for model selection is generality. A model
should be general enough to be valid for both present and
future, or "design" states of a system under study. The former
is necessary for calibration and verification, the latter for
application purposes. The parameter estimates and, in some
cases, the functional forms of empirical models are based
directly upon system response data. These models are at most
as general as the data used to derive them. Thus, the nature
of the data base and characteristics and dependencies of model
performance both inside and outside of the data base are keys
to assessing empirical model generality. The more elaborate
theoretical models are supposedly independent of particular
system response data, because both their formulations and their

2.6 <u>General Comments and Conclusions</u> (continued)

parameter estimates are usually derived independently, Thus,
if the formulations and parameters estimates are correot, a
theoretical model could be constructed to have high generality.
However, as discussed in Section 2.4.2, this generality has
not as yet been demonstrated. It has not been shown that a
single model with a unique set of parameter values can be an
equally valid representation of lakes at different trophic
stages. The evidence presented in Seotion 2.4.2 suggested
that this lack of generality might be reduced by increasing
complexity, specifically to incorporate many species at each
level of the aquatic food chain. However, this would
exponentially increase parameter estimation difficulties
which are already formidable.

The net result of these considerations is that theoretical
models still need to be tailored to specific applications.
This creates an empirical link between the "theoretical" model
and system response data and reduces the supposed gap in
generality between the "empirical" and "theoretical" types
of models. Thus, while there are certainly some general
aspects of the behavior of thase systems, analysis of lake
water quality problems is still a relatively empirical exercise,
entailing both monitoring and modelling of the systems and their
boundary conditions.

REFERENCES - CHAPTER 2

1. Aitchison, J. and J.A.C. Brown, The Lognormal Distribution, Cambridge University Press, 1963.

2. Ashton, P.M. and R.C. Underwood, ed., Non-Point Sources of Water Pollution, Proceedings of a Southeastern Regional Conference, Virginia Water Resources Center, September 1975.

3. Bannister, T.T., "Production Equations in Terms of Chlorophyll Concentration, Quantum Yield, and Upper Limit of Production", Limnology and Oceanography, Vol. 20, No. 3, pp. 342-364, May 1975.

4. Bayley, G.J. and J.M. Hammersly, "The Effective Number of Independent Observations in an Autocorrelated Time Series", Journal of the Royal Statistical Society, Vol. 8 (1B), pp. 184-197, 1946.

5. Benjamin, J.R. and C.A. Cornell, Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill Book Co., 1970.

6. Bierman, V.J., F.H. Verhoff, T.L. Poulson, and M.W. Tenney, "Multi-nutrient Dynamic Models of Algal Growth and Species Competition in Eutrophic Lakes" in Modeling the Eutrophication Process, Ann Arbor Science, 1974.

7. Blanton, J.A., "Vertical Entrainment into the Epilimnion of Strati-fied Lakes", Limnology and Oceanography, Vol. 18, No. 5, pp. 697-704, September 1973.

8. Bormann, F.H., G.E. Likens, T.G. Siciama, R.S. Pierce, and J.S. Eaton, "The Export of Nutrients and Recovery of Stable Conditions Following Deforestation of Hubbard Brook", Ecological Monographs, Vol. 44, No. 3, pp. 255-277, Summer 1974.

9. Cahill, T.H., P. Imperato, and F.H. Verhoff, "Evaluation of Phosphorus Dynamics in a Watershed", Journal of the Environmental Engineering Division, American Society of Civil Engineers, Vol. 100, No. EE2, pp. 439-454, April 1974.

10. Chen, C.W. and G.T. Orlob, "Ecologic Simulation for Aquatic Environments", Final Report, OWRR C-2044, Water Resources Engineers, Walnut Creek, California, 1972.

11. Cooke, G.E., J.N. Bhargara, M.R. McCemas, M.C. Wilson, and R.J. Heath, "Some Aspects of Phosphorus Dynamics of the Twin Lakes Watershed", in Modeling the Eutrophication Process, Ann Arbor Science, 1974.

REFERENCES - CHAPTER 2 (continued)

12. Cooley, W.W. and P.R. Lohnes, Multivariate Data Analysis, John Wiley & Sons, New York, 1971.

13. Correll, D.L., J.W. Pierce, and M.A. Faust, "A Quantitative Study of the Nutrient, Sediments and Coliform Bacterial Constituents of Water Runoff from the Rhode River Watershed", in ref. 2, 1975.

14. Dillon, P.J. and P.H. Rigler, "The Phosphrous-Chlorophyll Relationship in Lakes", Limnology and Oceanography, Vol. 19, No. 1, pp. 767-773, September 1974.

15. Dillon, P.J., "A Critical Review of Vollenweider's Nutrient Budget Model and Other Related Models", Water Resources Bulletin, American Water Works Association, Vol. 10, No. 5, pp. 969-989, October 1974.

16. Dillon, P.J. and F.H. Rigler, "A Test of a Simple Nutrient Budget Model for Predicting the Phosphorus Concentration in Lake Water", Journal of Fisheries Board Research of Canada, Vol. 13, pp. 1771-1778, 1974.

17. Dillon, P.J., "A Manual for Calculating the Capacity of a Lake for Development", Limnology & Toxicity Section, Water Res. Branch, Ontario Ministry of the Environment, October 1974.

18. Dillon, P.J., "The Phosphorus Budget of Cameron Lake, Ontario: The Importance of Flushing Rate to the Degree of Eutrophy of Lakes", Limnology and Oceanography, Vol. 20, No. 1, pp. 28-45, January 1975.

19. Dillon, P.J. and W.B. Kirchner, "The Effects of Geology and Land Use on the Export of Phosphorus from Watersheds", Water Research, Vol. 9, pp. 135-148, 1975.

20. DiToro, D.M., D.J. O'Connor, and R.V. Thomann, "A Dynamic Model of Phytoplankton in Natural Waters", Environmental Engineering and Science Program, Manhattan College, New York, June 1970.

21. DiToro, D.M., D.J. O'Connor, and R.V. Thomann, "A Dynamic Model of the Phytoplankton Population in the Sacramento-San Joaquin Delta", in Nonequilibrium Systems in Natural Waters, Adventures in Chemistry Series, No. 106, American Chemical Society, pp. 131-180, 1971.

22. DiToro, D.M., D.J. O'Connor, R.V. Thomann, and J.R. Mancini, "Preliminary Phytoplankton-Zooplankton-Nutrient Model of Western Lake Erie", Systems Analysis and Simulation in Ecology, Vol. 3, Academic Press, 1973.

REFERENCES - CHAPTER 2 (continued)

23. DiToro, D.M., D.J. O'Connor, and R.V. Thomann, "Preliminary Model of Potomac Estuary Phytoplankton", Journal of the Environmental Engineering Division, American Society of Civil Engineers, Vol. 100, No. EE2, pp. 699-715, June 1974.

24. DiToro, D.M., D.J. O'Connor, R.V. Thomann, and R.P. Winfield, "Mathematical Modeling of Phytoplankton in Lake Ontario 1. Model Development and Verification", EPA-660/3-75-005, March 1975.

25. DiToro, D.M. and W.F. Matystik, "Phytoplankton Biomass Model of Lake Huron and Saginaw Bay", in ref. 90, 1976.

26. Dixon, W.J., ed., BMDP - Biomedical Computer Programs, University of California Press, 1975.

27. Enviro Control, Inc., "National Assessment of Trends in Water Quality", National Technical Information Service, Springfield, Virginia, June 1971.

28. Frink, C.R., "Septic Tanks as Non-Point Sources", presented at the Lakes Region Planning Commission Workshop on Lake Management and Related Non-Point Source Pollution Control, Weirs Beach, New Hampshire, May 1976.

29. Grenney, W.J., D.A. Bella, and H.C. Curl, "A Theoretical Approach to Interspecific Competition in Phytoplankton Communities", American Naturalist, Vol. 107, No. 955, May-June 1974.

30. Grenney, W.J., D.A. Bella, and H.C. Curl, "Effects of Intracellular Nutrient Pools on Growth Dynamics of Phytoplankton", Journal of the Water Pollution Control Federation, Vol. 46, No. 7, pp. 1751-1760, July 1974.

31. Guy, H.P., "An Overview of Non-Point Water Pollution from the Urban-Suburban Area", in ref. 2, 1975.

32. Harms, L.L., J.N. Dornbush, and J.R. Andersen, "Physical and Chemical Quality of Agricultural Land Runoff", Journal of the Water Pollution Control Federation, Vol. 46, No. 11, November 1974.

33. Hetling, L.J. and R.M. Sykes, "Sources of Nutrients in Canadarago Lake", Journal of the Water Pollution Control Federation, Vol. 45, No. 1, pp. 145-156, January 1973.

34. Hobbie, J.E. and G.E. Likens, "Output of Phosphorus, Dissolved Organic Carbon, and Fine Particulate Carbon from Hubbard Brook Watershed", Limnology and Oceanography, Vol. 18, No. 4, pp. 734-742, July 1973.

REFERENCES - CHAPTER 2 (continued)

35. Hoel, P.G., <u>Introduction to Mathematical Statistics</u>, Fourth Edition, Wiley & Sons, 1971.

36. Holmes, R.W., "The Secchi Disc in Turbid Coastal Waters", <u>Limnology and Oceanography</u>, Vol. 15, No. 5, pp. 688-694, September 1970.

37. Huck, P.M. and G.J. Farquhar, "Water Quality Models Using the Box-Jenkins Method", <u>Journal of the Sanitary Engineering Division, American Society of Civil Engineers</u>, Vol. 100, No. EE3, pp. 735-753, June 1974.

38. Hutchinson, G.E., "Eutrophication: The Scientific Background of a Contemporary Practical Problem", <u>American Scientist</u>, Vol. 61, pp. 269-279, May-June 1973.

39. Imboden, D.M., "Phosphorus Model of Lake Eutrophication", <u>Limnology and Oceanography</u>, Vol. 19, No. 2, pp. 297-304, March 1974.

40. Johnson, N.M., G.E. Likens, F.H. Bormann, D.W. Fisher, and R.S. Pierce, "A Working Model for the Variation in Stream Water Chemistry at the Hubbard Brook Experimental Forests, New Hampshire", <u>Water Resources Research</u>, Vol. 5, pp. 1353-1363, 1969.

41. Keup, L.E., "Phosphorus in Flowing Waters", <u>Water Research</u>, Vol. 2., pp. 373-386, 1968.

42. Kirchner, W.B. and P.J. Dillon, "An Emprical Method of Estimating the Retention of Phosphorus in Lakes", <u>Water Resources Research</u>, Vol. 11, No. 1, pp. 182-183, February 1975.

43. Kirchner, W.B., "An Examination of the Relationship Between Drainage Basin Morphology and the Export of Phosphorus", <u>Limnology and Oceanography</u>, Vol. 20, No. 2, pp. 267-270, March 1975.

44. Kilhus, S.P., J.D. LaPerriere, and R.W. Bachman, "Nutrients and Algae in Some Central Iowa Streams", <u>Journal of the Water Pollution Control Federation</u>, Vol. 47, No. 7, pp. 1870-1874, July 1975.

45. Larsen, D.P. and H.T. Mercier, "Lake Phosphorus Loading Graphs: An Alternate", US EPA, National Eutrophication Survey, Working Paper No. 174, 1975.

46. Lassiter, R.R. and D.K. Kearns, "Phytoplankton Population Changes and Nutrient Fluctuations in a Simple Aquatic Ecosystem Model" in <u>Modeling the Eutrophication Process</u>, Ann Arbor Science, 1974.

47. Lassiter, R.R., "Modeling Dynamics of Biological and Chemical Components of Aquatic Ecosystems", National Environmental Research

REFERENCES - CHAPTER 2 (continued)

Center, Office of Research and Development, EPA-660/3-75-012, May 1975.

48. Lee, G.F., "Factors Affecting the Transfer of Materials Between Water and Sediments", Literature Review No. 1, University of Wisconsin, Water Resources Center, 50 pp., July 1970.

49. Lehman, JJ., D.B. Bodkin, and G.E. Likens, "The Assumptions and Rationales of a Computer Model of Phytoplankton Population Dynamics", Limnology and Oceanography, Vol. 20, No. 3, pp. 342-264, May 1975.

50. Lettenmaier, D.P., "Detection of Trends in Water Quality Data from Records with Dependent Observations", Water Resources Research, Vol. 12, No. 5, pp. 1037-1046, October 1976.

51. Likens, G.E., F.H. Bormann, N.M. Johnson, D.W. Fisher, and R.S. Pierce, "Effects of Forest Cutting and Herbicide Treatment on Nutrient Budgets in the Hubbard Brook Watershed-Ecosystem", Ecological Monographs, Vol. 40, No. 1, pp. 23-47, Winter 1970.

52. Likens. G.E. and R.H. Borman, "Linkages Between Terrestrial and Aquatic Ecosystems", Bioscience, Vol. 24, No. 8, pp. 447-456, August 1974.

53. Lehr, R.C., "Characteristics and Comparative Magnitudes of Non-Point Sources", Journal of the Water Pollution Control Federation, Vol. 46, No. 8, pp. 1849-1872, August 1974.

54. Lorenzen, M.W. and R. Mitchell, "Theoretical Effects of Artificial Destratification in Impoundments", Environmental Science and Technology, Vol. 7, No. 10, pp. 939-944, October 1973.

55. Lorenzen, M.W. and R. Mitchell, "An Evaluation of Artificial Destratification for Control of Algal Blooms", Journal of the American Water Works Association, Vol. 67, p. 373, 1975.

56. Lorenzen, M.W., D.J. Smith, and L.V. Kimmel, "A Long Term Phosphorus Model for Lakes: Application to Lake Washington", presented before the Division of Environmental Chemistry, American Chemical Society, Philadelphia, April 1975.

57. Matalas, M.C. and W.B. Langbein, "Information Content of the Mean", Journal of Geophysical Research, Vol. 67, No. 9, pp. 3441-3448, 1962.

58. Mauleg, K.W., D.P. Larsen, D.W. Schultz, and H.T. Mercier, "A Six-Year Water, Phosphorus, and Nitrogen Budget for Shagawa Lake",

REFERENCES - CHAPTER 2 (continued)

Journal of Environmental Quality, Vol. 4, No. 2, pp. 236-242, 1975.

59. McMichael, F.C. and J.S. Hunter, "Stochastic Modeling of Temperature and Flow in Rivers", Water Resources Research, Vol. 8, No. 1, pp. 87-98, February 1972.

60. Mehta, B.M. and R.C. Ahlert, "Stochastic Variation of Water Quality of the Passaic River", Water Resources Research, Vol. 11, No. 2, pp. 300-308, April 1975.

61. Meta Systems Inc., "Assessment of National Small Rural Watersheds Program", prepared for the U.S. Department of Transporation, Federal Highway Administration, 1975.

62. Meta Systems Inc., "Cross-Florida Barge Canal Project - Nutrient Budget", for U.S. Army Corps of Engineers, Jacksonville District, Contract 17-75-C-0077, March 1976.

63. O'Melia, C.R., "An Approach to the Modeling of Lakes", Schweitz Zeitschrift fur Hydrologie, Vol. 34, pp. 1-33, 1972.

64. Park, R.A. et al., "A Generalized Model for Simulating Lake Ecosystems", Eastern Deciduous Forest Biome, U.S. International Biological Program, Contribution No. 152, May 1974.

65. Patten, B.C., "Need for Ecosystem Prospective in Eutrophication Modeling", in Modeling the Eutrophication Process, Ann Arbor Science, 1974.

66. Process Research, Inc., "The Impact of the Federal Water Pollution Control Act on the Charles River and Boston Harbor", Report to the National Commission on Water Quality, Washington, D.C., June 1975.

67. Process Research, Inc., "Final Report on the Storrow Lagoon Demonstration Project", for the Metropolitan District Commission, Commonwealth of Massachusetts, 1975.

68. Reckow, K., Environmental Systems Program, Harvard University, unpublished results, 1975.

69. Resource Planning Associates, "208 Modeling Approach - Working Paper", prepared for Lakes Region Planning Commission, Meredith, New Hampshire, August 1975.

70. Scavia, D., B.J. Eadie, and A. Robertson, "An Ecological Model for the Great Lakes", in ref. 90, 1976.

REFERENCES - CHAPTER 2 (continued)

71. Schindler, D.W., "Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management", Science, Vol. 184, pp. 897-899, May 1974.

72. Schofield, W.R. and R.G. Krutchkoff, "Deterministic Model of a Dynamic Eutrophic Estuary", Journal of the Environmental Engineering Division, American Society of Civil Engineers, Vol. 100, No. EE4, pp. 979-996, August 1974.

73. Shannon, E.E. and P.L. Brezonik, "Eutrophication Analysis: A Multivariate Approach", Journal of the Sanitary Engineering Division, American Society of Civil Engineers, Vol. 98, No. SA1, pp. 37-57, February 1972.

74. Shannon, E.E. and P.L. Breznok, "Relationships Between Lake Trophic State and Nitrogen and Phosphorus Loading Rates", Environmental Science and Technology, Vol. 6, No. 8, pp. 719-735, August 1972.

75. Sheldon, A.L., "A Quantitative Approach to the Classification of Inland Waters", in Natural Environments, Krutilla, J.V. (ed.), Johns Hopkins University Press, Baltimore, Maryland, pp. 205-261, 1972.

76. Smith, F.E., "Effects of Enrichment in Mathematical Models", in Eutrophication: Causes, Consequences, Correctives, National Academy of Sciences, Washington, 1969.

77. Snedecor, G.W. and W.G. Cochran, Statistical Methods, Sixth Edition, Iowa State University Press, 1967.

78. Snodgrass, W.J. and C.R. O'Melia, "A Predictive Model for Phosphorus in Lakes - Development and Testing", Ph.D. Thesis, Department of Environmental Sciences and Engineering, University of North Carolina, 1974.

79. Steele, R.J., "Environmental Control of Photosynthesis in the Sea", Limnology and Oceanography, Vol. 7, pp. 137-150, 1962.

80. Stumm, W. and J.A. Leckie, "Phosphate Exchange with Sediments; Its Role in the Productivity of Surface Waters", Advances in Water Pollution Research, San Francisco and Hawaii, Vol. 2, Paper III-26, 16 pp., 1970.

81. Sykes, R.M., "The Prediction of Lacustrine Trophic Status", Civil Engineering Department, Ohio State University, unpublished manuscript, 1975.

REFERENCES - CHAPTER 2 (continued)

82. Thomann, R.V., "Variability of Waste Treatment Plant Performance", Journal of the Sanitary Engineering Division, American Society of Civil Engineers, Vol. 96, No. SA3, pp. 819-837, June 1970.

83. Thomann, R.V. and R.P. Winfield, "On the Verification of a Three-Dimensional Phytoplankton Model of Lake Ontario", in ref. 90, 1976.

84. Tyler, J.E., "The Secchi Disc", Limnology and Oceanography, Vol. 13, No. 1, January 1968.

85. US EPA, "Proposed Criteria for Water Quality", Washington, D.C., 1973.

86. US EPA, Pacific Northwest Environmental Research Laboratory, National Environmental Research Center, Corvallis, Oregon, National Eutrophication Survey, Working Paper Series, 1972-75.

87. US EPA, "The Relationships of Phosphorus and Nitrogen to the Trophic State of Northeast and North-Central Lakes and Reservoirs", National Eutrophication Survey, Working Paper No. 23, December 1974.

88. US EPA, "An Approach to a Relative Trophic Index System for Classifying Lakes and Reservoirs", National Eutrophication Survey, Working Paper No. 24, 1974.

89. US EPA, "Relationships Between Drainage Area Characteristics and Non-Point-Source Nutrients in Streams", National Eutrophication Survey, Working Paper No. 25, Pacific Northwest Environment Research Lab, Corvallis, Oregon, August 1974.

90. US EPA, Environmental Modeling and Simulation, Proceedings of a Conference, Cincinnati, Ohio, April 1976.

91. Uttormark, P.D., J.D. Chapin, and K.M. Green, "Estimating Nutrient Loadings of Lakes from Non-Point Sources", Office of Research and Monitoring, US EPA, EPA-660/3-74-020, August 1974.

92. Uttormark, P.D. and J.P. Wall, "Lake Classification- A Trophic Characterization of Wisconsin Lakes", National Environmental Research Center, Office of Research and Development, US EPA, Ecological Research Series, EPA-660/3-75-003, June 1975.

93. Vollenweider, R.A., "Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, with Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication", Organization of Economic Cooperation and Development, Paris, September 1970.

94. Vollenweider, R.A., "Possibilities and Limits of Elementary Models Concerning the Budget of Substances in Lakes", Arch. Hydrobiol., Vol. 66, No. 1, pp. 1-36, April 1969.

REFERENCES - CHAPTER 2 (continued)

95. Vollenweider, R.A., "Input-Output Models", unpublished manuscript, 1973.

96. Wang, W.C. and R.L. Evans, "Dynamics of Nutrient Concentration in the Illinois River", Journal of the Water Pollution Control Federation, Vol. 42, p. 2117, 1970.

97. Weibel, S.R., R.J. Anderson, and R.L. Woodward, "Urban Land Runoff as a Factor in Stream Pollution", Journal of the Water Pollution Control Federation, Vol. 36, No. 7, pp. 914-924, 1964.

98. Welch, P.S., Limnology, First Edition, McGraw-Hill, New York and London, 1935.

3.0 EXPLORATORY ANALYSIS OF ONONDAGA LAKE WATER QUALITY DATA

3.1 Introduction

3.1.1 General Description

Onondaga Lake is situated at the northern edge of the City of Syracuse in Central New York State (Figure 3.1-1). It is a relatively small lake with a surface area of 11.7 km^2 and with average and maximum depths of 12 and 20.5 meters, respectively. The lake's 600 km^2 drainage basin (Figure 3.1-2) contains approximately 325,000 people and essentially all of Onondaga County's 140 industries[24]. It discharges to the Seneca River at an average rate of 19.3 cubic meters per second (W.Y. 1968-74) and has a mean hydraulic residence time of 84 days.

The lake's primary use has long been as a receptacle for municipal and industrial wastes. About 14% of the average outflow volume consists of effluent from Onondaga County's primary sewage treatment facility. A steel mill and a solvay process plant discharge thermal and chemical wastes directly into the lake or its immediate tributaries. Land use in the watershed is about 15% urban and 34% agricultural. About 35% of the urban land is served by combined sewers. Thus, the impact of diffuse sources may be appreciable, although such effects are currently screened by the predominance of point sources.

The location of the lake renders it a potentially valuable

Figure 3.1-1

Onondaga Lake
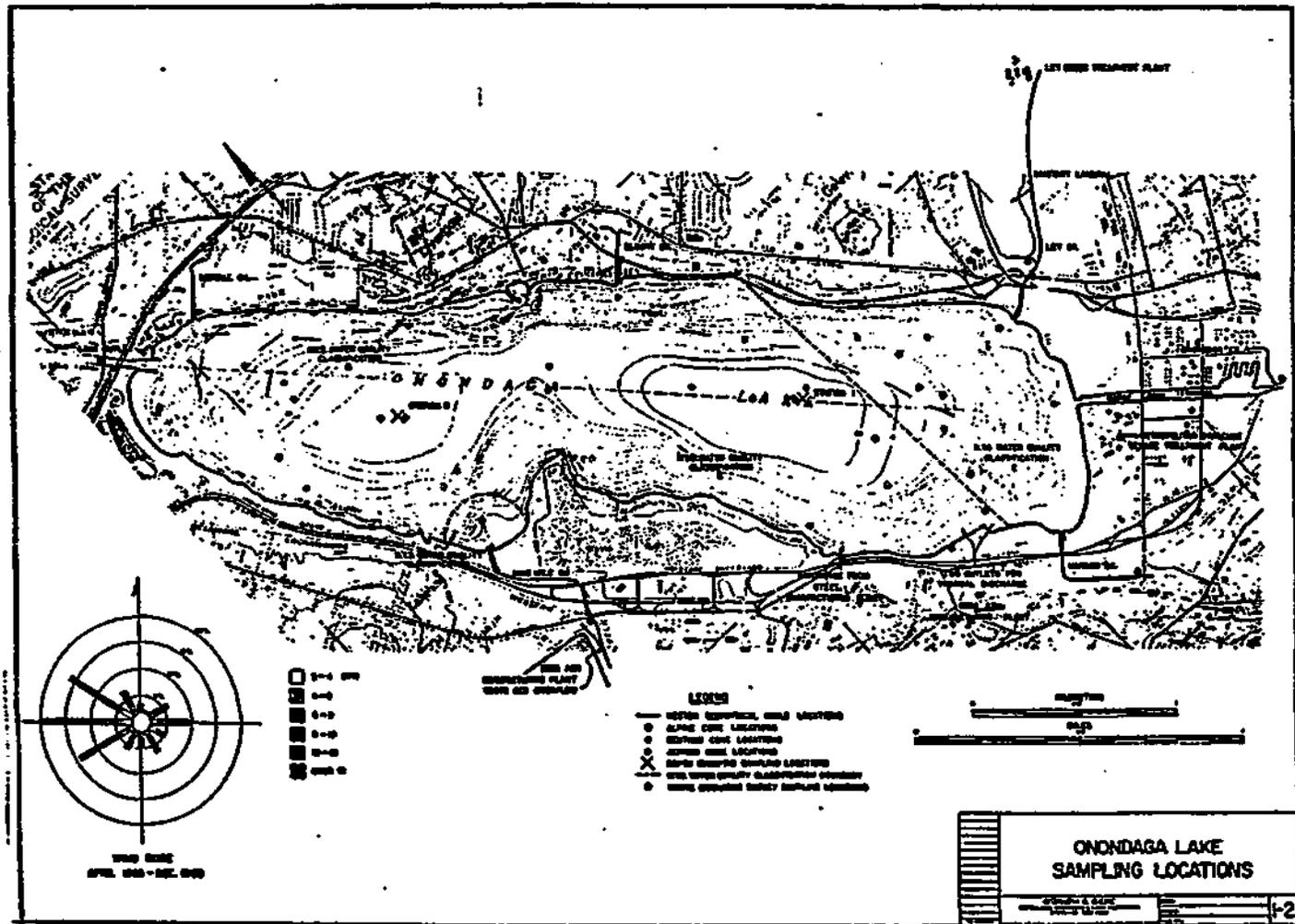
Figure 3.1-2
Onondaga Lake/Watershed

LEGEND

- ▓ BELOW 600'
- ▒ 600' TO 1200'
- ☐ 1200' TO 1800'
- ▓ ABOVE 1800'
- ⚫⚫⚫ DRAINAGE BASIN BOUNDARY

SCALE : MILES
0 1 2 3 4 5

ONONDAGA LAKE
CONTOURS
AND DRAINAGE BASIN

O'BRIEN & GERE

4-2
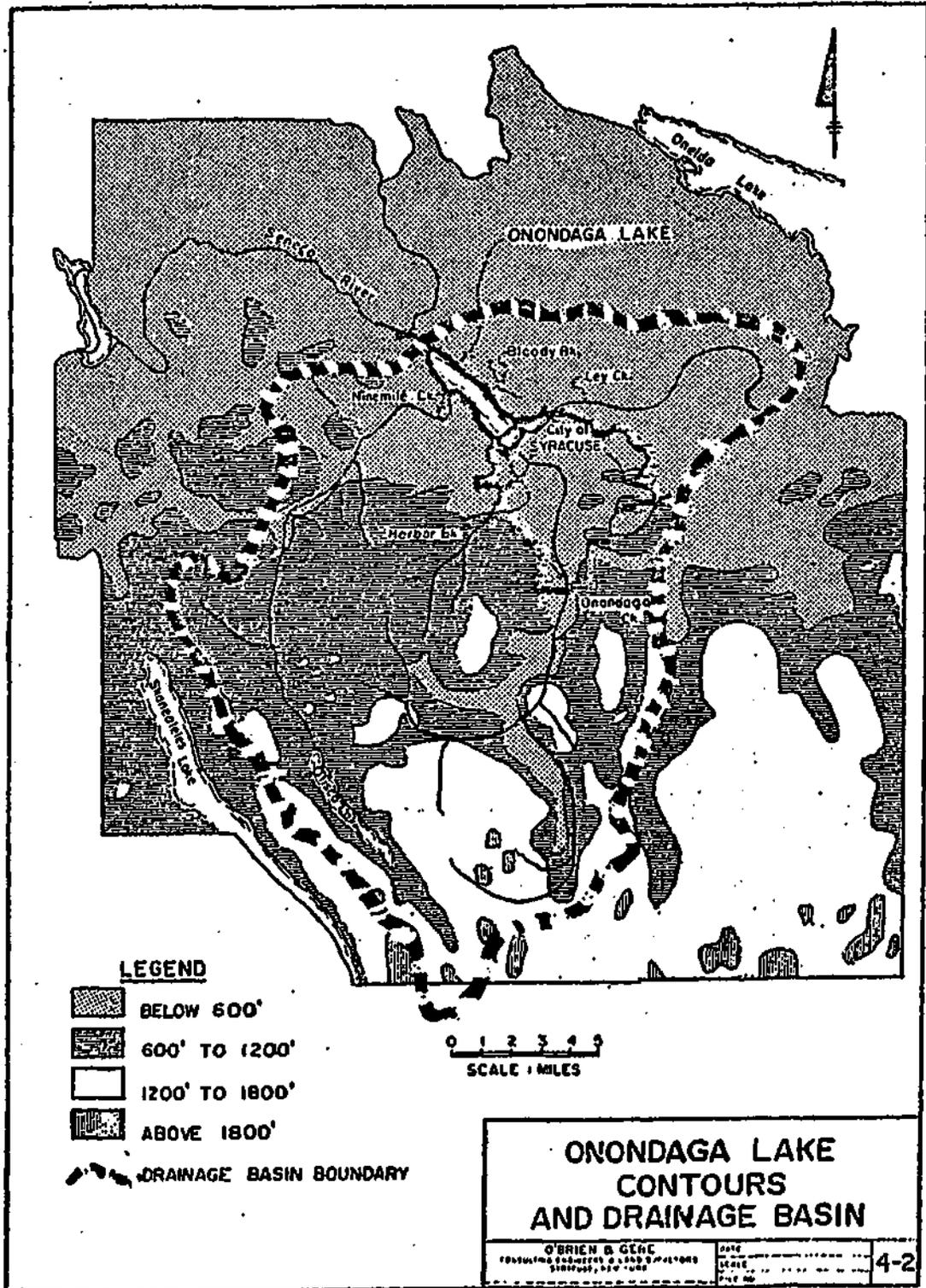
3.1.1 General Description (continued)

recreational asset to Onondaga County's residents. However, like most urban lakes and rivers, severe water quality problems have precluded realization of this potential. In addition to waste assimilation, the current beneficial uses of the lake are restricted to industry (cooling and process waters), transportation (the lake discharges to the New York State barge canal system), and some recreation (boating and picnicking along nothern shores). Fishing has been outlawed since 1970. The future use of Onondaga Lake for contact recreation is one of Onondaga County's major environmental goals and an issue of great public concern.

3.1.2 Water Quality Issues

The quality of Onondaga Lake water is unique[24]. A mean dissolved solids concentration of 3500 ppm and chloride content of 1700 ppm, corresponding roughly to 10% seawater, render it one of the most saline lakes in the Northeast. Most of the salinity is due to waste discharges from the Allied Chemical Company solvay process plant, although there is historical and geological evidence indicating that natural sources of salinity may also be appreciable. This aspect of the lake's water quality is thought to have important physical, chemical, and biological effects on struotural and functional aspects of the Onondaga Lake ecosystem. Physically, during stratified periods,

3.1.2 Water Quality Issues (continued)

density gradients are markedly enhanced, and vertical mixing hindered, as a result of vertical salinity gradients. The lake does undergo two periods of vertical circulation yearly, normal for lakes of this depth and at this latitude. Calcium and sodium, the major cations in the lake, are present in about equal mass concentrations. Chemical precipitation of calcium salts is marked, the lake being supersaturated with respect to calcium carbonate and calcium phosphate most of the year[24]. Biologically, salinity is thought to be an important factor in determining and regulating the dominant species of biota at various trophic levels, ranging from phytoplankton to fish.

Onondaga Lake has been classified as extremely eutrophic. The abundance of primary nutrients has permitted the development of algal blooms to objectionable levels. In 1968-69, the average concentration of total phosphorus in the epilimnion was 2.34 mg/l[24], over one hundred times that considered "dangerous" from a eutrophication standpoint[39]. During the same period, time- and volume-averaged dissolved oxygen was 30% of saturation and the hypolimnion of the lake was anaerobic between early March and late October.

Potential toxic effects due to extreme levels of ammonia, chromium, copper, and mercury have also been of concern. The problem of mercury contamination in Onondaga Lake fish has been sufficiently prominent to merit a paragraph of discussion in the 1975 Report of the National

3.1.2 Water Quality Issues (continued)

Council on Environmental Quality[5]. The geometric mean fecal coliform density was twice the New York State bathing standard in the epilimnion during Onondaga County's baseline monitoring period, 1968-69, reflecting primarily both dry- and wet-weather overflows from the Syracuse combined sewer system[24]. The existence of extensive sludge deposits in the vicinity of the sewage treatment plant outfall site has raised questions about the feasibility of substantially improving lake water quality by reducing pollution sources alone[24].

3.1.3 Plans for Pollution Abatement

The plan for abating the pollution of Onondaga Lake centers on the construction of a facility to provide tertiary treatment of Onondaga County's sewage. The facility, currently under construction and scheduled for completion in 1980, is expected to remove 95 percent of the influent BOD and 90 percent of the influent total phosphorus. The unique aspect of this plan is that the 0.29 m^3/sec waste bed overflow from the Allied Chemical plant will be used as a precipitating agent to effect phosphorus removal from the municipal waste stream (design flow 3.78 m^3/sec). The industrial stream is currently the major source of the lake's salinity, which is not expected to be influenced by this plan. Allied's waste beds will be used in turn for disposal of digested sludge.

3.1.3 Plans for Pollution Abatement (continued)

Reduction of the municipal sewage phosphorus loadings alone is not generally anticipated to bring about a reversal of the eutrophic conditions in the lake[38]. Although diffuse nutrient sources (combined sewer overflows and urban and agricultural runoff) have not been accurately quantified, they are considered potentially of sufficient magnitude to support abundant algal activity in the lake after reduction of municipal sewage loadings. The EPA[38] has concluded: "The best reason for instituting phosphorus removal at the MSSTP is protection of Lake Ontario". Considerably more measurements and/or analysis may be required in order to quantify nonpoint phosphorus discharges and determine the degree of control necessary to significantly abate Onondaga Lake's eutrophic symptoms.

Another aspect of the basin plan is that the existing submerged sewage outfall will be abandoned during dry weather periods and a shoreline, surface outfall constructed. Since the combined municipal/industrial effluent is expected to have a chloride concentration in excess of 7000 g/m^3, it will be much more dense than ambient lake waters (averaging 1700 g/m^3). The surface outfall has been proposed "to provide maximum mixing of the plume with ambient water and some degree of bio-degradation before sinking into the hypolimnion", and "to prevent the tendency to create a permanent stratification within the lake"[21].

3.1.3 Plans for Pollution Abatement (continued)

The surface outfall plan has been criticized with regard to possible water quality and aesthetic implications[28,38]. The saline industrial stream currently discharges into Ninemile Creek, the lake's largest tributary, prior to entering the lake. This creek affords about twice as much dilution of the saline stream as will the municipal effluent. The physics of the situation dictate that the greater the difference in density between the effluent and ambient waters, the greater will be the tendency for stratification. Implementation of the plan may result in enhanced vertical density gradients in lake and concommitant water quality impact. These considerations suggest, minimally, that the design of the outfall to permit maximum dilution of the effluent may be critical. In addition, the mixing of the calcium-rich effluent with ambient lake waters is expected to induce calcium carbonate precipitation and result in a visible white plume, which may be aesthetically objectionable[38].

With the exception of salinity, industrial discharges have reportedly been brought under control[38]. In 1974, Crucible Inc., a steel mill formerly contributing substantial quantities of oil and grease, chromium, copper, and iron, installed treatment facilities to limit these discharges. In 1970, Allied Chemical was legally forced to reduce mercury discharges from its chlor-alkalai plant from 10 kg/ day to 0.4 kg/day[38]. The problem of mercury contamination in Onondaga Lake fish has persisted, however, according to an ongoing NYDEC fish

3.1.3 Plans for Pollution Abatement (continued)

monitoring program[38,5]. No continuing program for monitoring mercury discharges or ambient lake levels is known to the author.

The EPA[37] has recognized that the solvay process in general presents special waste treatment problems. Existing plants have been granted special exemption from future zero discharge limitations. Accordingly, there is no plan to reduce the dissolved solids content of Onondaga Lake. Allied Chemical has indicated that the expense of eliminating the saline discharge would entail closing of its solvay plant and thus a substantial economic loss to Onondaga County. The plan to utilize the process waste for phosphorus removal from the municipal waste stream represents a rather unique, compromising alternative.

Another aspect of the pollution abatement program concerns the control of combined sewage discharges during dry and wet weather. Onondaga Country has undertaken an extensive sewer maintenance program and most of the dry-weather overflows into tributary creeks are reported to have been eliminated[14]. An EPA Demonstration Grant has been awarded to O'Brien and Gere to investigate high-rate processes for bacteria, ammonia, and suspended solids removal from sewer over-flows. As yet, no final solution to the problem has been adopted, although "general abatement" of the overflows is required by the year 1985.

3.1.4 Scope of Chapter

Following is a preliminary analysis of available data relevant to the water quality of Onondaga Lake and its tributaries. This work is intended to provide a basis for more detailed modelling studies aimed at policy evaluation. The levels and variations of major water quality components are summarized and displayed along temporal and spatial dimensions in order to elucidate trends and seasonal variations and to determine the extents of vertical and horizontal mixing in the lake. Considerable effort has been spent in investigating or developing methods of summarizing and displaying the data. Information obtained from the U.S. Geological survey is used to develop a hydrologic balance, which is combined with tributary, point source, and lake water quality data to formulate mass balances on major water quality components. Trends in these mass balances and in ambient lake concentrations, and shifts in phytoplankton populations observed over the study period are interpreted mechanistically. The mass balances are used in combination with some of the models discussed and developed in Chapter 2 in a preliminary assessment of the potential for controlling the eutrophication problem through point and non-point source nutrient controls or through lake mixing. The results of this analysis, coupled with considerations of the water quality issues and management policies outlined above, serve as a basis for the modelling efforts described in Chapter 4.

3.2 Data Base

3.2.1 Lake and Tributary Water Quality

In 1967, the Onondaga County Department of Public Works obtained
an FWQA Research and Development Grant (11060PAE) to determine the
feasibility of a cooperative municipal industrial approach to the
wastewater problems. As part of the contract, in 1968, Onondaga
County initiated an extensive monitoring program of Onondaga Lake and
its tributaries. The survey, performed by O'Brien & Gere Engineers,
Inc., was the first systematic and comprehensive investigation of
water quality in the area. The stated objectives of the initial
phase of the program (1968-69) were to[24]:

"1.   Ascertain the present trophic status of the lake;

2.   Evaluate the impact of engineering programs;

3.   Provide baseline data for ongoing evaluations;

4.   Establish a program for continuous monitoring of
      the lake."

The baseline study emphasized that the most important objective was
"to evaluate the impact of various pollution abatement programs with
respect to the best uses of the lake."

3.2.1 <u>Lake and Tributary Water Quality</u> (continued)

This program has been continued on a more or less uniform basis through 1976[20]. The ongoing study has been "designed to gauge the reaction of the lake to various changes in activity surrounding the lake." The monitoring report for the 1970 survey also suggests that it is "on the basis of accumulated information that a predictive model can be developed and the water quality resulting from planned pollution abatement programs can be determined." Thus, the aims of this program appear to be to monitor ambient trends and to provide basic data for development and calibration of predictive planning models. The essential features of this program are discussed below.

Table 3.2-1 presents the sampling frequencies for each station and year. The stations can be located on Figure 3.1-1. Lake station 2 was abandoned in 1971, after enough data had been obtained to indicate that the concentration differences between stations 1 and 2 were relatively small, and thus that the lake was well-mixed on horizontal planes. This assumption is examined further in Section 3.4.1. A total of ten influent tributary stations have been established, seven of which have been sampled regularly on a biweekly basis. The stations on the major tributaries (Onondaga Creek, Ninemile Creek, Ley Creek, and Harbor Brook) have been located at USGS gauge sites to facilitate flow measurements. Grab samples of the three major point-source discharges (Allied Chemical, Crucible, and Metro STP) have also been taken regularly.

Table 3.2-1

Onondaga Lake Survey

Numbers of Samples by Station and Year

	Station	Depth	1968	1969	1970	1971	1972	1973	1974	1975	Total
LAKE	Station 1	0-18 m @ 3m intervals	23	40	25	11	19	17	18	15	168
	Station 2	0-18 m @ 3m intervals	21	41	11	0	0	0	0	0	73
TRIBUTARIES	Ley Creek	surface	0	15	24	24	24	23	25	23	158
	Onondaga Creek	surface	0	17	24	24	25	24	25	23	168
	Harbor Brook	surface	0	16	22	24	25	24	25	23	159
	Ninemile Creek	surface	0	16	24	24	25	24	25	23	161
	Bloody Brook	surface	0	15	22	0	0	0	0	0	37
	Sawmill Creek	surface	0	?	18	0	0	0	0	0	18
	Sanitary Landfill[a]	surface	0	?	16	0	0	0	0	0	16
	Allied Chemical	effluent	0	15	24	24	23	24	25	23	158
	Crucible	effluent	0	19	24	24	24	24	25	23	163
	Metro STP	effluent	0	15	24	24	25	24	25	23	160
	Outlet (in)[b]	surface	0	0	0	0	21	24	23	23	91
	Outlet (out)[b]	bottom	0	0	0	0	24	24	23	23	94
	Outlet (unspecified)[b]	?	0	0	20	22	0	0	0	0	42

a - discharges into Ley Creek

b - see text

3.2.1 Lake and Tributary Water Quality (continued)

Grab samples of tributaries and point sources may not have provided a sufficient basis for accurately estimating average mass fluxes of materials into the lake. In the cases of tributaries, periodic grab sampling has not taken into account the effects of intervening storms. This is a particular problem in streams which are subject to combined sewer overflows, as are Onondaga Creek and Harbor Brook. Periodic grab sampling of point sources at these frequencies has not accounted for probable weekly and diurnal periodicities in these sources, which are coupled with domestic and industrial activities.

Sampling the outlet of Onondaga Lake is difficult due to peculiar hydrodynamic conditions. Seneca River waters have been frequently observed to flow into the lake at the surface, while the dense lake waters flow out in the lower layers. The Seneca River is part of the New York State barge canal system and its levels are regulated. The pattern of bilaminar flows at the outlet may become more distinct in dry seasons when river levels may intermittently exceed lake levels. In 1970 and 1971, the outlet was sampled regularly, apparently without distinction as to flow direction. In 1972-75, attempts were made to sample both outflowing (bottom) and inflowing (surface) waters at the outlet station. This flow pattern renders the outlet data difficult to interpret.

3.2.1 <u>Lake and Tributary Water Quality</u> (continued)

Table 3.2-2 summarizes the components monitored in the lake
and tributary surveys. A total of thirty chemical species have been
included. The study has provided meteorological data on wind speed,
direction, and air temperature over the lake. In latter years, an
increasing emphasis has been placed on bacteriological measurements.
The phytoplankton population has been characterized by identifying,
counting, and sizing the dominant species. "Biomass" represents an
aggregate estimate of the total phytoplankton cell volume per volume
of sample, as calculated from algal counts and sizes. In the latter
years, measurements of chlorophyll have been used to characterize
total algal biomass. Zooplankton have been counted and identified
as to general class (rotifers, copepods, cladocerans) throughout
the survey period and as to species in the baseline study, which
also included sampling and identification of fish and benthic organ-
isms.

Two comments are in order relative to the selection of compon-
ents. The omission of mercury from the list prior to 1975 seems
unfortunate, considering that mercury toxicity has been a predominant
water quality issue. Phosphate detergent legislation in 1971 and
future tertiary treatment are measures aimed at controlling eutro-
phication by limiting phosphorus supplies. However, the monitoring
program has not included any measure of organic phosphorus in the
sources or lake. "Total inorganic phosphorus" has been substituted

Table 3.2-2

Onondaga Lake Survey
Components Monitored

Component	Lake Stations	Tributary Stations	(Baseline) 1968-69	1970	1971	1972	1973	1974	1975
Temperature	X	X	X	X	X	X	X	X	X
Alkalinity	X	X	X	X	X	X	X	X	X
pH	X	X	X	X	X	X	X	X	X
Total Inorganic Carbon	X		X	X				X	X
Total Organic Carbon	X		X	X			X	X	X
CO$_2$	X	X	X	X	X	X	X	X	X
Chloride	X	X	X	X	X	X	X	X	X
Dissolved Oxygen	X	X	X	X	X	X	X	X	X
BOD$_5$	X	X	X	X	X	X	X	X	X
Conductivity	X	X	X	X	X	X	X	X	X
Total Inorganic P	X	X	X	X	X	X	X	X	X
Ortho P	X	X	X	X	X	X	X	X	X
Organic N	X	X	X	X	X	X	X	X	X
NH$_3$-N	X	X	X	X	X	X	X	X	X
NO$_2$-N	X	X	X	X	X	X	X	X	X
NO$_3$-N	X	X	X	X	X	X	X	X	X
SiO$_2$	X	X	X	X	X	X	X	X	X
Sulfate	X		X	X	X	X	X	X	X
Sulfide	X	X	X	X	X	X	X	X	X
Calcium	X	X	X	X	X	X	X	X	X
Magnesium	X	X	X	X	X	X	X	X	X
Sodium	X	X	X	X	X	X	X	X	X
Potassium	X	X	X	X	X	X	X	X	X
Iron	X	X	X	X	X	X	X	X	X
Copper	X	X	X	X	X	X	X	X	X
Chromium	X		X						X
Zinc	X		X						X
Mercury	X		X	X	X	X	X		
Manganese	X		X	X	X	X	X		
Fluoride	X		X	X	X	X	X		
Suspended Solids	X		X	X	X	X	X		
Dissolved Solids	X		X	X	X	X	X		
Transparency	X		X	X	X	X	X	X	X
Wind Speed	X		X	X	X	X	X		
Wind Direction	X		X	X	X	X	X		
Air Temperature	X		X		X	X	X		X
Total Coliforms	X		X	X			X	X	X
Fecal Coliforms	X		X				X	X	X
Fecal Streptococci	X		X				X	X	X
Biomass	X		X	X	X	X	X	X	X
Chlorophyll	X		X	X	X	X	X	X	X
Zooplankton	X		X	X	X	X	X	X	X
Phytoplankton	X		X	X	X	X	X	X	X

3.2.1 <u>Lake and Tributary Water Quality</u> (continued)

for the "total phosphorus" measurements usually included in lake
surveys. In the lake, the organic fraction of total phosphorus is
likely to become dominant, should phosphorus supplies become limiting
to algal growth. Hydrolysis of organic phosphorus discharged into
the lake may serve as an additional nutrient source which would not
be accounted for in a source control program based upon inorganic
phosphorus data alone.

3.2.2 <u>Metro Sewage Treatment Plant Operating Data</u>

To supplement the O'Brien and Gere grab-sample data on the Metro
STP discharge, the Onondaga County Department of Drainage and Sanita-
tion has provided copies of treatment plant operating records for
the years 1968 to 1974[23]. These reports have provided information
on minimum, mean, and maximum daily flows and on daily mean concentra-
tions of suspended solids, settleable solids, pH, and BOD_5, based
upon 24-hour flow-weighted, composite samples of the plant influent
and effluent for 1972-72, and upon 8-hour composite samples for
1968-71. Sampling was generally not done on weekends, so that mean
lake loadings derived directly from these data may be somewhat biased.

Analysis of 24-hour composite samples for total inorganic phos-
phorus was done at approximately weekly intervals from 1972 to 1974.
Comparison of these data with the O'Brien and Gere morning grab-sample

3.2.2 <u>Metro Sewage Treatment Plant Operating Data</u> (continued)

data on the plant effluent over the same period has provided a means
of calibrating the latter measurements and correcting them for time-
of-day. Both surveys have used Onondaga County's laboratories for
analysis and have sampled on weekdays, so that differences in the
phosphorus data can likely be attributed to diurnal variations. The
data from each survey have been aggregated and averaged at one-month
intervals and compared on a month-by-month basis. The mean ratio of
the composite to morning grab-sample concentration was found to be
1.43 (standard deviation 0.28). This is in general agreement with
the analyses of Mauleg et al.[17] and Shannon[29], both of whom have
observed marked diurnal periodicities in total phosphorus concentra-
tions of sewage, with concentrations tending to be lowest in mid-morning
hours.

The Metro STP operating records have provided information of the
incidence of plant upsets due to operating or equipment problems,
potentially of use in interpreting residuals calculated in future
modelling studies. The records also note the incidence of major
storms requiring bypass of sewage around the plant directly into the
lake. No information on the quantity or quality of the bypassed flow
is available, however.

3.2.3 Hydrology

The US Geological Survey has provided tributary flow data for the development of a hydrologic balance. Table 3.2-3 summarizes the availability of streamflow data for a total of six gauging stations. Only two stations (Onondaga Creek at Dorwin Ave. and Ninemile Creek at Camillus) were operated for the entire period of interest (1968-74). In order to estimate missing flow data, relationships have been developed relating flows at stations with missing data to the flows at either of the stations with a complete record. The regression models employed and parameter estimates are given in Table 3.2-4.

No evaporation data for the region and period have been located. To estimate evaporation from the lake surface, regional monthly-average values given by Meyer[18] have been used. While year-to-year climatological variations would be expected to influence the evaporation rates, the overall hydrologic balance of the lake is not sensitive to direct precipitation or evaporation, since the lake's mean surface overflow rate is about 52 meters/year, as compared with mean precipitation and evaporation rates of 1.13 meters/year and 0.67 meters/year, respectively.

The above information has been used to formulate a seven-year hydrologic balance on Onondaga Lake on a monthly basis (Table 3.2-5, Figure 3.2-1). Since no continuous gauging at the outlet has been done, the total outflow for each month has been estimated by summing

Table 3.2-3

Availability of Flow Data for Onondaga Lake Hydrologic Balance

| Station | USGS Code | Location | Drainage Area (km^2) | Water Year | | | | | | |
				1968	1969	1970	1971	1972	1973	1974
A	04239000	Onondaga Creek @ Dorwin Ave.	229.2	X	X	X	X	X	X	X
B	04240010	Onondaga Creek @ Spencer St.	282.3				X	X	X	X
C	04240105	Harbor Brook @ Hiawatha	29.3				X	X	X	X
D	04240120	Ley Creek @ Syracuse	77.4						X	X
E	04240200	Ninemile Creek @ Camillus	218.3	X	X	X	X	X	X	X
F	04240300	Ninemile Creek @ Lakeland	297.8				X	X	X	

X = flow data available

Table 3.2-4

Regression Models Used to Estimate Missing Flow Observations

General Model :

$$\log_{10} Q_i = a + b \log_{10} Q_j + e$$

Q_i, Q_j = mean monthly flow for stations i and j (m^3/sec)

a, b = regression parameters

e = error term

Station i	Station j	Number of Observations	a	b	r^2	SEE
Onondaga Creek @Spencer Street	Onondaga Creek @Dorwin Avenue	49	.253	.834	.986	.037
Ninemile Creek @Lakeland	Ninemile Creek @Camilus	34	.422	.773	.974	.046
Ley Creek @Syracuse	Onondaga Creek @Dorwin Avenue	21	-.174	.555	.773	.108
Harbor Brook @Hiawatha Blvd.	Ninemile Creek @Camilus + .91 m^3/sec[a]	48	-.863	.827	.832	.098

a - .91 m^3/sec added to Camilus flow to account for upstream drinking water diversion from Otisco Lake and to linearize relationship between Harbor Brook and Ninemile Creek flows.

Table 3.2-5

Onondaga Lake Water Balance

Mean Flows in Cubic Meters Per Second

| | Drainage Area (km²) | Water Year | | | | | | | |
		1968	1969	1970	1971	1972	1973	1974	Mean
Onondaga Creek	282.3	5.42	6.67	4.82	5.72	6.56	6.51	5.86	5.94
Harbor Brook	29.3	.47	.55	.43	.45	.55	.60	.56	.52
Ley Creek	77.4	1.37	1.56	1.26	1.37	1.51	1.73	1.44	1.46
Ninemile Creek	297.8	6.96	8.35	6.18	8.06	7.62	8.79	6.95	7.57
Ungauged Inflow	39.6	.82	.99	.73	.90	.94	1.02	.85	.89
Metro STP	-	2.10	2.19	2.60	2.68	2.95	2.99	3.49	2.72
Precipitation	11.7	.38	.35	.35	.40	.46	.50	.52	.42
Evaporation	-	.25	.25	.25	.25	.25	.25	.25	.25
Outflow	738.1	17.28	20.42	16.14	19.33	20.34	21.88	19.42	19.26

Figure 3.2-1

Monthly Water Balance
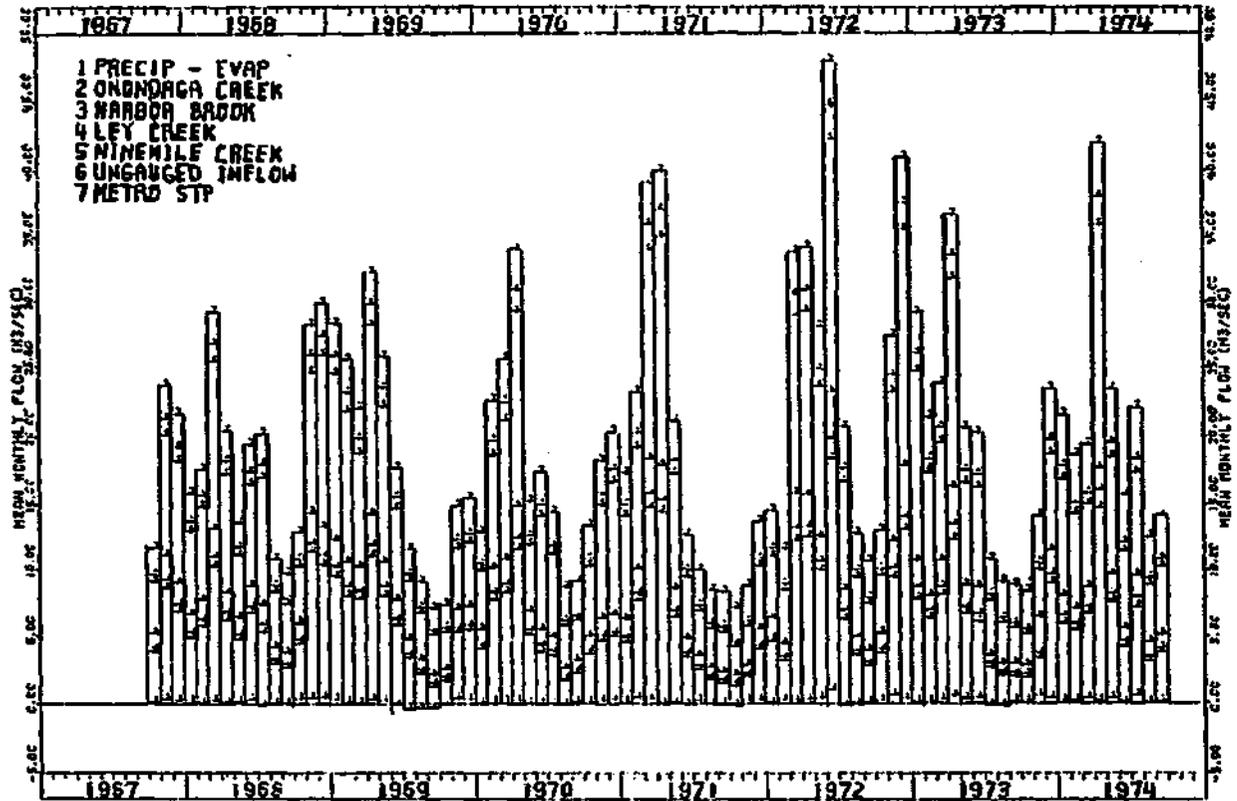
3.2.3 Hydrology (continued)

the inputs and subtracting evaporation. The ungauged portion of the
inflow has been estimated by multiplying the total gauged tributary
inflow by the ratio of ungauged to total gauged drainage areas (0.058).

Annual outflow ranged from 16.1 cubic meters per second in 1970
to 21.9 cubic meters per second in water year 1972, corresponding to
a range of 100 to 74 days in mean hydraulic residence time. The Metro
STP discharge averaged 14% of the mean outflow rate. The two industrial
discharges, Crucible and Allied Chemical, representing exchange flows
of 0.29 and 3.62 m^3/sec, respectively, did not influence the net water
balance of the lake.

3.2.4 Meteorology

Meteorologic data has been obtained from the Local Climatological
Data publication of the US Department of Commerce[36] for the weather
station at Hancock Airport, Syracuse, located about four miles north-
east of the lake. Table 3.2-6 summarizes relevant data on precipitation,
sunshine, cloud cover, and wind speed and direction over the period
1968-74.

Table 3.2-6

Summary of Meteorologic Data 1968-74

Hancock Airport, Syracuse, New York

Calendar Year	Precipitation (cm)	Percent of Possible Sunshine	Average Wind Speed (km/hr)	Resultant Wind Speed (km/hr)	Resultant Wind Direction[a]	Maximum Wind Speed (km/hr)
1968	112.3	45.3	15.1	5.1	26.0	69.4
1969	81.4	46.8	14.2	4.2	26.0	75.8
1970	97.1	48.8	15.0	3.7	25.2	69.4
1971	99.5	48.5	15.4	4.7	25.4	75.8
1972	140.7	42.3	14.7	3.7	25.5	83.9
1973	133.7	44.1	15.0	4.2	25.3	72.6
1974	127.6	42.4	15.5	4.8	25.8	96.8

a - 0 = 36 = True North

3.3 Methods

The techniques discussed below have been applied to summarize and display the water quality data along spatial and temporal dimensions. This work is essentially exploratory in nature. It is intended to elucidate essential temporal and spatial variations and associations within data. Results are contained in Appendices A through D.

3.3.1 t-Tests for Horizontal Mixing

Paired t-tests (Appendix A) have been made in order to test the statistical significance of the differences in the mean concentrations of various components between stations 1 and 2 (see Figure 3.1-1). Tests have been applied to data from each of the seven depths sampled and to epilimnion (0., 3 meters, 6 meters) and hypolimnion (12 meters, 15 meters, 18 meters) averages. The objective of this work is to determine the degree of horizontal mixing in the lake at each vertical level.

For each vertical level and component, all pairs of concentration measurements taken simultaneously have been selected from the general data matrix. For each station, the mean, standard deviation, and standard error of the mean have been calculated. A paired t-statistic has been computed according to:

3.3.1 t-Tests for Horizontal Mixing (continued)

$$t_k = \frac{\bar{c}_{1k} - \bar{c}_{2k}}{\sqrt{\sum_{i=1}^{N_k} (c_{1ki} - c_{2ki})^2 / N_k (N_k - 1)}} \qquad (3.3-1)$$

where,

c_{jki} = concentration of station j , depth k , and date i

N_k = number of paired observations at depth k

$\bar{c}_{jk}$ = mean concentration at station j and depth k

t_k = t-statistic at depth k

The t-statistic, with $N_k - 1$ degrees of freedom, has been used to test the significance of the differences in means between the two stations. Since station 2 was abandoned after 1970, the data for these tests spans the 1968-70 period, a total of 88 sampling dates.

3.3.2 GRID Displays

GRID[30], a program from the Harvard Laboratory for Computer Graphics and Spatial Analysis, has been used to display water quality

3.3.2 GRID Displays (continued)

variations as a function of time and depth (Appendix B). Variations
in concentration have been represented at six levels using symbols of
increasing visual intensities to indicate increasing concentrations.
For each water quality component, the original matrix of concentration
observations ($\sim$ 168 dates $\times$ 7 depths) has been transformed into a
96 $\times$ 19 matrix by aggregating and averaging the data at monthly
intervals and interpolating linearly with depth to provide estimates
at one meter intervals from 0 to 18 meters. Accordingly, a one-month
by one-meter GRID cell size has been used. Months with missing
profiles have been estimated by interpolating between adjacent
months. Profiles missing for more than three successive months have
been left blank in the displays. The six concentration levels have
been selected at equal intervals for components which appeared to be
distributed normally and at geometrically increasing intervals for
components which tended toward lognormal distributions. This type
of display has been found to be particularly helpful in identifying
seasonal stratification patterns.

3.3.3 Line Plots

Seasonal and long-term changes in concentration have been
illustrated by plotting raw and smoothed volume-averaged concentra-
tions as a function of time (Appendix C). For each sample profile,

3.3.3 Line Plots (continued)

the volume-averaged concentration has been computed as:

$$\bar{c} = \frac{\int_0^{z_{max}} c_z\, a_z\, dz}{\int_0^{z_{max}} a_z\, dz} \tag{3.3-2}$$

where,

a_z = lake surface area of depth z

c_z = observed concentration at depth z

The above integration has been performed using the trapezoidal rule, with surface areas derived from Figure 3.3-1.

In order to estimate seasonal and long-term components of the variations in each time series, a variety of "smoothing" techniques have been tried. One of the most flexible and generally satisfactory methods found was that of "cubic splines", due to Reinsch[27,41]. This technique consists of fitting piecewise cubic polynomials to each set of four successive observations in the time series. The coefficients of the polynomials are selected to maximize the smoothness of the total curve subject to a specified constraint on the mean squared deviation of the fitted curve from the observations. The measure of smoothness employed is minus the integral of the squared second

Figure 3.3-1

ONONDAGA LAKE

Surface Area and Volume Vs. Depth
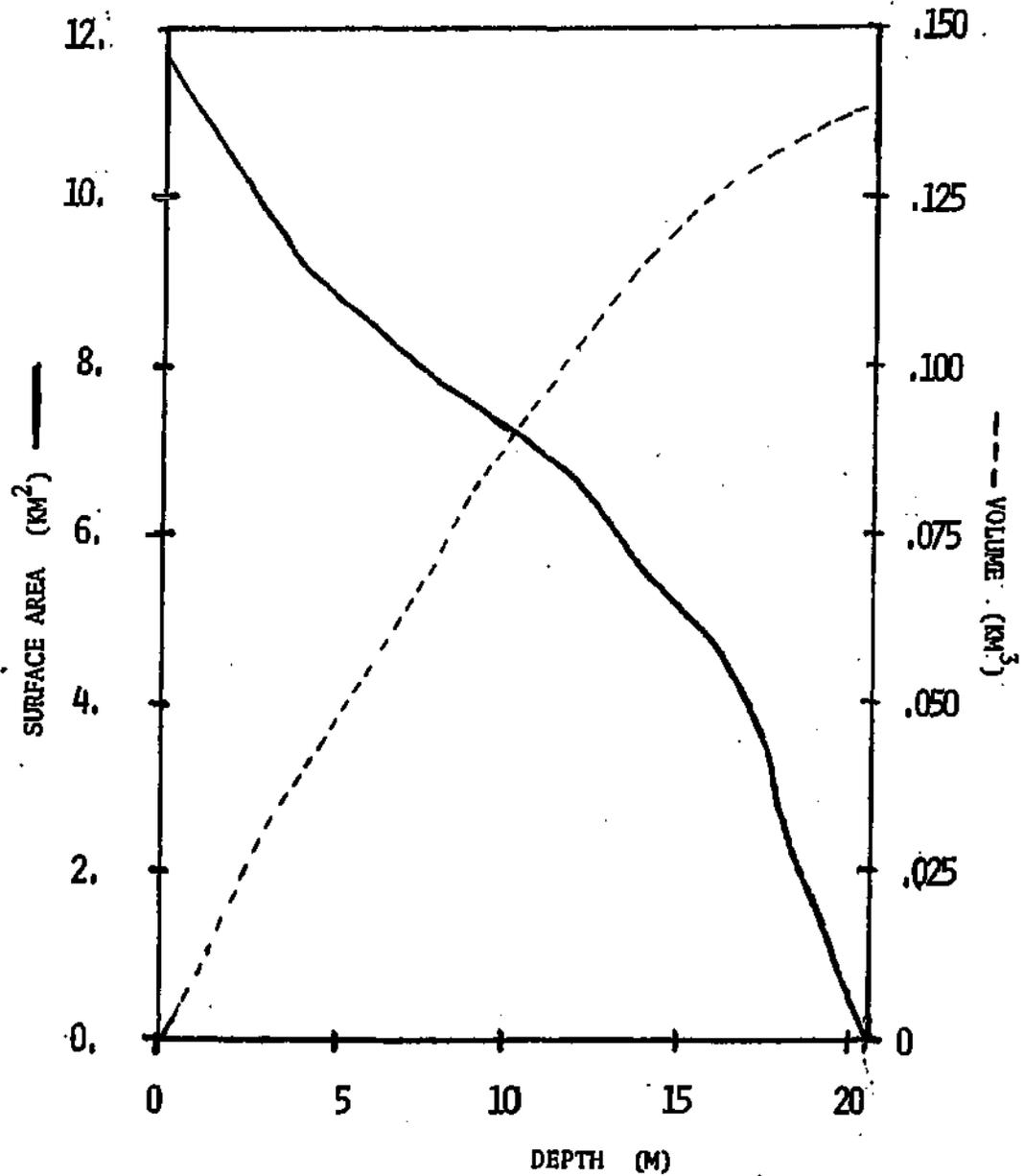
3.3.3 Line Plots (continued)

derivative. Mathematically, the problem is formulated as:

$$\text{maximize:} \quad F = - \int_{tmin}^{tmax} [S''(t)]^2 \, dt \qquad (3.3\text{-}3)$$

$$\text{subject to:} \quad \sum_{i=1}^{N} \left(\frac{S(t_i) - y_i}{\delta y_i} \right)^2 \leq N \qquad (3.3\text{-}4)$$

where,

$$S(t) = a_i + b_i(t-t_i) + c_i(t-t_i)^2 + d_i(t-t_i)^3 , \qquad t_i \leq t \leq t_{i+1}$$

a_i, b_i, c_i, d_i = coefficients for the time interval, $t_i \leq t \leq t_{i+1}$

$S(t)$ = "smoothed" value at time t

y_i = observed value of time t_i

δy_i = estimate of the standard deviation of observation y_i

N = total number of observations

Coefficients at the beginning and ends of the time series are subject to the following constraints:

3.3.3 Line Plots (continued)

$$c_1 = c_N = d_N = 0 \qquad\qquad (3.3-5)$$

The computations of the coefficients have been performed using a subroutine in the IBM SLMATH Program Library[12]. This method selects the smoothest possible curve that fits the data to within the specified accuracy. By varying the specified levels of the δy_i , varying degrees of smoothness can be achieved. After some experimentation, most satisfactory results have been reached by first transforming the data to render variance independent of level, which, in many cases required log transformation. In such cases, a small number has been added to each observation before taking logarithms to allow inclusion of zeroes. The standard deviation of the transformed data has been computed and each element in the δy_i vector set equal to a specified fraction, f , of the computed standard deviation. In most cases f level of 1.0 has been found to give smooth curves which seem to follow trend, while an f level of 0.5 has been found to give curves which follow seasonal variation, as well as trend. In a series with a large trend, such as phosphorus, there was little difference in the smoothed curves at f levels of 1.0 and 0.5, and f levels of 0.5 and 0.25 have been used to depict trend and seasonal variations, respectively. Plotting the raw data as points and each of the two smoothed curves on the same graph has

3.3.3 Line Plots (continued)

been found to give a reasonable display of the various components of the time series variation. This technique is similar to that employed by Spirtas and Levin[31] in examining air pollution data.

The smoothing techniques described above are a form of exploratory data analysis and have been employed only as an objective means of summarizing the data and elucidating essential aspects. Since second moments have not been considered, no notion of the statistical significance of trend or seasonal components has been implied. The method has been employed as a partial means of suggesting time series model formulations, for which parameters might be estimated and tested for significance. The cubic spline method selected for smoothing compares favorably with other techniques, such as moving averages with various weighting schemes. The latter methods generally require equally spaced observations and give smooth curves whose variation tends to lag behind variations in the data, unless optimal filtering coefficients are used.

3.3.4 Mass Balances

Lake and tributary quality data have been combined with hydrologic data in order to formulate mass balances on major water quality components for water years 1970 through 1974 (Appendix D). As discussed in Section 3.2, only approximate estimates of mass fluxes can be derived,

3.3.4 Mass Balances (continued)

because only grab-sample data was available for the tributaries and point sources. Estimation of confidence limits for the mass fluxes according to the methods outlined in Section 2.3.2, would be essential to an engineering evaluation of potential impacts of source control measures, but has been excluded from this preliminary analysis.

For each tributary or point source, continuous stream flow and grab-sample concentration data have been used to estimate average mass fluxes at one-month intervals. The algorithm employed to perform these calculations is based upon the following equality:

$$\overline{W} (t_2 - t_1) = \int_{t_1}^{t_2} w_t \, dt = \int_{t_1}^{t_2} q_t \, c_t \, dt = \int_{v_1}^{v_2} c_t \, dv \qquad (3.3\text{-}6)$$

$$v_2 - v_1 = \int_{t_1}^{t_2} q_t \, dt \qquad (3.3\text{-}7)$$

where,

$\overline{W}$ = mean mass flux between time t_1 and time t_2 (mass/time)

w_t = instantaneous flux at time t (mass/time)

q_t = instantaneous flow at time t (vol/time)

3.3.4 Mass Balances (continued)

c_t = instantaneous concentration at time t (mass/vol)

v = integrated flow (vol)

According to the scheme employed, for a given set of concentration and flow data, the coefficients of variation of sample concentration and sample flux are computed and compared. If concentration is found to be more variable than flux, calculations are performed on the flux versus time scale, otherwise they are performed on the concentration versus volume scale. In either case, the series is first smoothed by fitting piecewise linear least-squares segments to each set of three successive observations. This smoothing operation is performed three times. The series is then integrated numerically using the trapezoidal rule and the integral is evaluated at monthly intervals by linear interpolation. Generally, concentration has been found to be less variable than flux in the tributary and lake data and more variable than flux in the point source data.

The fundamental equation employed in the mass balance is:

$$\text{Input} - \text{Output} = \text{Change in Storage} + \text{Accumulation} \qquad (3.3-8)$$

3.3.4 Mass Balances (continued)

For each component and month, an estimate of the total input flux
has been derived by summing the individual tributary and point source
contributions. Ungauged inputs have been estimated based upon gauged
tributary inputs and drainage area. Due to the difficulty in sampling
the lake outlet (see Section 3.2.1), concentration data from the epilim-
nion (average of 0, 3, and 6 meter samples at station 1) have been
used to estimats outflow. The change in storage term in equation (8)
has been determined from the volume-averaged concentration data at
station 1. Finally, the accumulation term has been calculated by
difference.

The mass balances have been plotted in cumulative form in Appen-
dix D. For each component, the cumulative input plot depicts the
integrals of the mass flux from each source. The cumulative balance
plot depicts the integrals of the total input, output, change in
storage, and accumulation terms. Changes in the slopes of these
curves reflect flux rate changes.

3.4 Results and Discussion

3.4.1 Horizontal Mixing

Table 3.4-1 summarizes the results of the tests for significant differences in means between stations 1 and 2 in the epilimnion and hypolimnion. The calculations are tabulated in detail in Appendix A. In Table 3.4-1, components have been grouped according to whether the tests indicated that the mean for station 1 was higher, that there was no significant difference between the stations, or that the mean for station 2 was higher. In addition, the differences in means have been computed as percents of the grand means.

Figure 3.1-1 shows that station 1 is located in the south basin of the lake, closer to Syracuse and the Metro STP discharge. Results of these tests indicate that the epilimnion at station 1 is generally a more reduced environment, significantly higher in BOD_5, ammonia, nitrogen, and organic nitrogen, and lower in dissolved oxygen and NO_3-N than the epilimnion at station 2. Higher rates of primary productivity at station 2 are indicated by higher levels of biomass, dissolved oxygen, and pH and lower levels of alkalinity. The slightly higher epilimion mean temperature at station 1 may reflect the industrial cooling water discharges at the south end of the lake.

Differences between stations in the hypolimnion are less easily interpreted. Alkalinity and pH show the same pattern as in the epilimnion. Higher levels of BOD_5 at station 2 may reflect sedimentation

Table 3.4-1

Summary of Results of
t-Tests for Horizontal Mixing

	STATION 1[e] SIGNIFICANTLY[a] HIGHER COMPONENT Δ[b]	NO SIGNIFICANT[a] DIFFERENCE BETWEEN STATIONS COMPONENT Δ[b]	STATION 2[e] SIGNIFICANTLY[a] HIGHER COMPONENT Δ[b]
EPILIMNION[c]	BOD$_5$ (28.58%) Temperature (.73%) Alkalinity (2.63%) Ammonia N (9.10%) Organic N (11.83%)	NO$_2$-N (.78%) Calcium (-.60%) SiO$_2$ (3.29%) Chloride (-.14%) Ortho-P (-.10%) TIP (-7.95%) Iron (2.70%) Chromium (-.62%) Copper (4.16%)	Dissolved Oxygen (-19.52%) Biomass (-2.69%) NO$_3$-N (-8.50%) pH (-1.27%)
HYPOLIMNION[d]	Alkalinity (1.92%) SiO$_2$ (6.85%) Chloride (2.10%)	Dissolved Oxygen (4.24%) Biomass (8.04%) Temperature (-2.51%) NO$_2$-N (20.9 %) NO$_3$-N (-20.91%) Calcium (.20%) Ammonia-N (4.27%) Organic-N (6.36%) Ortho-P (1.84%) TIP (9.93%) Iron (-5.52%) Chromium (-5.37%) Copper (7.05%)	BOD$_5$ (-11.64%) pH (-.53%)

a-95% Confidence level

b-$\Delta = \dfrac{2(x_1 - x_2)}{x_1 + x_2} \times 100\%$

c-epilimnion = mean of depths 0m, 3m, 6m;

d-hypolimnion = mean of depths 12m, 15m, 18m.

e-for station locations, see Figure 3.1-1

3.4.1 Horizontal Mixing (continued)

of sewage or algal solids. Growth and sedimentation of diatoms may account for slightly higher SiO_2 concentrations at both levels of station 1, although only hypolimnion differences are statistically significant. The slightly higher chloride levels in the hypolimnion of station 1 are unexplained.

Because prevailing winds are from the west and the lake's orientation is northwest to southeast (see Figure 3.1-1), one would expect considerable backmixing in the system in the direction from station 2 to station 1, the opposite direction from net advective outflow. These results indicate that there is some "plug flow" behavior of the system, or that, on the average, backmixing rates do not totally dominate over advection and reaction rates. The "reactions" involved include oxidation or settling of BOD, nitrification, and algal nutrient uptake and growth. Despite the fact that statistically significant differences can be discerned, the percentage differences between the stations are generally small, especially in comparison to vertical differences. It is also possible that some of the differences between the stations can be accounted for by exchange flows with Seneca River waters at the outlet.

3.4.2 Vertical Mixing

GRID displays in Appendix B illustrate marked vertical strati-
fication patterns in the lake during summer months. Table 3.4-2
indicates the components tending to concentrate in the epilimnion,
those exhibiting no marked vertical stratification, and those tending
to concentrate in the hypolimnion. The pattern generally reflects
the effects of atmospheric exchange and primary production in the
epilimnion, settling of algal and non-algal particulates, and
subsequent decomposition and release soluble compounds in the reduced
conditions of the hypolimnion.

The tendencies of BOD_5, SiO_2, ortho-P, TIP, and NH_3-N to con-
centrate in the hypolimnion are consistent with the above mechanisms.
Higher levels of NO_2-N and NO_3-N in the epilimnion reflect higher
dissolved oxygen levels there and possible denitrification reactions
in the hypolimnion. The pH and alkalinity patterns are consistent
with the following reactions[33]:

$$
\begin{matrix}
\text{(alkalinity)} & \text{(calcite)} \\
Ca^{++} + 2\ HCO_3^- \ \rightleftharpoons \ CaCO_3 \downarrow + CO_2 + H_2O
\end{matrix}
\qquad (3.4\text{-}1)
$$

$$
CO_2 + H_2O \ \rightleftharpoons \ H_2CO_3 \ \rightleftharpoons \ H^+ + HCO_3^-
\qquad (3.4\text{-}2)
$$
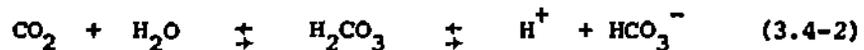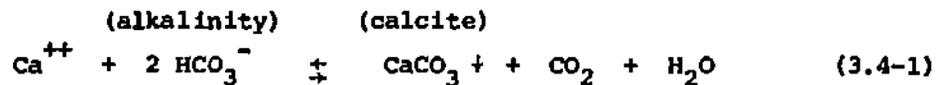
Table 3.4-2

Summary of Mid-Summer Vertical Stratification
Patterns in Onondaga Lake, by Component

Components Concentrating in Epilimnion	Components Showing No Stratification	Components Concentrating in Hypolimnion
Dissolved Oxygen	Chromium	BOD_5
Organic N	Copper	SiO_2
NO_2-N	Iron	TIP
NO_3-N		Ortho-P
pH		NH_3-N
Temperature		Alkalinity
Chlorophyll		Calcium
		Chloride

3.4.2 Vertical Mixing (continued)

Carbon dioxide is removed from the epilimnion as a result of photo-synthesis and added to the hypolimnion as a result of respiration and organic matter decay. Removal of CO_2 from the epilimnion would tend to drive reaction (1) to the right, resulting in increased removal of alkalinity through calcium carbonate precipitation. CO_2 removal would also drive reaction (2) to the left, leading to a decrease in hydrogen ion concentration and an increase in pH. The temperature and chloride stratification patterns are chiefly density effects. The metals chromium, copper, and iron do not seem to stratify, possibly indicating that they are not involved heavily with the chemical and biological processes described above.

The lake does not remain vertically stratified year-round. Density gradients caused by chloride stratification are apparently not sufficiently strong to prevent fall and spring overturns induced by thermal variations. The density structure of Onondaga is further examined in Figures 3.4-1 to 3.4-5.

A GRID display of density is shown in Figure 3.4-1. For each cell in the grid, density has been computed from the corresponding temperature and chloride value using the following relationship[13,9]:

$$\rho = (1 - k_1(T-4)^2)(1 + k_2C) \qquad (3.4-3)$$

where,

$$\rho = density \ (g/cm^3)$$

Figure 3.4-1
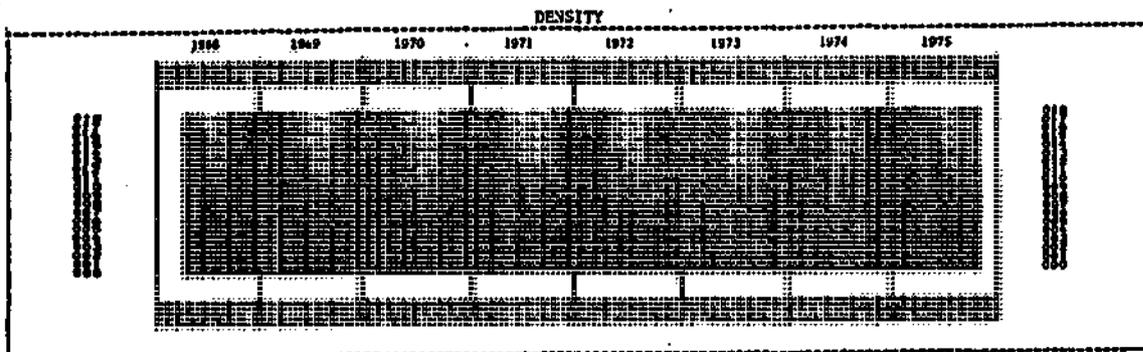
DENSITY

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : DENSITY , (P=1.0)*1000. = P IN G/CM3

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF -1.00 AND 2.40 MEAN = 1.40 ST. DEV. = 0.90

Figure 3.4-2

DENSITY GRADIENT

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : DENSITY GRADIENT ((MG/CM3)/M)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.0 AND 0.24 MEAN = 0.09 ST. DEV. = 0.06

3.4.2 Vertical Mixing (continued)

T = temperature (degrees C)

C = chloride concentration (g/m^3)

k_1 = 6.57×10^{-6} [(g/cm^3)/(deg C)^2]

k_2 = 1.2×10^{-6} (g/m^3)$^{-1}$

A GRID display of vertical density gradient is shown in Figure 3.4-2. For each month (column), a fourth degree polynomial (ρ versus z) has been fit to the seven density values computed from the corresponding seven temperature and chloride observations. The polynomial has been differentiated and evaluated to provide estimates of the vertical density gradient at one-meter intervals. Because of the relative numerical instability of the derivatives of fitted polynomials, estimates of density gradient obtained in this way are only approximate, but sufficient for display purposes. The objective of these calculations is to illustrate the location, strength, and seasonal variation of the pycnocline, the center of which corresponds to the depth of the maximum density gradient. The magnitude of the maximum density gradient may be taken as an indication of the degree of resistance to vertical mixing.

Figure 3.4-2 illustrates the downward movement of the pycnocline from near the surface in late spring to around 12 meters in late

3.4.2 Vertical Mixing (continued)

summer, prior to overturn. Such migration is typically attributed
to evaporative cooling of surface waters and input of kinetic energy
at the surface due to wind action[3,11,32]. Density gradients appear
to be decreasing in strength somewhat during this period, particularly
in comparing the early spring months of successive years. In 1974,
the pycnocline was considerably weakened relative to other years.

Figure 3.4-3 displays raw and smoothed variations of the vertical
temperature and chloride gradients. For each sampled profile and
component, the volume-weighted gradient has been computed as:

$$\overline{\frac{dx}{dz}} = \frac{\int_0^{z_m} (x_z - \overline{x})(z - \overline{z}) a_z \, dz}{\int_0^{z_m} (z - \overline{z})^2 a_z \, dz} \tag{3.4-4}$$

$$\overline{z} = \frac{\int_0^{z_m} z \, a_z \, dz}{\int_0^{z_m} a_z \, dz} \tag{3.4-5}$$

$$\overline{x} = \frac{\int_0^{z_m} x_z \, a_z \, dz}{\int_0^{z_m} a_z \, dz} \tag{3.4-6}$$

3.4.2 Vertical Mixing (continued)

where,

x_z = concentration at depth z (g/m^3)

z = surface area of lake at depth z (m)

z_m = maximum depth (m)

The trapezoidal rule has been used to perform the above integration numerically. The computation of the gradient according to equation (4) is analogous to the computation of a linear regression coefficient of x_z on z, weighted according to surface area. Figure 3.4-3 indicates that both temperature and chloride gradients were less pronounced in the later years of the survey. The absolute values of the thermal gradients generally appear to peak slightly after mid-year, corresponding to peak lake temperatures (Appendix C). The absolute values of chloride gradients generally appear to peak in early spring.

The relative contributions of thermal and chemical gradients to the total density gradient are plotted in Figure 3.4-4. The volume-average density gradient has been approximated as:

Figure 3.4-3

Line Plots of Temperature and Chloride Gradients

Figure 3.4-4

Chemical and Thermal Components of Total Density Gradient
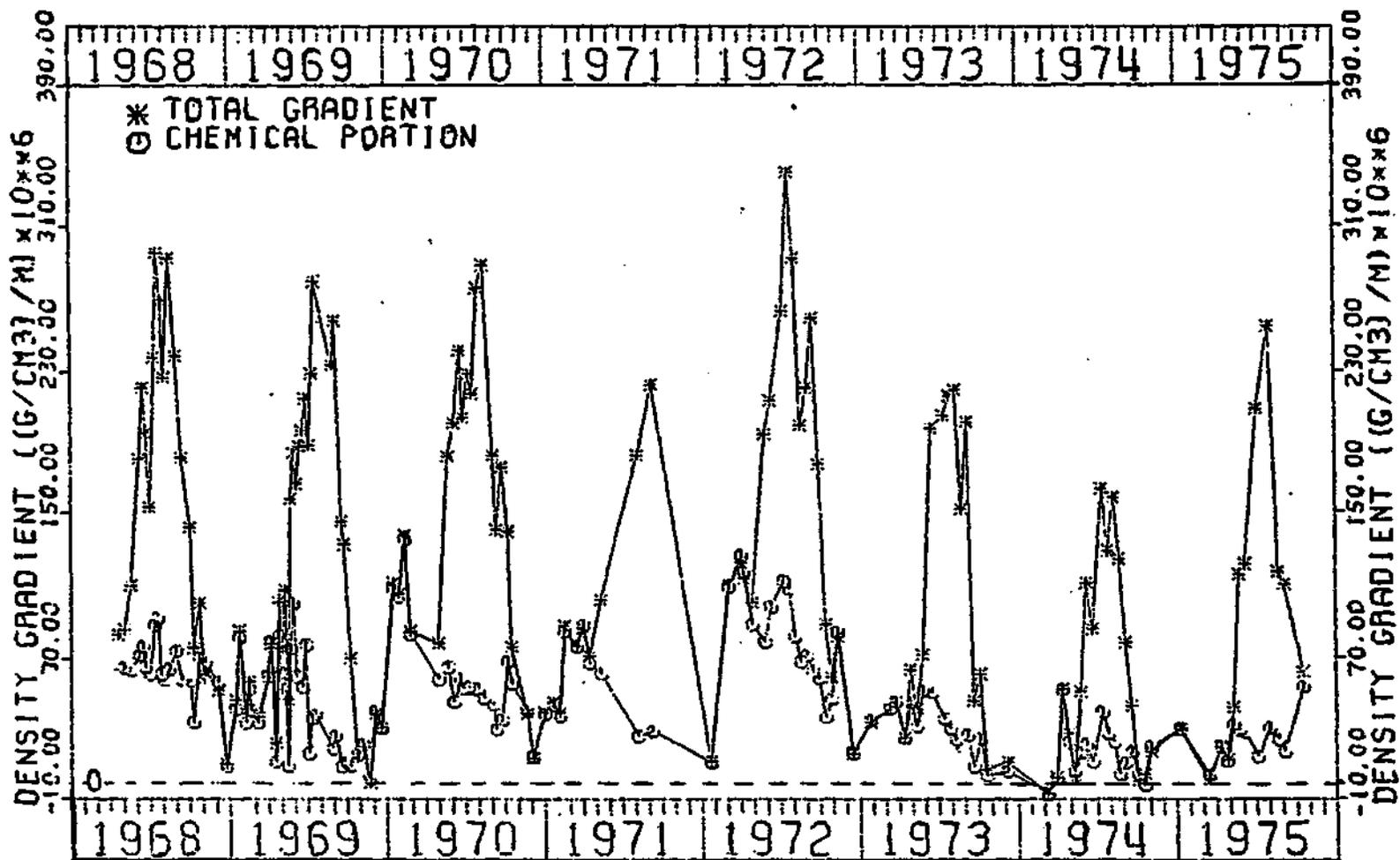
3.4.2 Vertical Mixing (continued)

$$\overline{\left(\frac{\partial \rho}{\partial z}\right)} \approx \overline{\left(\frac{\partial T}{\partial z}\right)} \overline{\left(\frac{\partial \rho}{\partial T}\right)} + \overline{\left(\frac{\partial c}{\partial z}\right)} \overline{\left(\frac{\partial \rho}{\partial c}\right)} \qquad (3.4\text{-}7)$$

total   thermal   chemical
density   portion   portion
gradient

From equation (3):

$$\overline{\left(\frac{\partial \rho}{\partial T}\right)} = -2 k_1 (\overline{T} - 4)(1 + k_2 \overline{\overline{C}}) \qquad (3.4\text{-}8)$$

$$\overline{\left(\frac{\partial \rho}{\partial c}\right)} = [1 - k_1 (\overline{T} - 4)^2] k_2 \qquad (3.4\text{-}9)$$

The means and gradients of chloride and temperature have been calculated according to equations (4), (5), and (6). The two lines in Figure 3.4-4 represent variations of the total density gradient and of the chemical portion of the total density gradient. The distance between the two lines represents the thermal portion. It is evident that thermal effects dominate during mid-summer and chemical effects are most important during other periods. The figure also indicates that the decreasing trend in early spring density gradients noted previously is associated primarily with

3.4.2 <u>Vertical Mixing</u> (continued)

decreasing chloride stratification.

An essentially equivalent representation of the average density gradient is shown in Figure 3.4-5. The "buoyant potential energy deficit" represents the theoretical energy requirement to completely mix the lake vertically[32]. It has been computed from each sample profile according to:

$$PE = 9.8 \times 10^9 \left[\int_0^{z_m} \rho_z \, z \, a_z \, dz - \overline{\rho z} \int_0^{z_m} a_z \, dz \right] \qquad (3.4\text{-}10)$$

$$\overline{\rho} = \frac{\int_0^{z_m} \rho_z \, a_z \, dz}{\int_0^{z_m} a_z \, dz} \qquad (3.4\text{-}11)$$

$$\overline{z} = \frac{\int_0^{z_m} z \, a_z \, dz}{\int_0^{z_m} a_z \, dz} \qquad (3.4\text{-}12)$$

where,

PE = buoyant potential energy deficit (joules)

ρ_z = density at depth z (g/cm^3)

Figure 3.4-5

Buoyant Potential Energy Deficit
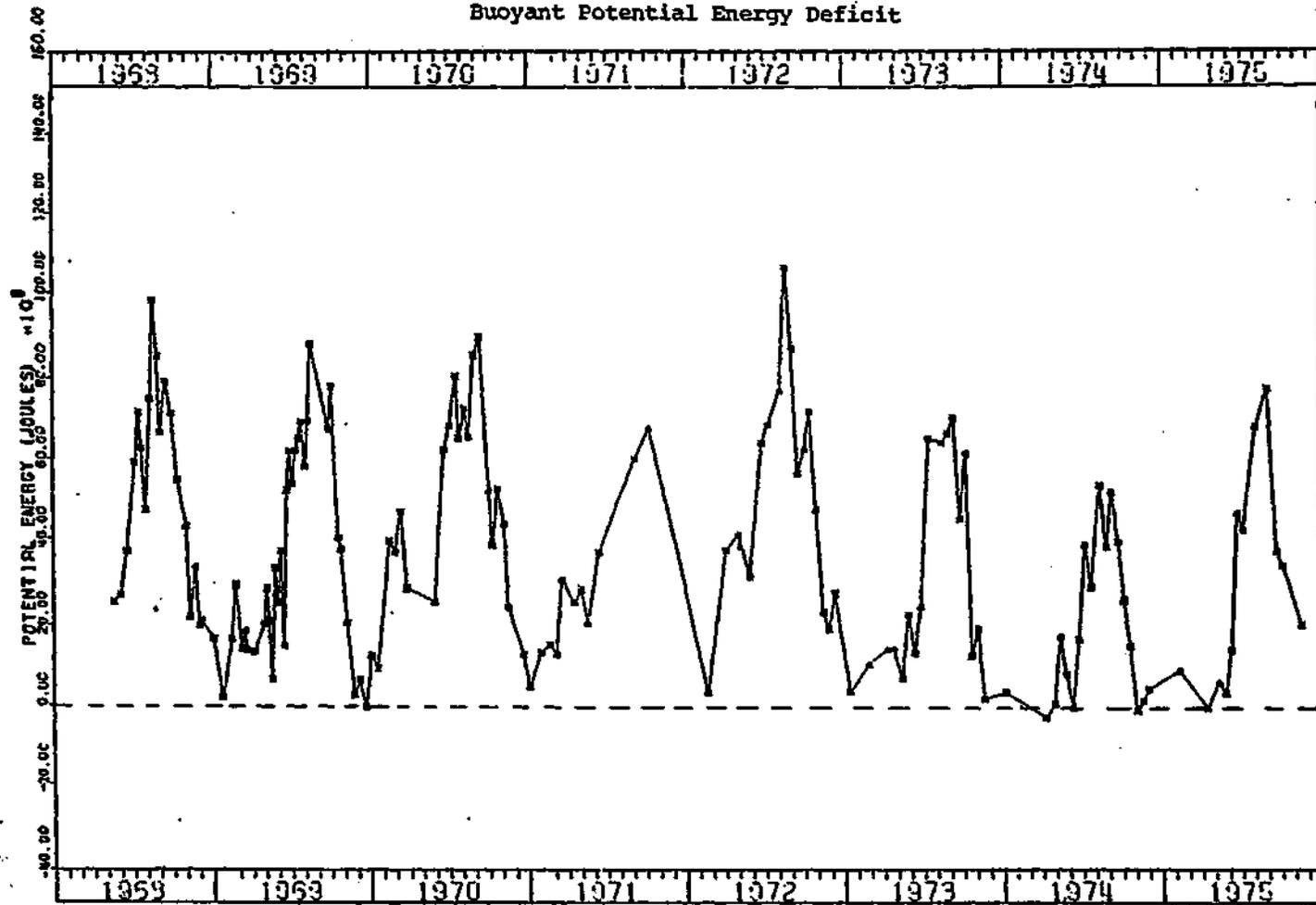
3.4.2 Vertical Mixing (continued)

Again, the trapezoidal rule has been used to perform the above numerical integration. A PE value of zero or less indicates instability in the water column, i.e. well-mixed conditions. A decreasing trend in PE is apparent, with yearly maximum values decreasing and longer periods of instability occurring between maxima.

The above analysis suggests that Onondaga Lake underwent a change in density stratification during the study period. Calculations in Appendix D do not indicate any change in the mean rates of chloride flux into the lake over this period. It seems most likely that the observed changes can be attributed to hydrologic and/or climatologic variations. While mean annual flows did not vary much from year to year (Table 3.2-5), maximum mean monthly flows were generally higher in later years (Figure 3.2-1). Possible climatologic variations affecting thermal or kinetic energy exchanges at the lake surface may also have influenced the degree of density stratification. Modelling studies will be required to determine whether the decreasing trend in density stratification can be attributed to such factors.

3.4.3 Phytoplankton and Nutrients

The vertical and seasonal distributions of nutrients in Onondaga Lake have been discussed previously (Section 3.4.2). The pattern has been shown to be consistent with algal growth in surface regions and subsequent settling and release of available nutrients into the hypolimnion. During the study period, it is unlikely that nutrients kinetically limited algal growth rates, although they may have played a role in determining dominant species. Figure 3.4-6 depicts variations of the epilimnion- and hypolimnion-average chlorophyll concentrations between 1972 and 1975. Annual cycles appear to be bimodal and no trend in peak chlorophyll concentrations can be identified.

SiO_2 depletion in the epilimnion, particularly in 1974 and 1975 (Appendix B), may have controlled diatom populations. The transition from spring diatom populations to summer greens, as mediated by SiO_2, temperature, and possibly other factors, may be reflected in the bimodal chlorophyll cycles.

The patterns of epilimnion- and hypolimnion-average concentrations of total inorganic P and ortho-P are shown in Figures 3.4-7 and 3.4-8. The response to detergent phosphate legislation in 1971 is evident. The relatively low degree of vertical stratification with regard to phosphorus in the earlier years of the study may reflect the fact that phosphorus levels were too high during this period to have been

Figure 3.4-6

Line Plot - Chlorophyll
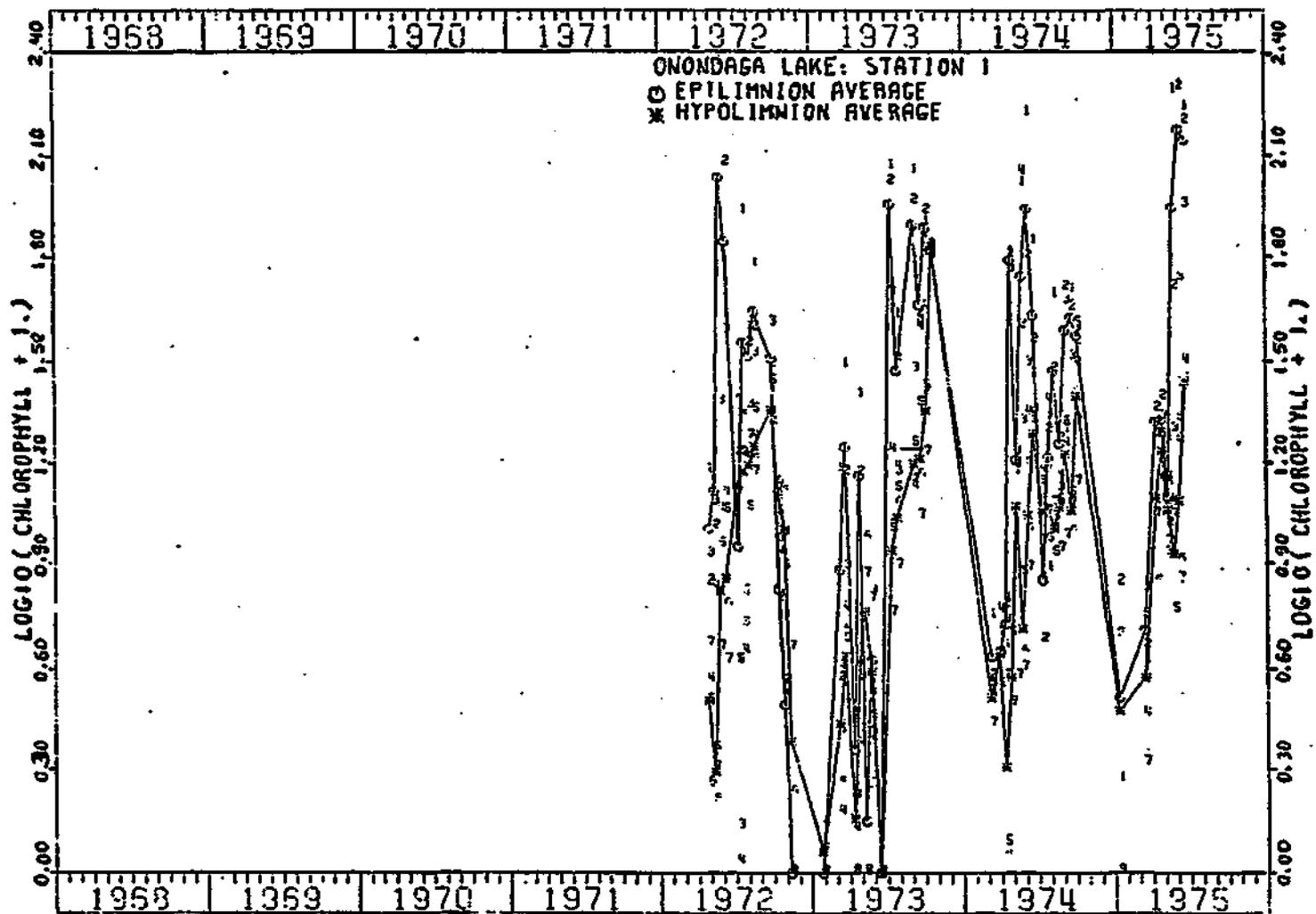
Figure 3.4-7

Line Plot - Total Inorganic Phosphorus

ONONDAGA LAKE: STATION 1
⊕ EPILIMNION AVERAGE
✳ HYPOLIMNION AVERAGE
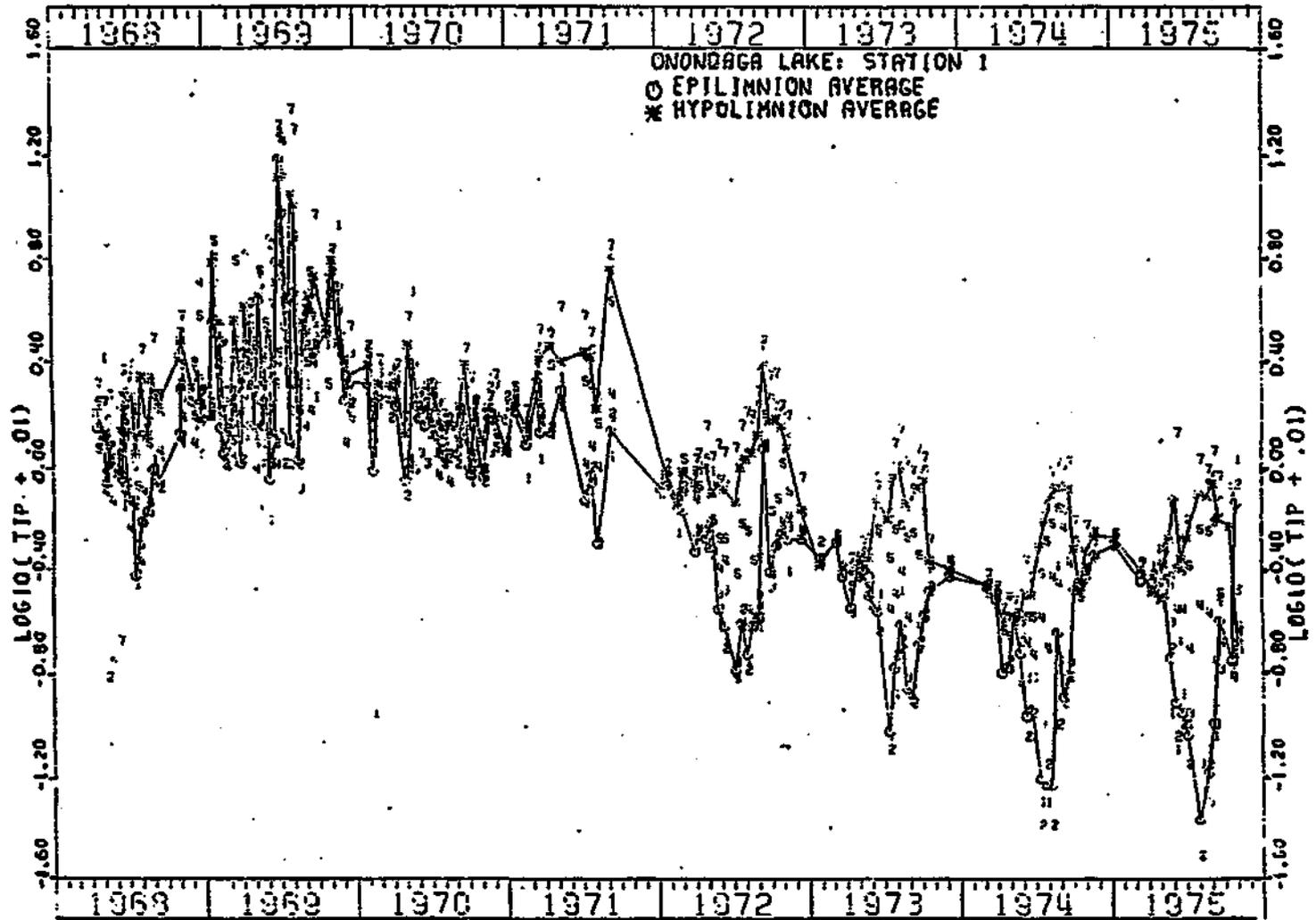
Figure 3.4-8

Line Plot - Ortho-Phosphorus
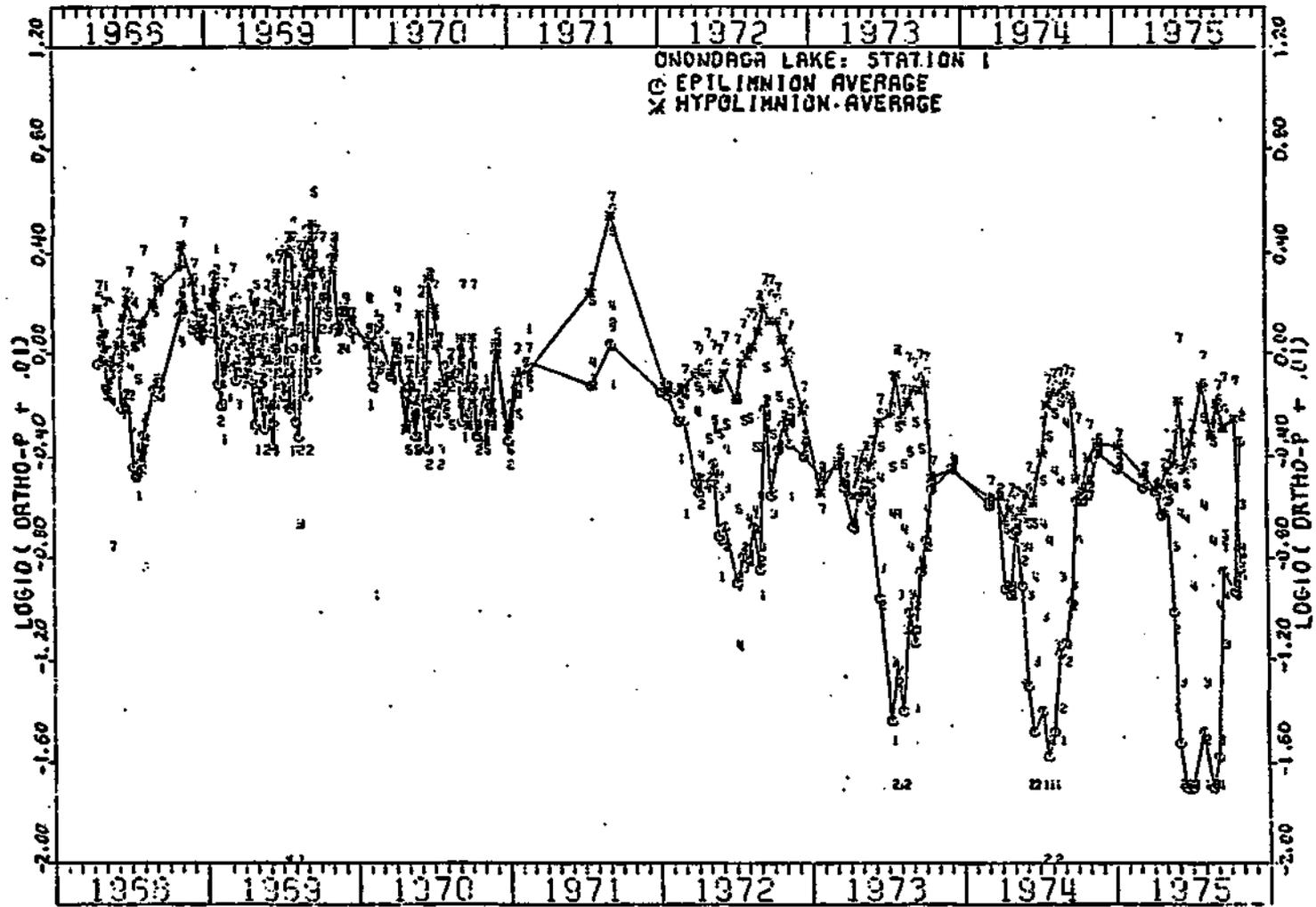
3.4.3 Phytoplankton and Nutrients (continued)

influenced significantly by algal uptake. In 1975, ortho-phosphorus averaged 0.01 g/cm^3 in the epilimnion on three mid-summer sampling dates, the lowest levels observed throughout the study period. Thomann et al.[35] have used a half-saturation constant of 0.005 g/cm^3 P in simulating the response of algal growth rate to phosphorus concentration. Using this value and the Michaelis-Mentem rate expression, a level of 0.01 g/cm^3 P would result in a 33% reduction in maximum algal growth rate due to phosphorus stress. Acknowledging the variability of the 0.005 g/cm^3 parameter, it seems possible that available phosphorus levels may have partially controlled algal growth rates in 1975.

Levels of ammonia nitrogen and carbon dioxide remained well above rate-limiting levels throughout the study period. It is evident that light was the primary rate-limiting resource. Mid-summer transparencies were generally on the order of one half to one meter.

The abundance of primary nutrients has supported a diverse collection of phytoplankton. Figure 3.4-9 presents semi-quantitative data on the phytoplankton and zooplankton populations between 1968 and 1974, as derived from the Onondaga Lake Study reports[20,24] and Sze[34]. Because of the wide variations in cell size from specie to specie, the cell number data in Figure 3.4-9 cannot be used alone to determine biomass distribution. In 1968, species were recorded as

Figure 3.4-9

Phytoplankton and Zooplankton Populations

PHYTOPLANKTON	1968	1969	1970	1971	1972	1973	1974	PHYLUM
CHLAMYDOMONAS SPP								I
SCHROEDERIA SETIGERA								
PEDIASTRUM BORYANUM								
PEDIASTRUM DUPLEX								
PEDIASTRUM SIMPLEX								
CHLORELLA VULGARIS								
OOCYSTIC PARVA								
ANKISTRODESMUS FALCATUS								
SCENEDESMUS BIJUGA								
SCENEDESMUS OBLIQUUS								
SCENEDESMUS QUADRICAUDA								
CLOSTERIUM GRACILE								
STAURASTRUM PARADOXUM								
EUGLENA PROXIMA								II, III
SYNURA UVELLA								
MELOSIRA GRANULATA								
CYCLOTELLA GLOMERATA								
CYCLOTELLA MENEGHINIANA								
COSCINODISCUS SUBTILIS								IV
DIATOMA TENUE								
FRAGILARIA CROTONENSIS								
SYNEDRA ULNA								
ASTERIONELLA FORMOSA								
AMPHIPRORA ALATA								
NITZSCHIA PALEA								
CERATIUM HIRUNDINELLA								V
GLENODINIUM PULVISCULUS								
CHROOMONAS NORDSTETII								VI
CRYPTOMONAS OVATA								
MICROCYSTIS AERUGINOSA								VII
ANABAENA CIRCINALIS								
ANABAENA FLOS-AQUAE								
APHANIZOMENON FLOS-AQUAE								
ZOOPLANKTON								
ROTIFERS								
COPEPODS								
CLADOCERANS								
VORTICELLA								
	1968	1969	1970	1971	1972	1973	1974	

SYMBOL LOG10 (ANIMALS/100 L)
- 0.0 - 0.9
- 1.0 - 1.9
- 2.0 - 2.9
- 3.0 - 3.9
- 4.0 - 4.9

SYMBOL LOG10 (CELLS/ML)
- PRESENT
- 2.0 - 2.9
- 3.0 - 3.9
- 4.0 - 4.9
- 5.0 - 5.9

PHYLUM
I	Chlorophyta	(Greens)
II	Euglenophyta	
III	Crysophyceae	
IV	Bacillariophyceae	(Diatoms)
V	Dinophyceae	(Dinoflagellates)
VI	Cryptomonadineae	
VII	Myxophyceae	(Blue-Greens)

3.4.3 Phytoplankton and Nutrients (continued)

present/absent only.

The phytoplankton population is shown to consist mainly of greens and diatoms. From 1968 to 1971, blue-greens (chiefly Aphanizomenon flos-aquae) dominated biomass in late summer, a fact not obvious in Figure 3.4-9 because of the greater size of blue-green cells relative to other types. The near-disappearance of blue-greens in latter years has been interpreted as a "positive" sign that conditions in the lake have been improving. Mechanistically, this has been attributed to lower levels of available phosphorus encountered after phosphate detergent legislation in 1971[20], the general hypothesis being that blue-greens were less able to compete for lower levels of phosphorus than were other algal types. Alternative explanations are discussed below.

The apparent decreasing trend in SiO_2 levels may reflect an increase in diatom activity. Two factors may have contributed to this. Diatoms have been shown to be particularly sensitive to chromium[10], the levels of which decreased markedly in 1973 (Appendices B and C). The lowering of chromium levels may have given the diatoms a competitive advantage over other algal types. Secondly, simulation studies by Bella[2] have shown that the degree of vertical mixing in a lake may influence interspecific competition among algal types. In particular, enhanced mixing tends to favor

3.4.3 Phytoplankton and Nutrients (continued)

faster-sinking algae (diatoms) over slower-sinking varieties (blue-greens). Thus, the increasing diatom activity and the near-disappearance of blue-greens may also be explained by the apparent decrease in vertical stratification, as discussed previously (Section 3.4.2).

Aside from the disappearance of blue-greens, no trend in algal species diversity is evident. On the order of twelve algal species were generally present in mid-summer samples throughout the monitoring period. In view of the abundance of nutrients and peculiar aspects of the lake (high heavy metal concentrations and salinity), the diversity of the phytoplankton population is considered surprisingly high[24,38]. Fish populations are likewise surprisingly diverse[24,38]. O'Brien and Gere[24] examined the temporal sequences of phytoplankton and zooplankton densities during the 1968-69 baseline study, noting that major shifts or drops in the phytoplankton population could not be explained by concomitant zooplankton increases. It was concluded that factors other than predation (possibly toxicity) were chiefly controlling the algal populations.

3.4.4 Dissolved Oxygen

The GRID display of dissolved oxygen in Appendix B shows the effects of aeration and primary production in the surface waters and

3.4.4 Dissolved Oxygen (continued)

respiration and oxygen depletion in the hypolimnion. The oxygen

data have been re-plotted as percent of saturation in Figure 3.4-10.

For each GRID cell, the oxygen saturation level has been computed

from the corresponding temperature and chloride level using the

following relationship derived from data in Fair, Geyer, and Okun[8]:

$$(DO)_{SAT} = \frac{448.4}{T + 30.04} + \frac{3.8 \times 10^{-3}C}{T + 22.8}$$

where,

T = temperature (degrees C)

C = chloride concentration (g/m^3)

The time- and space-averaged dissolved oxygen concentration during

the period amounted to 31% of saturation, reflecting the dominance

of heterotrophic activity induced by external and internal BOD

sources. Super-saturated levels of DO in the mid-summer epi-

limnion indicate regions of intense photosynthesis.

The data indicate a general trend toward improved conditions

in the hypolimnion, represented by shorter annual periods of

anaerobic conditions. At 18 meters, oxygen levels were below

Figure 3.4-10

DISSOLVED OXYGEN , PERCENT OF SATURATION

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (18 = SURFACE)

COMPONENT : DISSOLVED OXYGEN , PERCENT OF SATURATION

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF   4.00 AND   100.00 MEAN =   31.16 ST. DEV. =   28.39

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
MINIMUM   4.00   28.00   52.00   76.00
MAXIMUM   28.00   52.00   76.00   100.00

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL
                    25.00   25.00   25.00   25.00

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL
        LOW VALUES   HIGH VALUES
LEVELS   0   1   2   3   4   5

SYMBOLS

FREQUENCY   311   583   527   199   65   44

3.4.4 Dissolved Oxygen (continued)

1 g/m^3 for ten months of the year in 1970, compared with a five-month anaerobic period in 1974. This trend can be most easily seen by comparing hypolimnic dissolved oxygen levels in the early spring of successive years in Figure 3.4-10.

Line plots of volume-averaged dissolved oxygen levels are shown in Figure 3.4-11, smoothed according to the method outlined in Section 3.3.3. Oxygen levels plotted as DO , DO/DO_{SAT} , and $DO_{SAT}-DO$ all indicate an improving trend. Seasonal patterns are also evident and especially regular when the data are plotted as DO deficit ($DO_{SAT}-$ DO). The regular periodicity of these data may reflect the importance of photosynthetic oxygen sources, since minimum deficits generally occur during seasons of peak solar intensity and photosynthetic rates. Temperature effects on DO_{SAT} could also partially account for this periodicity.

Dissolved oxygen levels are generally determined by the balance, or imbalance between oxygen sources and sinks. Trends in the DO data should therefore result from trends in one or more of the oxygen sources or sinks. In general, no trends can be seen in the ambient levels of BOD or in the external loading rates. It is unlikely that the internal sources and sinks of DO have been altered much, since photosynthetic nutrient supplies generally remained in excess, except, possibly, in the case of ortho-phosphorus in 1975. The DO trend can

Figure 3.4-11

Line Plots - DO, DO/DO$_{sat}$, and DO$_{sat}$ - DO
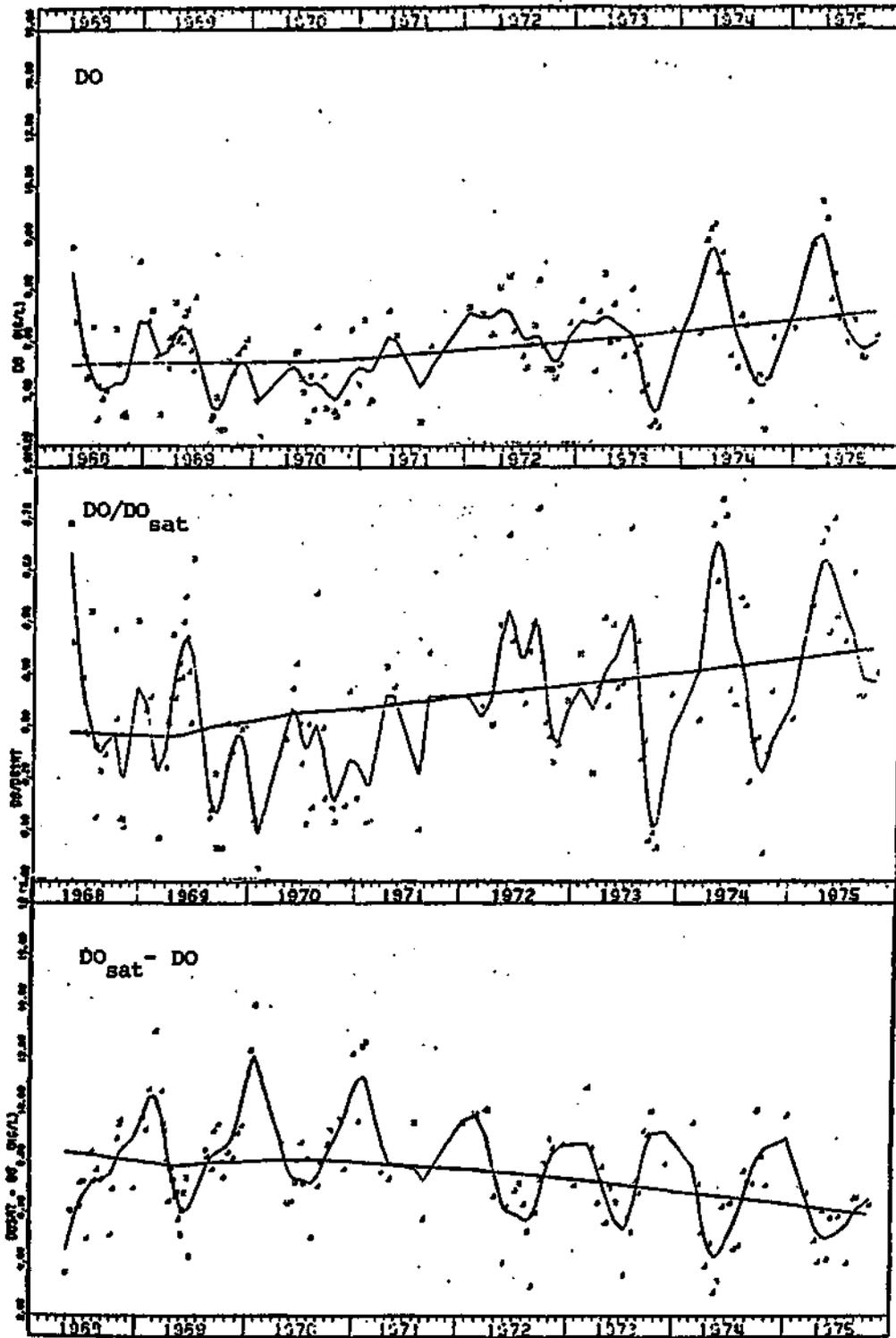
3.4.4 Dissolved Oxygen (continued)

best be explained by improved aeration of the bottom waters accompanying the decreasing trend in vertical stratification. Figure 3.4-12 compared the GRID displays of vertical density gradient and dissolved oxygen, as percent of saturation. The inverse relationship between density gradient and hypolimnic DO levels is apparent.

3.4.5 Mass Balances

The results of mass balance calculations performed on major water quality components are displayed in Appendix D, and summarized in Table 3.4-3. In the latter, the four terms of each mass balance (input, output, change in storage, and accumulation) are expressed in units of grams per square meter of lake surface per year. The retention coefficient (= accumulation/total input) and the fraction of the total input attributed to the Metro STP discharge are also given. Based upon examination of the mass balance figures in Appendix D, the component balances which exhibited trends have been broken into two or more time periods. "Trends" are indicated in the balance plots by changes in slope of the cumulative input, output, and/or accumulation lines. Division has been done at yearly intervals to incorporate seasonal effects.

As noted in Section 3.3.4, the epilimnion-averaged concentrations

Figure 3.4-12

DISSOLVED OXYGEN , PERCENT OF SATURATION[a]

DENSITY GRADIENT[b]
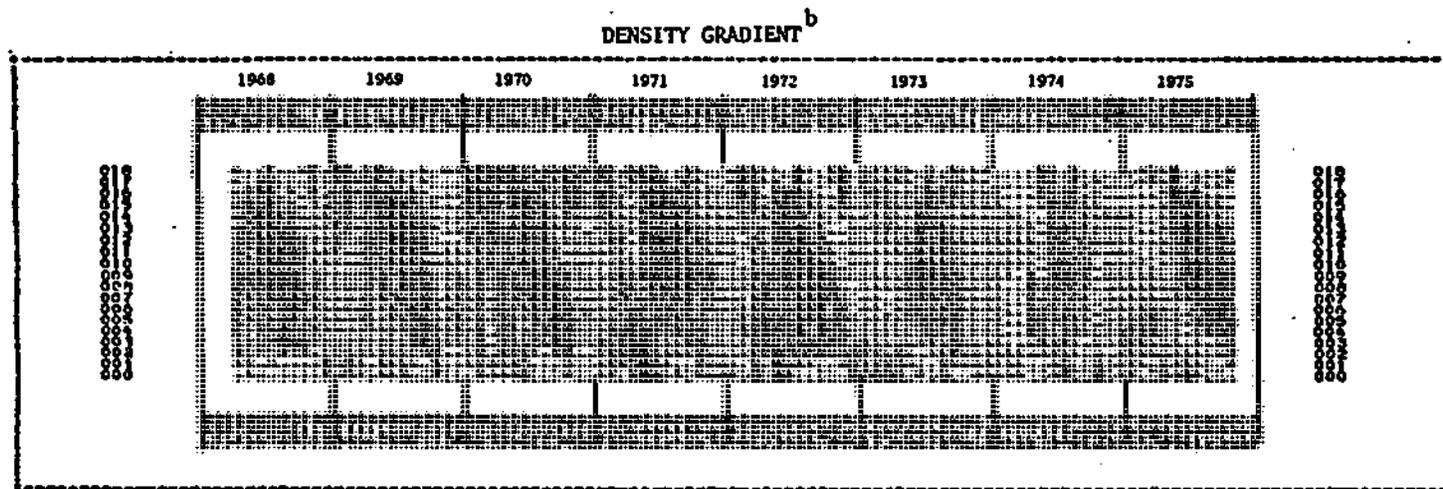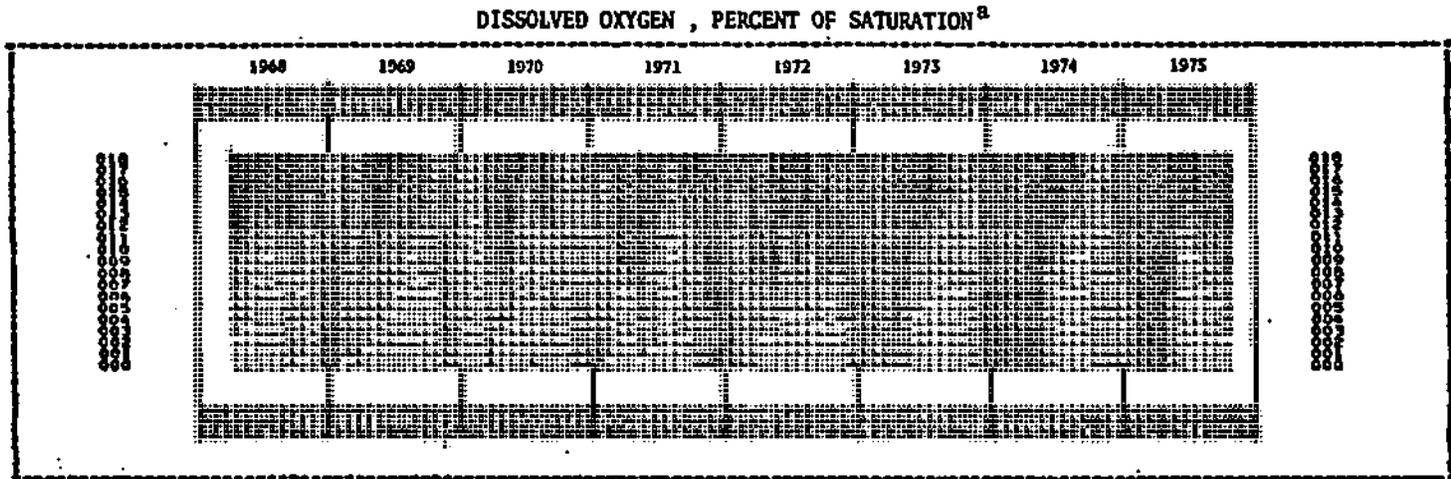
a - see Figure 3.4-10 for scale
b - see Figure 3.4-2 for scale

Table 3.4-3

Mass Balances - Onondaga Lake

Fluxes in grams per square meter of lake surface per year

Component	Water Years	Total Input	Total Output	Change in Storage	Accumulation	Retention[a] Coefficient	Fraction of Input due to Metro STP
Chloride	1970-74	72814.	70753.	322.	1739.	.024	.055
Nitrate and Nitrite N	1970-72	47.4	25.7 (28.0)[b]	.2	21.5 (19.2)	.453 (.405)	.120
	1973-74	24.5	20.8 (22.7)	-.3	4.0 (2.1)	.163 (.086)	.150
Total Kjeldahl N	1970-71	273.8	267.3 (241.8)	-.2	6.7 (32.2)	.025 (.118)	.434
	1972-74	262.3	189.1 (171.1)	+.8	72.4 (90.4)	.227 (.345)	.516
Total Nitrogen	1970-71	321.5	288.7 (263.8)	1.3	31.5 (56.4)	.098 (.175)	.387
	1972-74	294.9	214.9 (196.4)	0.0	80.0 (98.5)	.271 (.334)	.471
Silica (SiO$_2$)	1970-74	394.2	235.8	-6.3	164.7	.417	.247
Alkalinity (as Ca CO$_3$)	1970-74	10500.	9038. (8799.)	37.	1425. (1659.)	.136 (.158)	.162
Total Inorganic P	1970	156.7	79.4	-20.8	98.1	.626	.431
	1971	68.2	70.3	10.1	-12.2	-.179	.545
	1972-74	36.3	20.9	-7.6	23.0	.634	.743
Ortho-P	1970	48.0	39.5	-14.3	22.8	.477	.613
	1971	32.5	40.0	11.7	-19.2	-.593	.592
	1972-74	23.7	15.7	-5.2	13.2	.560	.697
Condensed Inorganic P	1970	108.7	39.9	-6.5	75.3	.693	.351
	1971	35.7	30.3	-1.6	7.0	.196	.502
	1972-74	12.6	5.2	-2.4	9.8	.778	.831
5-Day BOD	1970-74	1389.	289. (217.)	-5.	1105. (1177.)	.796 (.847)	.792

a Retention Coefficient = Accumulation/Total Input

b Corrected for lack of complete mixing; multiplied by ratio of Station 2 to Station 1 epilimnion - averaged concentrations (Table 7)

3.4.5 <u>Mass Balances</u> (continued)

at station 1 (see Figure 3.1-1) have been used to estimate lake outputs. Tests for horizontal mixing have indicated, however, that the lake is not completely mixed with regard to all components. The output fluxes in parenthesis in Table 3.4-3 have been corrected for this lack of complete horizontal mixing by multiplying the station 1 estimate by the ratio of the mean concentration at station 2 (closer to the lake outlet) to the mean concentration at station 1, as given in Table 3.4-1. This was not done for silica or the phosphorus components, since the differences between station 1 and 2 were not found to be statistically significant in these cases.

For each component, the accumulation term in the balance has been computed by subtracting output and change in storage from total input. This term thus represents the net sum of all fluxes which have not been measured or estimated directly, including sediment or atmospheric exchanges and chemically- or biologically-mediated transformations within the lake. The accumulation term also reflects possible errors in the input flux estimates due to sampling strategy relative to periodicities in point sources or intervening storm events in tributaries, as discussed in Section 3.2.1.

Over the five-year period, estimated chloride inputs balanced estimated outputs to within 2.4%. Since no mechanisms for chloride

3.4.5 Mass Balances (continued)

trapping within the lake are known to exist, the salt balance tends
to verify the assumptions, techniques, and data used in developing
the hydrologic balance and in estimating mass fluxes from grab-
sample concentration data. Ninemile Creek, which accepts the waste
bed overflow from the Allied Chemical Company's solvay plant, accounted
for 76.1% of the total chloride input to the lake over the five-year
period. Assuming that the background concentrations of chloride in
Ninemile Creek were the same as those in Onondaga Creek (averaging
275 g/m^3), about 68.3% of the total chloride flux can be attributed
to the solvay plant discharge. No trends in the chloride balance
are evident.

Fluxes of reduced nitrogen forms (Kjeldahl) dominate over those
of oxidized forms (nitrate and nitrite) by about an order of magni-
tude. Trends in the balances of each nitrogen form are evident.
Nitrate and nitrite inputs were reduced by about a factor of two
in 1973 and 1974, attributed primarily to reductions in loadings
from Onondaga Creek and Ninemile Creek. Nitrate and nitrite may be
produced in aerobic regions of the lake through nitrification and
lost in anaerobic regions through denitrification. Algal uptake may
be another important nitrate sink. The reduced net accumulation rates
of the oxidized nitrogen forms in the latter years may reflect in-
creased oxygen levels and enhanced nitrification.

3.4.5 <u>Mass Balances</u> (continued)

Trend in the Kjeldahl nitrogen balance is toward reduced output
rates and increased net accumulation rates. During water years 1970
and 1971, outputs balanced inputs to within 11.8%. During the 1972-
1974 period, however, 34.5% of the influent Kjeldahl nitrogen was
trapped with the lake or converted to oxidized forms. Enhanced
aeration could have accounted for this increase. The disappearance
of blue-green algae over this same period may also have eliminated
an important internal Kjeldahl nitrogen source.

The trends and relationships in the total nitrogen balance are
similar to those of the Kjeldahl nitrogen balance. The increase in
the percent of total nitrogen loading due to the Metro STP discharge
from 38.7% to 47.1% may be partially attributed to the Onondaga
County combined sewer maintenance program, which reportedly elimi-
nated several dry-weather sewer overflows into Onondaga Creek and
Harbor Brook during this period[14]. In addition, the Ley Creek
Sewage treatment plant discharge was diverted from Ley Creek to the
Metro STP at the end of 1969[24].

The silica balance was relatively constant throughout the
five-year period. An estimated 43.7% of the influent silica was
trapped within the lake. Mechanisms for silica removal include
uptake by diatoms and subsequent deposition, and formation and
sedimentation of silica-containing minerals[24].

3.4.5 Mass Balances (continued)

The alkalinity balance likewise appears to be relatively stable. On the average, 15.8% of the influent alkalinity was removed. The primary reaction accounting for this is:

$$Ca^{++} + 2 HCO_3^- \; \underset{\textstyle \sim}{\overset{\textstyle +}{\rightleftharpoons}} \; \underset{\textstyle calcite}{CaCO_3 \downarrow} + CO_2 + H_2O \qquad (3.4-13)$$

with "alkalinity" labeling $2 HCO_3^-$ and "calcite" labeling $CaCO_3\downarrow$.
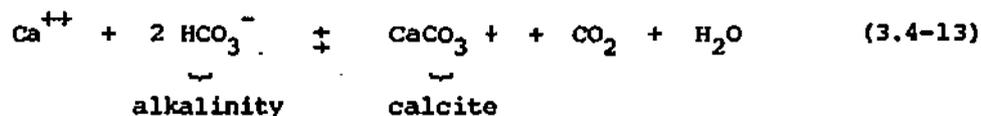
The high levels of calcium in the lake drive this reaction to the right in nearly all regions and seasons, except where pH falls below 7[24].

Table 3.4-3 presents balances on three phosphorus forms: TIP, ortho-P, and condensed inorganic P (= TIP - ortho-P). The response to phosphate detergent legislation in 1971 is evident. The time period has been broken into three parts: pre-legislation (1970), transition (1971), and post-legislation (1972-74). Comparison of pre- and post-legislation periods indicates a reduction of 77% in the TIP loadings*. The TIP retention coefficients, however, remained

---

* This 77% reduction is surprisingly high, in view of the amount of total phosphorus which has been attributed to detergents in sanitary wastes ($\sim$ 56%)[29]. Some of this reduction could have been due to the county's combined sewer maintenance program which was being carried out simultaneously with the detergent legislation. In addition, the pre- and post-legislation phosphorus measurements do not include the organic phosphorus fraction, which would not be influenced by detergent formulation.

3.4.5 <u>Mass Balances</u> (continued)

relatively constant, at values of 62.6% and 63.4%, respectively. During the transition period, internal sources of TIP and ortho-P are apparent, in that the computed retention coefficients are negative. This could have resulted from exchange from sediments or with organic phosphorus forms in the water column during this period. This type of transient response is in tune with a model proposed by Lorenzen et al.[16] for simulating the response at Lake Washington to phosphorus diversion. Mechanisms for phosphorus removal include algal uptake and deposition and chemical precipitation. The lake is supersaturated with calcium phosphate, and formation and precipation of hydroxyapatite and fluorapatite are expected to be important mechanisms for phosphorus removal[24]. The relative stability of such minerals to changes in redox potential (unlike iron-phosphorus compounds) tends to indicate that release of soluble phosphorus from the sediments in significant amounts would be unlikely. The fact that the pre- and post-legislation TIP retention coefficients are similar indicates first-order behavior of the phosphorus removal mechanism and suggests that chemical precipitation may be more important than algal uptake and sedimentation. The former mechanism is first-order and the latter, zero-order in phosphorus concentration, under conditions in which phosphorus is not limiting algal growth.

The BOD_5 balance appears to have been relatively stable over

3.4.5 Mass Balances (continued)

the five-year period. The Metro STP discharge, determined from the
Onondaga County plant performance records[23], accounted for 79.2%
of the total external loading of BOD_5. Average outflow amounted to
15.3% of the external loading. Consideration of internal loadings
resulting from primary production can provide a much more complete
picture of the organic matter and oxygen balances of the lake.

Gross primary production has been estimated using data provided
by O'Brien and Gere[20]. Light/dark bottle studies were done on four
days between May and September in 1974. The average daily gross
producitivity was estimated at 6.1 grams of carbon per square meter
per day (observed range 3.8 - 7.4 $g-C/m^2$-day). Assuming that 85% of
the yearly production occurred in the 150-day May - September period[4],
an annual gross production of 1080 $g-C/m^2$-year is estimated. Using
conversion factors given by Byrlinsky and Mann[4], this corresponds
to 2900 $g-O_2/m^2$-yr and 10150 Kcal/m^2-yr. At an annual average visible
solar radiation intensity of 3200 kcal/m^2-day[22], this corresponds to
a photosynthetic efficiency of 0.9%.

Brylinski and Mann[4] have studied the relationships among phyto-
plankton standing crop, gross productivity, and photosynthetic
efficiency using the International Biological Program data base.
The results of the above calculations have been compared with those
of Brylinski and Mann in Figure 3.4-13. While Onondaga Lake is shown
to be highly productive, the relationships among phytoplankton standing

Figure 3.4-13

Relationships Among Phytoplankton Standing Crop, Gross Photosynthesis,
and Photosynthetic Efficiency for Onondaga Lake Compared with Results
of Brylinsky and Mann[a]

M. BRYLINSKY AND K. H. MANN
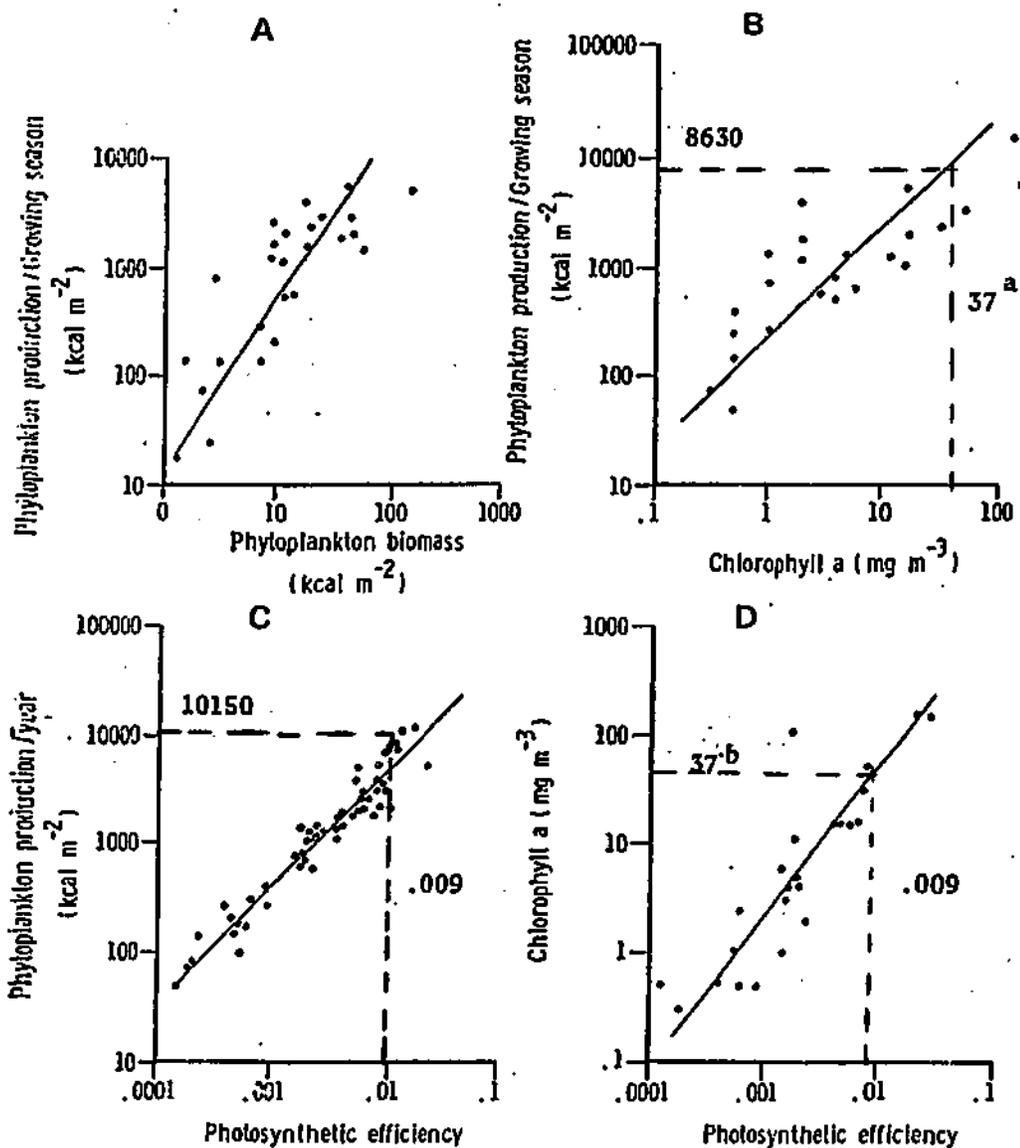
Fig. 10.   Relationship of (A) phytoplankton production to phytoplankton biomass, (B) phytoplankton
production to chlorophyll a concentration, (C) phytoplankton production to photosynthetic efficiency,
and (D) chlorophyll a to photosynthetic efficiency.

a - Epilimnion average concentration in Onondaga Lake, May-Sept., 1974.

3.4.5 Mass Balances (continued)

crop (as chlorophyll), primary production, and photosynthetic effi-
ciency are close to what one would expect based upon Brylinski and
Mann's studies of a wide range of lakes. Besides providing perspec-
tive on Onondaga Lake, this lends some strength to the productivity
estimate derived above.

Using this figure to represent internal sources of oxygen
demand, an annual-average ultimate oxygen demand balance has been
formulated (Table 3.4-4). The ratio of 5-day to ultimate carbon-
aceous BOD has been assumed to be 0.65, corresponding to a first
order decay rate of 0.21 day^{-1}. Nitrogenous oxygen demands have
been calculated from the TKN balances, assuming the following
stoichiometry for the nitrification reaction:

$$NH_3 + 2 O_2 \rightarrow NO_3^- + H^+ + H_2O \qquad (3.4-14)$$
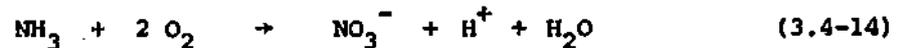
Table 3.4-4 shows that external sources of oxygen demand only
slightly exceed internal (photosynthetic) sources on an annual
average basis. During productive seasons, internal sources would
be expected to dominate. The total net loading to the lake, after
subtracting outflow, amounts to 14 $g-O_2/m^2$-day, or, assuming an
average depth of 12 meters, about 1.2 $g-O_2/m^3$-day. Since gross

Table 3.4-4

Onondaga Lake
Ultimate Oxygen Demand Balance

Average Annual Conditions (1970-74)

	Areal Basis $(g\text{-}O_2/m^2\text{-day})$	Volumetric Basis $(g\text{-}O_2/m^3\text{-day})$
External Sources		
Carbonaceous	5.85	0.49
Nitrogenous	3.28[*]	0.27
Internal Source		
Gross Photosynthesis	7.91	0.66
Total Input	17.04	1.42
Outflows		
Carbonaceous	0.92	0.08
Nitrogenous	2.14[*]	0.18
Total Outflow	3.06	0.26
Net Demand	13.98	1.16

* 1972-74

3.4.5 Mass Balances (continued)

photosynthesis would supply a maximum of 7.9 $g-O_2/m^2$-day, assuming no atmospheric losses due to super-saturation, remaining oxygen sources or organic matter sinks would have to account for a minimum of 6.1 $g-O_2/m^2$-day or 0.5 g/m^3-day of oxygen demand. Oxygen sources include aeration, and nitrate and sufate reduction. Sedimentation of particulate organic matter may be an important oxygen demand sink, depending upon the extent to which this oxygen demand is exerted in a benthic form.

O'Brien and Gere[24] have attempted to quantify benthic oxygen demands by measuring oxygen uptake in disturbed sediment cores. Measured values ranged from 0.0024 to 0.0034 $g-O_2/m^2$-day, well below the magnitude of external and other internal sources of oxygen demand. However, these data are suspect; Newbold and Ligget[19] reviewed the literature on benthic oxygen demands in lakes and streams and concluded that a value in the range of 0.1 to 1.0 $g-O_2/m^2$-day was appropriate for Cayuga Lake. This range is about 50 times higher than that measured in Onondaga Lake. It seems unlikely that the Onondaga values would be so much lower, particularly considering its reportedly extensive organic sludge deposits[24]. On the other hand, the precipitation of calcium carbonate in Onondaga may serve to blanket the sludge deposits and prevent exertion of oxygen on demand in the overlying water. Calcite precipitation may also be an important mechanism for removal of yellow and humic organic compounds from the water column

3.4.5 Mass Balances (continued)

via an adsorption/sedimentation process[25]. If Newbold and Ligget's
figures are used, the benthic component of the oxygen demand balance
may be appreciable. However, since the source of the organic matter
exerting the benthic demand is in the water column, the net balance
in Table 3.4-4 is not affected.

Tertiary treatment of the Metro STP effluent to remove BOD and
phosphorus is a measure designed to reduce both external and internal
(photosynthetic) sources of oxygen demand. Based upon the calcula-
tions outlined above, the maximum impact of this plan on the oxygen
demand budget of Onondaga Lake is presented in Table 3.4-5. In
order to put an upper limit on the effects of phosphorus reduction,
gross photosynthesis has been assumed to be reduced in proportion to
the total phosphorus loading. This assumes that production is phos-
phorus-limited and ignores nutrient recycling effects. Accordingly,
maximum reduction of 59% in the total oxygen demand loading is
estimated.

3.5 Eutrophication Assessment

3.5.1 Phosphorus as a Controlling Factor

In order to develop some perspective on Onondaga Lake from a
eutrophication viewpoint, its phosphorus balance has been compared

Table 3.4-5

Onondaga Lake

Present and Projected Ultimate Oxygen Demand Loadings

(grams O_2/m^2-day)

	Average Conditions (1970-74)	Future Conditions (1980)
External Sources		
Carbonaceous	5.85	1.53
Nitrogenous	3.28	3.28
Internal Source		
Gross Photosynthesis	7.91	2.15*
Total	17.04	6.96

* Assuming primary production reduced in proportion to total
  inorganic phosphorus loading; represents a lower limit.

3.5.1 Phosphorus as a Controlling Factor (continued)

with balances of a wide variety of other northern temperate lakes.
This comparison has been done in the contexts of empirical models
for predicting lake trophic state as a function of phosphorus loading
and morphologic indices. The models employed have been discussed
and evaluated in detail in Chapter 2. They include the two models
of Vollenweider[39,40], that of Dillon and Rigler[6], and the model
developed from the stepwise discriminant analysis in Chapter 2.

Table 3.5-1 presents the "permissible" (oligotrophic/mesotrophic)
and the "dangerous" (mesotrophic/eutrophic) phosphorus loading levels
for Onondaga Lake as estimated by these various models. These levels
can be compared with estimated past, present, and future TIP loading
also presented in Table 3.5-1. The projected TIP loading for 1980 is
still over twice the highest "dangerous" level estimated by any of
the models.

Two factors render these empirical models of limited value for
application to Onondaga. First, the models are all based upon
measurements of total phosphorus, whereas only total inorganic
phosphorus readings were available in this case. Second,   the
uniquely-high calcium content of Onondaga promotes phosphorus removal
via calcium phosphate precipitation and sedimentation[24]. This re-
moval mechanism was generally not present to the same degree in the
lakes upon which the empirical models were based. For this reason,

Table 3.5-1

Allowable Phosphorus Loadings for Onondaga Lake According to Various Models

(grams / m^2 - year)

Model	"Permissible"	"Dangerous"
Vollenweider I	0.11	0.22
Vollenweider II	0.74	1.48
Dillon and Rigler	1.48	2.96
ESP	0.66	1.90

Observed and Projected Loadings

Water Year	TIP Loading	% METRO STP
1970	156.7	43 %
1971	68.9	54 %
1972	42.2	76 %
1973	38.7	67 %
1974	28.7	81 %
Projected*	7.8	30 %

* assuming 90 % reduction in 1974 METRO STP loading

3.5.1 Phosphorus as a Controlling Factor (continued)

the permissible and dangerous loadings specified by the various models may be too conservative. To some extent, Dillon's model may account for the effects of the additional phosphorus removal mechanism by employing the observed retention coefficient (0.63). Accordingly, the loadings specified by that model are higher than those specified by the others. The fact that the observed retention coefficients before and after detergent legislation were similar (see Table 3.4-3) suggests the validity of the first-order assumption inherent in the Dillon model. However, it is still uncertain whether the relationship between lake trophic state and phosphorus concentration holds in the unusual chemical environment of Onondaga Lake, i.e., the target levels of 0.02 and 0.01 $g-P/m^3$ may not be realistic.

An indication of the importance of calcium phosphate precipitation as a phosphorus removal mechanism can be derived by comparing the observed retention and phosphorus sedimentation coefficients with those estimated using the retention model developed in Chapter 2:

$$1 - RE = 1/(1 + 0.824 \ T^{.454}) = 0.70$$

$$KE = 0.824 \ T^{-.546} = 1.86 \ year^{-1}$$

The corresponding observed values are given by:

3.5.1 Phosphorus as a Controlling Factor (continued)

$$1 - RO = 0.37$$

$$KO = RO/[(1 - RO)T] = 7.53 \text{ year}^{-1}$$

The standard error of estimate for the above retention coefficient model is 0.137. The following t-test suggests that the observed and estimated retention coefficients are significantly different:

$$t = (RO - RE)/SEE_R = (0.63 - 0.30)/0.137 = 2.41$$

These calculations indicate that the retention coefficient in Onondaga is about twice what one would expect, based upon the model developed on data from other northern temperature lakes. The effective first-order sedimentation coefficient is about 4 times the expected value, suggesting the importance of additional phosphorus removal mechanisms in Onondaga. Some of these differences may be attributed to the fact that only inorganic phosphorus measurements were available. Since the organic faction in the lake's outflow has been ignored, the actual total phosphorus retention coefficient may be somewhat lower.

3.5.1 Phosphorus as a Controlling Factor (continued)

As noted above, the Dillon model roughly accounts for any unusual phosphorus removal mechanisms in Onondaga. The projected loading for 1980 is still over twice the dangerous level predicted by that model. Since the Metro STP will account for only 30 percent of the projected loading, some rather extreme non-point source control measures may have to be implemented in order to reach acceptable loading levels. Estimation of confidence limits on the tributary and point-source loadings and consideration of combined sewer overflows (which may not be adequately reflected in the tributary loading estimates) would be required in order to provide an adequate basis for projecting the impacts of non-point source control measures. Generally, the results of this analysis supports the EPA's conclusion that "the best reason for instituting phosphorus removal at the Metropolitan Syracuse STP is the protection of Lake Ontario"[38].

3.5.2 Light as a Controlling Factor

In Chapter 2, the rationale behind the Lorenzen-Mitchell model[15] for predicting the potential effects of mixing upon algal production in lakes and impoundments was discussed. A scheme for estimating the parameters of the model from transparency observations was also presented. The model is applied below to estimate the potential effects of mixing under present and future nutrient loading regimes.

3.5.2 Light as a Controlling Factor (continued)

According the development in Chapter 2, peak, light-limited biomass, expressed in terms of grams of chlorophyll-a per square meter, is given by:

$$C_L^{max} \; Z \;\; = \;\; \frac{\mu^m F}{r\beta} - \frac{\alpha}{\beta} \; Z \qquad\qquad (3.5-1)$$

The corresponding expression for nutrient-limited biomass is:

$$C_N^{max} \; Z \;\; = \;\; \frac{N_o}{\gamma} \; Z \qquad\qquad (3.5-2)$$

where,

C_L^{max} = light-limited biomass (g Chl-a/m^3)

C_N^{max} = nutrient-limited biomass (g Chl-a/m^3)

Z = mixed depth (m)

μ^{max} = maximum growth rate of algae (day^{-1})

r = respiration rate of algae (day)$^{-1}$

β = incremental light extinction coefficient due to algae (m^2/g Chl-a)

α = background extinction coefficient (m^{-1})

F = light/depth integral

N_o = limiting nutrient concentration at spring overturn (g/m^3)

3.5.2 Light as a Controlling Factor (continued)

$$y = \text{nutrient content of algae} \\ \text{(g nutrient/g Chl-a)}$$

The model suggests that peak biomass·will be controlled by light or by nutrients, according to whether $C_L^{max} z$ is less than or greater than $C_N^{max} z$, respectively. In Chapter 2, it was shown that the parameters of the model could be estimated from the following relationships:

$$F = 2.718\lambda/24 \qquad\qquad (3.5-3)$$

$$\alpha = 1.44/z_s^{max} \qquad\qquad (3.5-4)$$

$$\frac{\mu^m}{r} = 1.44 \, z/F \, z_s^{min} \qquad\qquad (3.5-5)$$

where,

$$\lambda = \text{day length at peak biomass (hours)}$$

z_s^{max} = maximum Secchi depth observed in the course of a year (m)

z_s^{min} = minimum Secchi depth observed in the course of a year (m)

3.5.2 Light as a Controlling Factor (continued)

Transparency observations in Onondaga Lake over the 1968-74 period are plotted in Figure 3.5-1. These generally range from about 0.5 meter to 2.5 meters, although some year-to-year variations in these ranges are apparent. During this period, nutrient concentrations were in excess, so the assumption is made that a 0.5 meter Secchi depth corresponds to peak, light-limited biomass. For a thermocline depth of 9 meters, the average epilimnion depth is estimated as the ratio of volume to surface area, or 7.15 (see Figure 3.3-1). Accordingly, the following parameters are assumed:

$$Z_s^{min} = 0.5 \text{ meter}$$

$$Z_s^{max} = 2.5 \text{ meters}$$

$$Z = 7.15 \text{ meters}$$

$$\lambda = 15 \text{ hours}$$

Substituting into equations (3) to (5):

$$F = 1.70$$

$$\alpha = 0.58 \text{ m}^{-1}$$

$$\frac{\mu^m}{r} = 12.1$$

Figure 3.5-1

Onondaga Lake Transparency Measurements
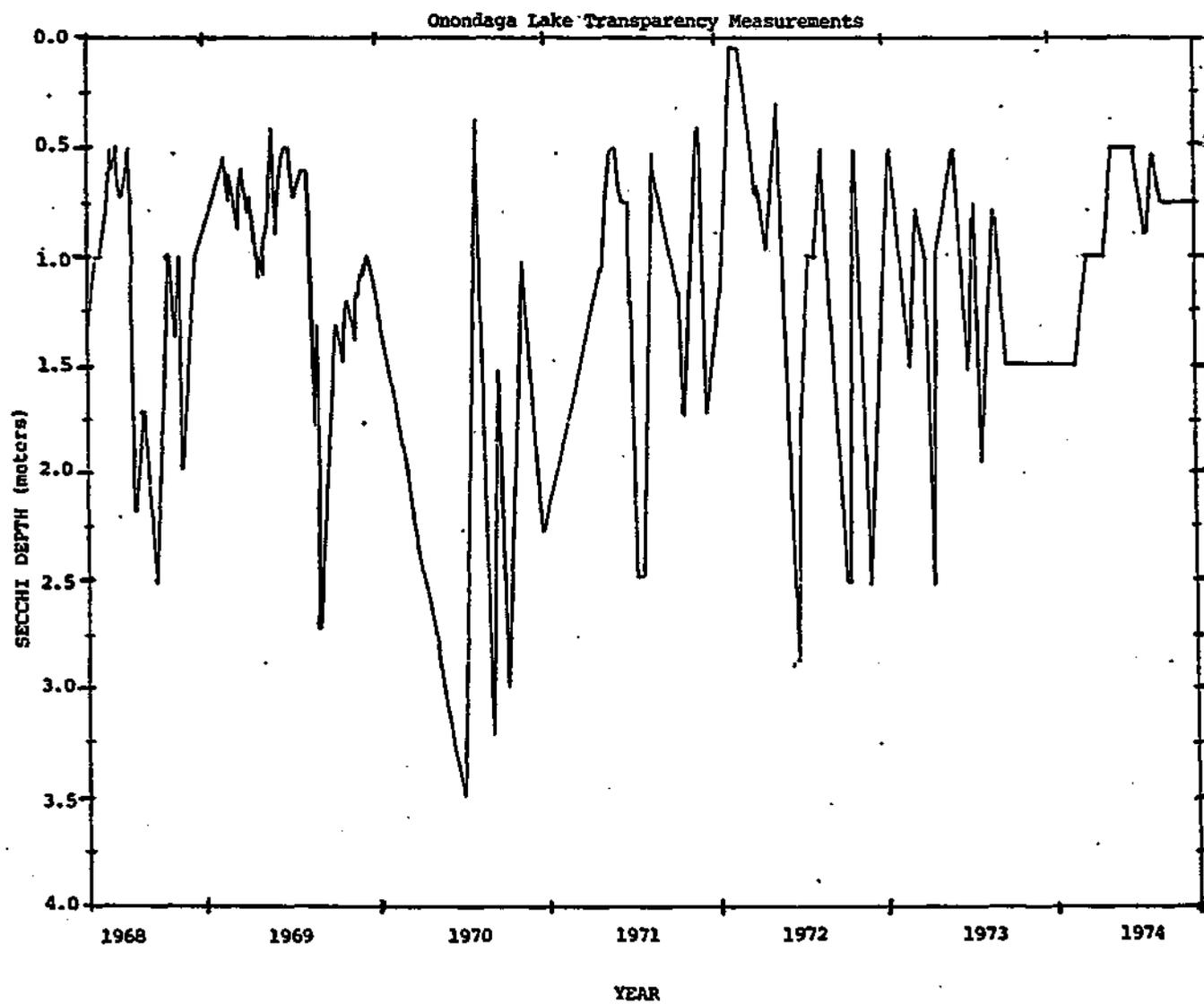
3.5.2 Light as a Controlling Factor (continued)

The estimated value of $\dfrac{\mu_m}{r}$ is within the range of values reported in the literature[7]. Assuming that phosphorus is the potential limiting nutrient, the remaining parameters have been derived from the literature:

$$\beta \;=\; 20 \; m^2/g \; Chl\text{-}a \qquad\qquad (ref.:\; 1,\; 26)$$

$$\gamma \;=\; 1 \; g\text{-}P/g \; Chl\text{-}a \qquad\qquad (ref.:\; 7,\; 26)$$

Thus, the expressions for light- and nutrient-limited biomass are:

$$C_L^{max} \; Z \;=\; 1.03 - 0.029 \; Z \qquad\qquad (3.5\text{-}6)$$

$$C_N^{max} \; Z \;=\; N_o \; Z \qquad\qquad (3.5\text{-}7)$$

These lines are plotted in Figure 3.5-2 for values of N_o, the total inorganic phosphorus concentration at spring overturn, of 1.58, 0.32, and 0.086 g/m^3. These concentrations correspond to 1970, 1974, and 1980 conditions, respectively. The figure for 1980 was estimated as the observed value for 1974 (Figure 3.4-7) times the ratio of 1980 to 1974 TIP loadings (Table 3.5-1).

Figure 3.5-2

Light- and Nutrient-Limited Biomass under Various Conditions
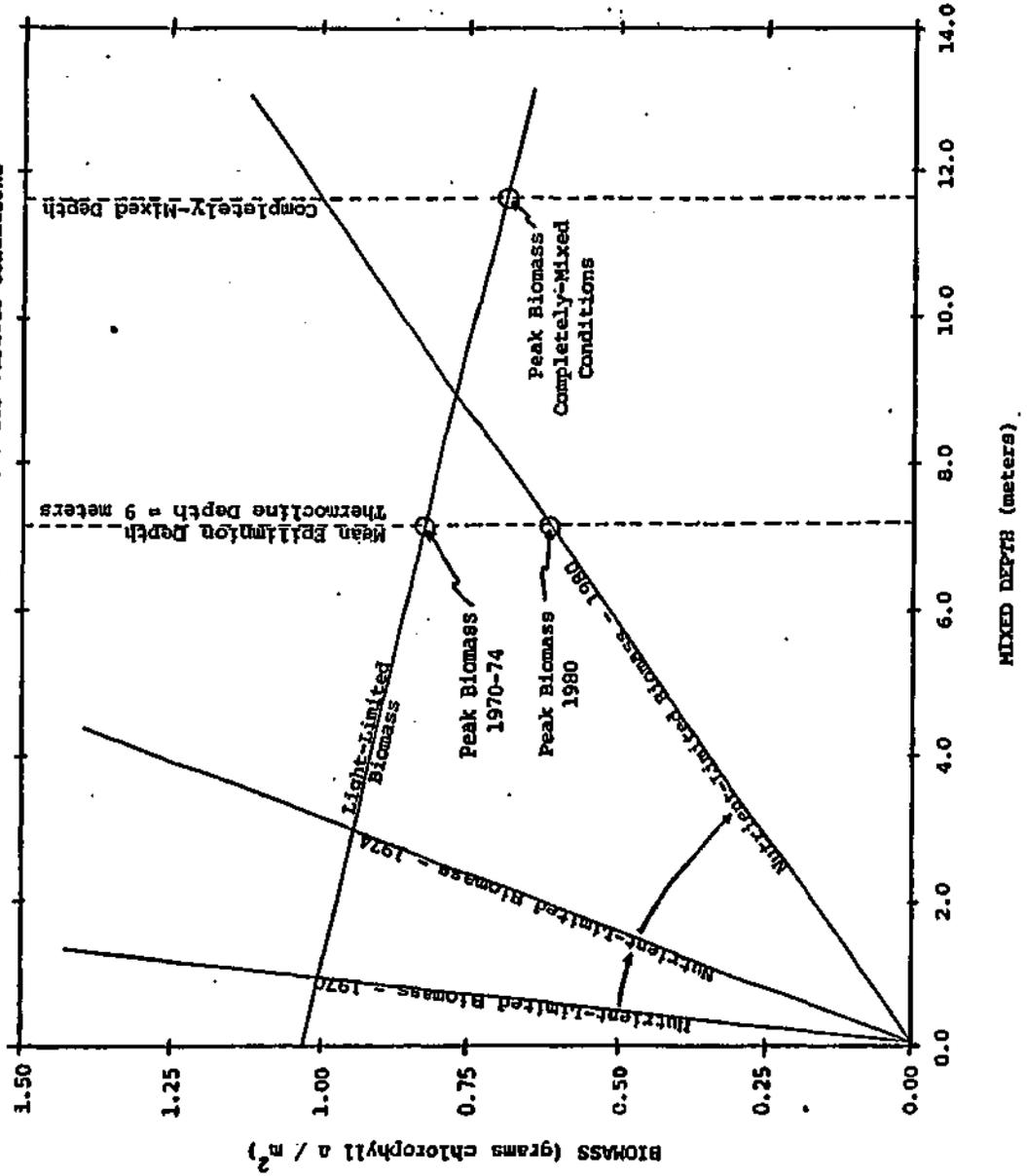
3.5.2 Light as a Controlling Factor (continued)

Figure 3.5-2 indicates that under current mixing regimes
(Z = 7.12 meters) and past and present phosphorus loading, peak
biomass should be light-limited. The intersection of the Z = 7.15
meters line with the light-limited biomass line occurs at a CZ
value of 0.82 g Chl-a/m^2. This suggests that peak biomass concen-
trations should be in the vicinity of 0.82/7.15 = 0.115 g Chl-a/m^3,
or 115 micrograms Chl-a/liter. Observed peak biomass levels for
the 1972-75 period ranged from about 90 to 150 and averaged 110
micrograms per liter (Figure 3.4-6). These observations are not
inconsistent with model predictions.

The model indicates that the 1980 loading reduction should
result in a nutrient-limited peak algal biomass level of 0.62 g Chl-a/
m^2, corresponding to a volumetric concentration of 87 micrograms/liter,
or a 24% reduction of current levels. In order to bring peak biomass
concentrations down to less than 30 micrograms Chl-a/liter, another
66% reduction in phosphorus loadings would be required. These conclu-
sions are not greatly different from those derived from use of the
Dillon model in the previous section, which suggested that a 62%
reduction in the 1980 loadings would be required to reach "dangerous"
loading levels (Table 3.5-1).

The fact that peak biomass will be nutrient-limited under 1980
conditions suggests that increases in mixing rates may enhance

3.5.2 Light as a Controlling Factor (continued)

productivity somewhat. If it is assumed that enhanced mixing effectively increases mixed depth, Onondaga would lie somewhere between current conditions ($Z = 7.15$ meters) and the completely-mixed conditions ($Z = 11.7$ meters). A maximim production of 0.77 g Chl-a/ m^2 would be realized at an effective depth of 9 meters, while the completely-mixed condition would result in a light-limited peak biomass of 0.70 g Chl-a/m^2. Thus, the model indicates that, from the point of view of total biomass production, no benefits would be derived from mixing the lake under the estimated 1980 loading conditions.

Under light-limited conditions, the expression for Secchi depth is:

$$
\left.\begin{aligned}
(Z_s^{min})_L &= \frac{1.44 \, rZ}{\mu^m F} \quad , \quad Z \le Z_c \\
&= \frac{1.44}{\alpha} \quad , \quad Z > Z_c
\end{aligned}\right\} \quad Z_c = \frac{\mu^m F}{r\alpha} \qquad (3.5\text{-}8)
$$

Under nutrient-limited conditions,

$$
(Z_s^{min})_N = \frac{1.44}{\alpha + \dfrac{\beta N_o}{\gamma}} \qquad (3.5\text{-}9)
$$

3.5.2 Light as a Controlling Factor (continued)

For the parameter values estimated above:

$$(z_s^{min})_L = 0.070 \ z \ , \qquad z \leq 35.5 \qquad \qquad (3.5\text{-}10)$$

$$\qquad \qquad = 2.50 \qquad , \qquad z > 35.5$$

$$(z_s^{min})_N = \frac{1.44}{0.58 + 20 \ N_O} \qquad \qquad (3.5\text{-}11)$$

Under 1980 conditions and the current mixing regime, it was estimated above the peak biomass would be nutrient-limited. Accordingly, the minimum Secchi depth would be given by:

$$z_s^{min} = (z_s^{min})_N = \frac{1.44}{0.58 + 20(0.086)} = 0.63 \ meter$$

This would represent a slight improvement over current conditions ($z_s^{min} = 0.5$ meter). It was also estimated above that if the lake were completely mixed ($z = 11.7$ meters), peak algal biomass would be light-limited. Accordingly, under these conditions:

$$z_s^{min} = (z_s^{min})_L = 0.070 \ (11.7) = 0.82 \ meters$$

3.5.2 Light as a Controlling Factor (continued)

Thus, mixing the lake under 1980 phosphorus loading conditions would increase minimum transparency by about 30%. This change would probably be difficult to measure with a Secchi disc in this range, although it could be detected with light transmission measurements. The two conditions examined above correspond to volumetric biomass concentrations of 87 and 59 micrograms Chl-a per liter, respectively, or a potential 31% decrease as a result of mixing.

Generally, in order for mixing alone to be successful as a means of restricting peak algal biomass to less than 30 micrograms $Chl-a/m^3$, equation (1) indicates that the following criterion must be satisfied:

$$C_L^{max} = \frac{\mu^m F}{r\beta \bar{Z}} - \frac{\alpha}{\beta} \leq 0.030 \qquad (3.5-12)$$

where,

$\bar{Z}$ = mean epilimnion depth (m)

= mean lake depth, under completely mixed conditions

For the parameter estimates employed above, this expression reduces to:

3.5.2 Light as a Controlling Factor (continued)

$$\bar{z}\,(1 + 1.67\alpha) \geq 34.3 \qquad\qquad (3.5\text{-}13)$$

or,

$$\bar{z}\,(1 + \frac{2.40}{z_s^{max}}) \geq 34.3 \qquad\qquad (3.5\text{-}14)$$

Lakes with greater mean depths or with greater background (non-algal) extinction coefficients would benefit more from artificial destratification. In the case of Onondaga, the value of the left side of equation (14) is 22.9, suggesting that additional controls on nutrient inputs would be necessary in order to restrict peak algal biomass to less than 30 micrograms of chlorophyll-a per liter.

3.6 Conclusions

1. Significant differences in concentrations of some water quality components between the two lake monitoring stations indicate that the epilimnion at station 1, closer to the Metro STP discharge, is a relatively reduced environment. These differences are small compared with vertical and seasonal variations.

2. Seasonal vertical stratification patterns reflect the effects of density and of chemical and biological reactions occurring in the epilimnion and hypolimnion.

3. A trend toward reduced density (temperature and chloride) stratification in the lake is possibly attributed to climatologic variations. This trend would indicate enhanced vertical mixing rates, in turn consistent with the following observations:

    a. enhanced hypolimnic oxygen levels;
    b. virtual disappearance of blue-green algae.

4. The apparent reduction in the blue-green algal population over the course of the monitoring period might be attributed to a number of factors possibly influencing interspecific competition, including:

    a. reduction in phosphorus levels;
    b. reduction in chromium levels;
    c. enhanced vertical mixing rates.

    If the last factor is chiefly responsible, there is no reason why the blue-green population would not return in

3.6 Conclusions (continued)

the future, since the enhanced mixing rates have been attributed to climatologic variations.

5. While nutrients may have controlled species dominance with season, the algal population appears to have been light-limited under the nutrient loading regimes of 1968-74.

6. Estimates of chloride inputs developed from water quality and quantity data balance outputs to within 2.4% over water years 1970-74. 68.3% of the total chloride flux can be attributed to Allied Chemical's waste bed overflow into Ninemile Creek.

7. Over this same period, fluxes of reduced nitrogen forms dominated over those of oxidized forms by about an order of magnitude. Apparent trends toward higher net accumulation rates of Kjeldahl N and lower accumulation rates of Nitrate and Nitrite N may reflect: (1) enhanced nitrification, conceivably a result of higher oxygen levels; (2) lower nitrogen fixation rates, resulting from the disappearance of blue-green algae.

8. Phosphorus balances indicate the following:

a. The average loading in water years 1972-74 was 23% of that observed in water year 1970. This 77% reduction can be attributed to the effects of:

(1) phosphorus detergent legielation in 1971;
(2) Onondaga County's combined sewer maintenance program;

3.6 Conclusions (continued)

(3) diversion of raw sewage formerly discharged
into Ley Creek to the Metro STP in late 1969.

b. The lake apparently retained 63% of the influent TIP
during each of the above periods, suggesting the
importance of a first-order reaction for phosphorus
removal in the lake. Under steady-state conditions,
the TIP retention coefficient and effective first-
order sedimentation coefficient were two and four
times, respectively, what one could expect, based
upon analysis of phosphorus balance data from other
northern temperate lakes. These results, along with
chemical equilibrium considerations, suggest the
importance of calcium phosphate precipitation as a
dominant removal mechanism.

c. Apparently negative accumulation rates and retention
coefficients for TIP and ortho-P during water year
1971 may reflect the net release of inorganic P from
the sediment or from organic P fractions in the water
column during this transition period, in which loadings
and ambient concentrations were markedly reduced.
Sediment release is considered relatively unlikely
because of the stability of apatite.

9. Under average 1972-74 conditions, external and internal
(photosynthetic) sources of ultimate oxygen demand were
about equal and totalled 17 g-O_2/m^2-day, or 1.4 g-O_2/m^3-day.

10. Installation of secondary and tertiary treatment facilities
at the Metro STP is estimated to:

3.6 <u>Conclusions</u> (continued)

      a. reduce ultimate oxygen demand loadings by a maximum
of 59%;

      b. reduce TIP loadings by a maximum of 73%, resulting in
nutrient-limited peak algal biomass levels of about
76% of those observed under past and present light-
limited conditions.

11. Since future algal populations are projected to be nutrient-
limited, no reduction in total biomass production would be
realized by attempting to control available light through
lake mixing.

12. Additional controls on nutrient inputs may have to be
implemented in order to bring peak biomass levels down to
mesotrophic levels. In this effort, further data and
analysis are required in order to assess the importance of
phosphorus loadings originating in combined sewer overflows,
which may not be adequately reflected in the calculations
outlined above.

REFERENCES - CHAPTER 3

1. Bannister, T.T., "Production Equations in Terms of Chlorophyll Concentration, Quantum Yield, and Upper Limit of Production", *Limnology and Oceanography*, Vol. 20, No. 3, pp. 342-364, May 1975.

2. Bella, D.A., "Simulating the Effect of Sinking and Vertical Mixing on Algal Population Dynamics", *Journal of the Water Pollution Control Federation*, Vol. 42, No. 5, Part 2, pp. R140-R152, May 1970.

3. Blanton, J.O., "Vertical Entrainment into the Epilimnia of Stratified Lakes", *Limnology and Oceanography*, Vol. XVIII, No. 5, pp. 697-704, September 1973.

4. Brylinsky, M. and K.H. Mann, "An Analysis of Factors Governing Productivity in Lakes and Reservoirs," *Limnology and Oceanography*, Vol. 18 , No. 1, January 1973.

5. Council on Environmental Quality, *Environmental Quality*, Sixth Annual Report, U.S. Government Printing Office, p. 396, 1975.

6. Dillon, P.J., "The Phosphorus Budget of Cameron Lake, Ontario: The Importance of Flushing Rate to the Degree of Eutrophy of Lakes", *Limnology and Oceanography*, Vol. 20., No. 1, pp. 28-45, Jan. 1975.

7. DiToro, D.M., D.J. O'Connor, R.V. Thomann, and R.P. Winfield, "Mathematical Modelling of Phytoplankton in Lake Ontario 1. Model Development and Verification", EPA-660/3-75-005, March 1975.

8. Fair, G.M., O.C. Geyer, and D.A. Okun, *Water and Wastewater Engineering*, Vol. 2, John Wiley & Sons, New York, 1968.

9. Harleman, D.R.F. and A.T. Ippen, "Salinity Instrusion Effects in Estuary Shoaling", *Journal of the Hydraulics Division, American Society of Civil Engineers*, Vol. 95, No. HY1, pp. 9-27, January 1969.

10. Hervey, R.J., "Effect of Chromium on the Growth of Unicellular Chlorophyceae and Diatoms", *Botanical Gazette*, Vol. III, No. 1, September 1949.

11. Hutchinson, G.E., *A Treatise on Limnology*, Vol. 1, John Wiley and Sons, 1957.

12. International Business Machines, Inc., *IBM System/360 and System/ 370, Subroutine Library - Mathematics*, Program Product 5736-XM7, 1974.

REFERENCES - CHAPTER 3 (continued)

13. *International Critical Tables*, National Research Council, Edward W. Washburn, Ed., First Edition, McGraw-Hill, 1930.

14. Lawler, D., Onondaga County Department of Drainage and Sanitation, personal communication, 1975.

15. Lorenzen, M. and R. Mitchell, "Theoretical Effects of Artificial Destratification in Impoundments", *Environmental Science and Technology*, Vol. 7, No. 10, pp. 939-944, October 1973.

16. Lorenzen, M.W., D.J. Smith, and L.V. Kimmel, "A Long Term Phosphorus Model for Lakes: Application to Lake Washington", presented before the Division of Environmental Chemistry, American Chemical Society, Philadelphia, April 1975.

17. Malueg, K.W., P. Larsen, D.W. Schults, and H.T. Mercier, "A Six-year Water, Phosphorus, and Nitrogen Budget for Shagawa Lake, Minnesota", *Journal of Environmental Quality*, Vol. 4, No. 2, pp. 236-242, 1975.

18. Meyer, A.F., *Evaporation from Lakes and Reservoirs*, Minnesota Research Committee, St. Paul, Minnesota, June 1947.

19. Newbold, J.D. and J.A. Liggett, "Oxygen Depletion Model for Cayuga Lake", *Journal of the Environmental Engineering Division, American Society of Civil Engineers*, Vol. EE1, pp. 41-59, February 1974.

20. O'Brien and Gere, Engineers, "Onondaga Lake Monitoring Program", Annual Reports to Onondaga County, New York, Department of Public Works, 1970-74.

21. O'Brien and Gere, Engineers, "Supplement to Wastewater Facilities Report - Metropolitan Sewage Treatment Paint Expansion", prepared for Onondaga County, April 1971.

22. Odum, E.P., *Fundamentals of Ecology*, Third Edition, W.B. Saunders Co., 1971.

23. Onondaga County, New York, Department of Drainage and Sanitation, "Performance Reports for the Metropolitan Sewage Treatment Plant", 1968-74.

24. Onondaga County, New York, "Onondaga Lake Study", US EPA, Water Quality Office Publication 11000 FAE 4/71, April 1971.

25. Otsuki, A. and R.G. Wetzel, "Calcium and Total Alkalinity Budgets and Calcium Carbonate Precipitation in a Small Hard-Water Lake", *Arch. Hydrobiol.*, Vol. 73, No. 1, pp. 14-30, February 1974.

REFERENCES - CHAPTER 3 (continued)

26. Process Research, Inc., "Final Report on the Storrow Lagoon Demonstration Project", prepared for the Metropolitan District Commission, Commonwealth of Massachusetts, 1975.

27. Reinsch, C.H., "Smoothing by Spline Functions", Numerical Mathematics, Vol. 10, No. 117, 1967.

28. Rooney, J., "Evaluation of Interim Basin Plan, IBP-NY-07-07, for Lake Onondaga, N.Y. and Project WPC-NY-659 (Syracuse Metro STP)", EPA Region II Office, New York, 1973.

29. Shannon, E.E., "Effects of Detergent Formulation on Wastewater Characteristics and Treatment", Journal of the Water Pollution Control Federation, Vol. 47, No. 10, pp. 2371-2383, October 1975.

30. Sinton, D. and C. Steinitz, "GRID, Version 3", Laboratory for Computer Graphics and Spatial Analysis, Harvard University, October 1971.

31. Spirtas, R. and H.J. Levin, "Patterns and Trends in Levels of Suspended Particulate Matter", Journal of the Air Pollution Control Association, Vol. 21, No. 6, pp. 329-333, June 1971.

32. Stefan, H. and D.E. Ford, "Temperature Dynamics in Dimictic Lakes", Journal of the Hydraulics Division, American Society of Civil Engineers, Vol. HY1, pp. 97-114, January 1975.

33. Stumm, W. and J.J. Morgan, Aquatic Chemistry, Wiley-Interscience, New York, 1970.

34. Sze, P., "The Phytoplankton of Onondaga Lake, N.Y.", Ph.D. Thesis, Cornell Univeristy, Limnology, 1972.

35. Thomann, R.V., D.M. DiToro, and D.J. O'Connor, "Preliminary Model of Potomac Estuary Phytoplankton", Journal of the Environmental Engineering Division, American Society of Civil Engineers, Vol. EE3, pp. 699-715, June 1974.

36. US Department of Commerce, Weather Bureau, "Local Climatological Data", Syracuse, New York, Hancock Airport, 1968-74.

37. US EPA, "Effluent Limitation Guidelines and Standards of Performance and Pretreatment for Inorganic Chemicals Manufacturing, Point Source Category 40 CFR, Part 415", Federal Register 38/1967: 28174-28194, 1973.

38. US EPA, Region II, "Environmental Impact Statement on the Wastewater Treatment Facilities Construction Grants for the Onondaga Lake Drainage Basin", May 1974.

REFERENCES - CHAPTER 3 (continued)

39. Vollenweider, R.A., "Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, with Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication", Office of Economic Cooperation and Development, Paris, September 1970.

40. Vollenweider, R.A., "Input-Output Models", unpublished manuscript, 1973.

41. Wold, S., "Spline Functions in Data Analysis", Technometrics, Vol. 16, No. 1, pp. 1-11, February 1974.

4.0 A MODEL FOR VERTICAL STRATIFICATION IN ONONDAGA LAKE

The previous chapter has relied primarily upon monitoring data
as a basis for analysis of Onondaga Lake water quality problems.
This chapter employs some of these data in the context of a modelling
study which illustrates some of the general parameter estimation,
sensitivity analysis, and error analysis techniques discussed in
Chapter 1. The focus of the study has been selected: (1) to be based
upon some of the key relationships observed among the water quality
variables in Chapter 3, and (2) to provide input to specific design
decisions which will have to be made by Onondaga County in the
near future, as discussed below.

4.1 Objectives - The Outfall Design Issue

The dissolved oxygen standard for Class B (contact recreational)
waters in New York State is 4 mg/liter. This standard is often violated
in the epilimnion, as well as the hypolimnion of Onondaga Lake, and
compliance with this standard is a primary water quality management
objective. In Chapter 3, an apparently increasing trend in dissolved
oxygen levels over the course of the monitoring period was noted.
This trend was shown to correlate with an apparently decreasing trend
in density stratification. The suggestion that density stratification
was partially controlling hypolimnic oxygen levels is consistent with
results of a modelling study by Bella[5], who has shown that hypolimnic
DO levels in stratified levels are generally most sensitive to vertical

4.1 Objectives - The Outfall Design Issue (continued)

diffusion rates. These rates, in turn, depend upon density structure.

In Chapter 3, it was also estimated that future installation of tertiary treatment facilities would result in a maximum reduction of about 60% in the total oxygen demand loadings to the lake, including both external and internal (photosynthetic) sources. The response of ambient DO levels to this reduction will depend upon the availability of oxygen resources to satisfy this demand, i.e.,upon reaeration rates, which, in turn, will be sensitive to density stratification through its controlling influence on vertical mixing rates. This raises concern over the potential impacts of the plan to combine the Metro STP effluent with the saline waste from Allied Chemical in order to effect phosphorus removal. Currently, the latter is discharged to Ninemile Creek, which affords about twice as much dilution on the average as will the Metro STP effluent. The plan will double the average density difference between the saline stream and ambient waters at the point of entry into the Lake. By increasing the tendency of the effluent to sink into the hypolimnion, the plan may influence density stratification and vertical mixing rates. This tendency to sink could be reduced by discharging through a diffusing system which would force immediate dilution of the effluent in the ambient waters and thereby reduce the gravitational driving force for a density current.

The proposed method of discharging the tertiary effluent from the Metro STP is through a surface, shoreline outfall[22]. Because of

4.1 Objectives - The Outfall Design Issue (continued)

the salinity of the Allied Chemical waste, the combined municipal/
industrial effluent is expected to have a chloride concentration in
excess of 7000 mg/l, compared with an average lake concentration of
1700 mg/l. Under all realistic temperature regimes, the effluent will
be much more dense than the ambient lake waters at the point of dis-
charge. The tendency of the effluent to sink into the hypolimnion
will depend upon the degree of initial dilution of the effluent in
the ambient waters in the vicinity of the outfall site, as well as
upon lake and discharge temperatures. The shoreline, surface outfall
design has grown partially out of concern for the potential water
quality impacts of discharging a relatively dense effluent[22]. Economic
considerations have also been of importance, the shoreline discharge
alternative being undoubtedly the least expensive one[22].

James Rooney of EPA Region II[26] has reviewed and criticized this
aspect of the Syracuse plan. In Rooney's discussion, reproduced below,
the environmental reasons for the proposed shoreline discharge are
cited and an opposing case for an offshore outfall design is made:

"The IBP and reference material do not adequately substantiate
the desirability of the proposed surface-shoreline discharge. The
items raised in support of this discharge method were as follows:

   a) '...to ensure distribution of the treatment plant effluent
       in the epilimnetic waters of the lake,' and thereby provide
       maximum mixing of the plume with ambient water and some degree
       of bio-degradation before sinking into the hypolimnion.

   b) '...to prevent the accumulation of any organic material in
       the hypolimnetic waters of the lake' and thereby preclude the
       possibility that the productive epilimnetic volume of the lake
       would decrease.

4.1 Objectives - The Outfall Design Issue (continued)

    c) '...to prevent the tendency...to create a permanent stratification within the lake.'

The studies completed to date, however, are inconclusive with respect to the water quality impact of the proposed discharge and do not demonstrate the potential for any of the aforementioned responses. On the contrary, it appears that the discharge alternative consisting of a submerged outfall would probably be less detrimental to the lake quality and aesthetics due to the following considerations:

    a). A submerged discharge would lessen the severity of the existing hypolimnetic conditions by the introduction of relatively high DO waters (5 mg/l) into a region which is essentially devoid of oxygen for approximately 7 months out of the year; the discharge, if located in the general vicinity of the thermocline (9 - 12 meters depth), would through proper design, also enhance the entrainment of some hypolimnetic waters into the epilimnion; this could possibly provide for a limited decomposition and subsequent reduction of the residual organic load in the hypolimnion where mean BOD_5 values are on the order of 9.0 mg/l.

    b). A submerged outfall will preclude the potential for future aesthetic degradation which may present a considerable problem due to the proposed introduction of the Allied Chemical wastewater into the municipal system; the plume visibility aspect and the related possibility of nearshore discoloration were not discussed in any of the reference material when evaluating the merits of surface-shoreline discharge.

    c). A submerged outfall will, by its very nature, provide greater initial dilutions than a surface discharge of comparable design; thus, the area affected by the relatively high concentrations of the effluent parameters (pH, ammonia, chlorides, total dissolved solids [TDS]) will be minimized under the submerged discharge alternative; in addition, the area most severely affected by the potentially adverse concentrations of these constituents will be located in the relatively unproductive waters of the hypolimnion.

    d). The submerged discharge alternative will preclude the possibility of any accumulation of wastewater constituents along the shallow region (3 feet depth) adjacent to the proposed plant site; the effluent dispersion model adapted for the plume analysis[22] does not indicate that this may be a problem, however, the utilization of this model is questionable in that (1) the model itself and its application to non-thermal discharges is unverified, (2) the model does not account for the limiting effects of impingement and reduced entrainment of ambient waters due to the shallow nature of the lake ($\leq$ 3 feet) near the discharge point, and, (3) the model evaluates the proposed discharge under quiescent conditions and does not address potentially more critical conditions which may

4.1 Objectives - The Outfall Design Issue (continued)

occur due to specific meteorologic and/or hydrologic conditions.

In summary, there is insufficient data available at the present time concerning the lake hydrodynamics, in general, and circulation patterns in the vicinity of the proposed shoreline discharge, in particular, to permit adequate projections on the water quality response from the surface discharge alternative. It is likely, however, that due to the presence of the shoal area adjacent to the plant site, the natural assimilative capacity and the potential for significant entrainment of ambient waters into the proposed effluent plume will be minimal while the possibility for aesthetic damage will be maximized. Based on these conclusions, it appears that a submerged (or surface) discharge located further out in the lake would be the most practical (and conservative) alternative for the enhancement of water quality in the lake."

In an Environmental Impact Statement on the project, the EPA[37] discussed the advantages and disadvantages of discharging to the epilimnion and to the hypolimnion of the lake. The greater availability of oxygen in the surface waters was considered an important justification for the surface discharge; however, it was considered likely that the dense effluent would sink into the hypolimnion before bio-oxidation of the relatively inert organic matter in the discharge. Because of uncertainty in the initial dilution rates and changes in lake and discharge temperatures, the report concluded that is was "impossible to predict how long the effluent would remain in the epilimnion".

The major disadvantage of the surface discharge was considered to be the plume visibility which would result from calcium carbonate precipitation. This reaction would occur when the excess calcium in the effluent mixes with alkalinity in the ambient lake waters. Control of effluent pH would limit this problem to some degree. The possibility was raised that a subsurface discharge could disturb the lake's bottom

4.1 Objectives - The Outfall Design Issue (continued)

sediments, releasing nutrients, organic material, and mercury, although
this was largely unsubstantiated.

The EPA[37] concluded that the major advantages of the surface outfall
were low construction and maintenance costs. It was estimated that
capital costs for a subsurface outfall would amount to an additional
one million dollars. The flexibility of the shoreline, surface discharge
alternative was also cited as an advantage. If problems (probably
aesthetic) were to arise with the design, the outfall could be extended
off-shore.

The outfall design problem has apparently been an issue of some
controversy. In this chapter, a model is developed and applied for the
purpose of addressing aspects of this issue. The study does not address
the effects of specific details of outfall design, but rather it is
concerned with general design specifications. The two major "decision
variables" studied are: (1) discharge location (epilimnion or hypolimnion),
and (2) degree of initial dilution. The model is concerned with vertical
mixing rates in the lake and formulates heat and mass balances on the
epilimnion and hypolimnion to simulate variations in lake temperatures
and chloride levels. It could serve as a basis for future development
of a more general model for predicting non-conservative water quality
components. As noted above, this work also provides a context in which
to illustrate some of the parameter estimation, sensitivity analysis,
and error analysis techniques discussed in Chapter 1.

4.2   Review of Lake Vertical Stratification Models

One-dimensional models for vertical density stratification in lakes and reservoirs can be grouped into two general categories: dispersion models and mechanical energy balance models. In the former, the vertical transport of heat is modelled as a combination of diffusive, convective, and advective processes, based upon the mass-transport equation. In the latter, the vertical mixing process is governed by the relationship between wind-induced kinetic energy in the surface layers and the buoyant potential energy deficit resulting from density stratification. Of the two groups, the former have generally seen more widespread application in lake and reservoir modelling. The various discrepancies in the assumed dominant mechanisms both within and between the two groups of models reflect the fact that there is still some general disagreement as to the relative importance of various vertical transport mechanisms.

4.2.1   Mass Transport Models

Three major models of the first type have been developed by people from MIT[8,11,13,14], Cornell University[30,31,32,33], and Water Resources Engineers (WRE)[42,24]. Table 4.2-1 is a summary of the dominant features of the models, as derived from an evaluation by Parker et al.[25]. The only mixing mechanism shared by all three of these models is the convective (instantaneous) mixing

Table 4.2-1

Essential Aspects of Deep Reservoir Temperature Models Evaluated by Parker et al.[25]

Aspect	Cornell Model[30,31,32]	M.I.T. Model[11,13]	W.R.E. Model[24,42]
Surface Boundary Conditions Specified	One of the following: (1) water surface temperature (2) heat flux at surface (3) equilibrium temperature Friction velocity (All B.C. input in sinusoidal form)	Meteorologic data:[b] (1) short-wave solar radiation (2) wind speed (3) cloud cover (4) relative humidity (5) air temperature	Meteorologic data :[b] (1) short-wave solar radiation (2) wind speed (3) cloud cover (4) relative humidity (5) atmospheric pressure (6) wet bulb temperature (7) dew point (8) air temperature
Diffusivity	Eddy diffusivity $K_H = K_{H_0} (1 + \sigma R)^{-1}$ $R = - (g z^2 \alpha_v / \omega_*^2) \, \partial T / \partial z$ $K_{H_0} = (c_1 + c_2 z) \omega_*$ ω_* = friction velocity z = depth α_v = coefficient of volume expansion σ, c_1, c_2 = empirical parameters	Molecular diffusivity	Empirical diffusion coefficient $D_c = \lambda_1$, $E < E_c$ $D_c = \lambda_2 \, E^{\lambda_3}$, $E \geq E_c$ $E = \frac{-1}{\rho} \frac{\partial \rho}{\partial z}$ = stability criterion $\lambda_1, \lambda_2, \lambda_3$ = empirical parameters E_c = critical stability parameter
Hydrologic Boundary Conditions	Constant volume No net inflows or outflows	Variable volume Influent streams enter at point of density equivalence, after initial dilution	Variable volume Influent streams enter at point of density equivalence
Time Step	Variable	Adjusted during simulation, based upon numerical stability criterion Typically 1 day	Constant for a given simulation Value based upon numerical stability criterion Typically 1 day
Depth Increment	Constant Maximum 100 intervals Typically 5 feet	Constant 20< Intervals < 50 Typically 1 - 2 meters	Constant Maximum 200 intervals Typically 1 - 2 meters
Running Time[a]	11.86 min/year Δy = 5 feet	1.7 min/year $\Delta y = 2$ m , $t = 1$ day 2.4 min/year $\Delta y = .1$ m , $t = 1$ day	3.2 min/year $\Delta y = 2$ m , $t = 1$ day 3.6 min/year $\Delta y = 1$ m , $t = 1$ day

a - CPU time exclusive of program loading ; computer model not specified, but the same for all.
b - Meteorologic data input on a daily basis; some variables computed internally as functions of others.

4.2.1 <u>Mass Transport Models</u> (continued)

which occurs in regions of instability, as determined by the sign of
the density gradient. Solutions to the governing partial differential
equations are achieved by finite difference techniques. All are
characterized by relatively high degrees of spatial and temporal
resolution (on the order of 1 meter and 1 day, respectively), and
are accordingly relatively expensive to implement. The three models
differ primarily in the mechanisms responsible for the vertical trans-
port of heat during stably stratified periods.

In the Cornell and WRE models, diffusion (eddy or "effective",
respectively) is the dominant transport mechanism. Application of
either of these models requires specification of a number of
parameters required to compute the time- and depth-variable diffusion
coefficient as a function of some measure of stability (the Richardson
number, in the case of the Cornell model, or normalized density gradi-
ent in the case of the WRE model). Neither of these models has been
applied extensively enough in order to establish the validity of
the functional forms or parameter values over a range of reservoirs.
Hence, the parameters must be viewed as empirical and may need to be
re-estimated for each new model application.

The MIT model, on the other hand, employs molecular diffusion
and advection as dominant vertical transport mechanisms during
stably stratified periods. Wind-induced mixing is ignored and mixing

4.2.1 Mass Transport Models (continued)

due to an unstable density profile accounts for convection in the
epilimnion. One advantage of the MIT model is that it does not
require the specification of empirical parameters for calculation
of diffusion coefficients, since the molecular diffusion coefficient
is constant.

In comparing the three models, Parker et al.[25] concluded that
the MIT model was preferable, based upon ease of application and
predictive capability. However, this conclusion may be biased,
since the predictive capabilities of the three models were compared
using data from Fontana Reservoir, the same reservoir used as a
data source in the development of the MIT model. After correcting
some apparent coding errors in the WRE diffusion coefficient calcu-
lation scheme, Parker et al. found that the WRE predictions were
relatively insensitive to an order of magnitude increase in the
effective diffusivity coefficient. Thus, the uncertainty in the
eddy diffusivity parameters may not be a serious drawback to appli-
cation of this model. The primary drawback of the WRE model was
its relative sensitivity to assumed depth increment over the range
tested (0.6 m - 2 m). The major disadvantages of the Cornell model
were found to be: (a) its failure to account for variable depth
versus surface area profile in the lake; (b) its lack of ability
to consider advective flows through the reservoir (and the vertical
mixing attributed to them); and (c) an apparent over-sensitivity of

4.2.1 Mass Transport Models (continued)

the calculated eddy diffusivity to wind velocities over the lake.

4.2.2 Mechanical Energy Balance Models

The second major group of models characterizes vertical mixing in a lake by considering the relationship between kinetic energy input at the surface due to wind shear stress and the change in buoyant potential energy which occurs with mixing. The development of density stratification is attributed to the differential absorption of thermal energy in the surface layer. The works of Turner et al.[34,35], Kato and Phillips[18], and Stefan and Ford[29] are representative of this group of models. The first two are supported primarily by data from laboratory experiments and achieve qualitative agreement with the observed behavior of stratification patterns in the ocean and lakes. Stefan and Ford's model has been developed specifically to simulate conditions in a lake.

The laboratory experiments conducted by Turner et al. and by Kato and Phillips involved measurements of the rate of entrainment across the interface of two superimposed layers of fluid of slightly different densities. Entrainment rates were studied as a function of density gradient and rate of mechanical energy input to one or both of the layers. The major difference between the series of

4.2.2 Mechanical Energy Balance Models (continued)

experiments by Turner et al. and by Kato and Phillips was in the mode
of mixing employed. Turner used grid stirring in one or both layers,
while Kato and Phillips introduced kinetic energy by applying a
known horizontal stress to the surface of the upper layer. The latter
method, according to the authors, is more representative of field
conditions (surface wind stress) and reduces the problem of having
to define the length and velocity scales of the turbulence induced
by grid stirring in order to apply experimental results to field
conditions. Because of this problem, in some respects, it is diffi-
cult to relate the results of the two series of experiments quantitively.

Both Turner et al. and Kato and Phillips utilized a model of the
following form to represent their data:

$$E = \frac{U_E}{U_*} = k R^n \qquad (4.2-1)$$

$$R = \frac{gZ\Delta\rho}{U_*^2 \rho} = \frac{g \, (\partial\rho/\partial Z) \, Z^2}{2\rho \, U_*^2} \qquad (4.2-2)$$

where,

    E = entrainment ratio (dimensionless)

    U_E = entrainment rate (measured) — (ℓ/t)

4.2.2 Mechanical Energy Balance Models (continued)

U_* = surface friction velocity - (ℓ/t)

ρ = density of surface layer - (m/ℓ^3)

$\Delta\rho$ = density jump across interface - (m/ℓ^3)

Z = depth of surface layer - (ℓ)

g = acceleration of gravity - (ℓ/t^2)

R = Richardson number (dimensionless)

n,k = empirical parameters (dimensionless)

As discussed above, Kato and Phillips employed the surface friction velocity and depth of the mixed surface layer as surrogates for the velocity and length scales of the induced turbulence, respectively. Turner, on the other hand, did not explicitly define these scales and utilized stirring rates (cycles/time) as a surrogate for the velocity scale. The unknown length scale and the proportionality constant between grid stirring rate and the time velocity scale are thus incorporated into the empirical parameter k, to be estimated from the experimental data. The slope, n, is, however, independent of the unknown scale factors and thus can be used as a partial basis for comparing the two sets of experimental results.

Turner et al.[35] found that the slope parameter, n, was approximately -1 in the lower range of Richardson numbers investigated and

4.2.2 Mechanical Energy Balance Models (continued)

approached -1.5 at higher Richardson numbers, where molecular dif-
fusion became important. The molecular diffusion effects were de-
duced from observations that, at high Richardson numbers, entrainment
rates were significantly lower when differences in salt concentration
were used to impose the density gradient, as compared to when tempera-
ture differences were used. This could be explained by the fact that
the molecular diffusivity of heat is on the order of one hundred times
that of salt. It is not possible to determine whether the divergence
between salt and heat transport is of consequence under field conditions
because of the scaling problem, i.e., the range of effective Richardson
numbers over which this phenomenon was observed may be higher than that
typically observed in natural systems.

The experimental results of Kato and Phillips are shown in
Figure 4.2-1. In fitting their data, the authors constrained the
slope, n , to -1 and derived a k value of 2.5, stating that the
latter is "uncertain to within about 30%". The authors noted that
a least-squares estimate of the slope would have been somewhat greater
than -1. The slope was constrained to illustrate the similarity of
the results to those of Turner et al. at low Richardson numbers and
to support the theoretical energy balance arguments discussed below.
In these experiments, no evidence was found of the -1.5 dependence
of E on R observed by Turner at high Richardson numbers. Because
of the scaling problem discussed above, it is uncertain whether this

Figure 4.2-1

Entrainment Ratio Versus Richardson Number - Experimental
Data of Kato and Phillips[18]
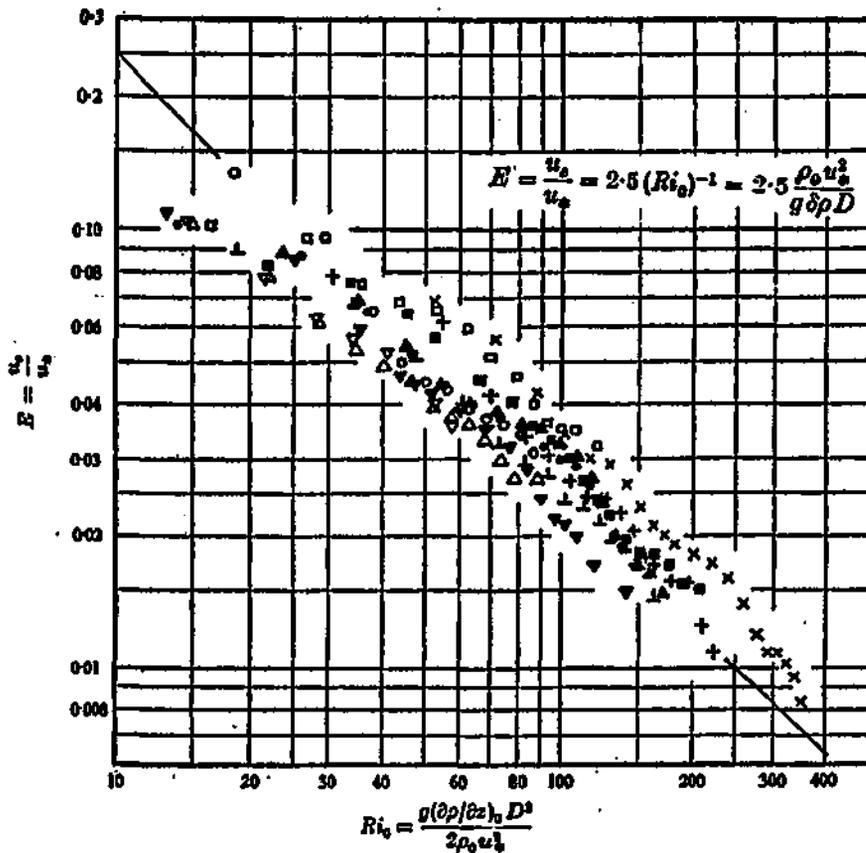
FIGURE 6. The entrainment coefficient E as a function of the overall Richardson number. The parameters represented by the symbols are indicated in the table below:

$r_0 = $	$d\rho/dz = 1.92 \times 10^{-4}$	3.84×10^{-3}	7.09×10^{-3} c.g.s.
0.995 c.g.s.	□	■	×
1.485	△	▲	
2.12	○	●	+
3.75	▽	▼	⊥

Equation shown in figure:

$$E' = \frac{u_e}{u_*} = 2\cdot 5\,(Ri_0)^{-1} = 2\cdot 5\,\frac{\rho_0 u_*^2}{g\,\delta\rho\,D}$$

$$Ri_0 = \frac{g(\partial\rho/\partial z)_0\,D^2}{2\rho_0 u_*^2}$$

4.2.2 Mechanical Energy Balance Models (continued)

constitutes a discrepancy between the two sets of experimental results.

As pointed out by Kato and Phillips, a slope of -1 in the E
versus R relationship is evidence that the rate of change of potential
energy in the system due to changes in the density stratification pattern
is proportional to the rate of kinetic energy dissipation in the
well-mixed surface layer. This is illustrated by calculations outlined
in Figure 4.2-2. For a slope of -1, the rate of change of potential
energy in the system is k/2 times the rate of kinetic energy input
to the surface layer. Thus, the k value of 2.5 derived from the
results of Kato and Phillips indicates that the rate of potential
energy increase in their system was 125% of the rate of kinetic energy
dissipation. (Their experimental system was adiabatic.) This apparent
problem was not discussed in their article, although the authors did
state that the k value was uncertain to within 30%, indicating that
the observed energy conversion efficiency was not significantly differ-
ent from 100%. However, this figure still seems higher than would be
acceptable if other kinetic energy sinks (primarily heat generated by
viscous damping) were taken into account. This suggests that the
velocity and/or length scales selected by Kato and Phillips may not
have been accurate and thus that the k value may not be appropriate
for direct application in other systems, although the slope would be
independent of such errors.

Figure 4.2-2

Mechanical Energetics of Entrainment

Initial State   Final State
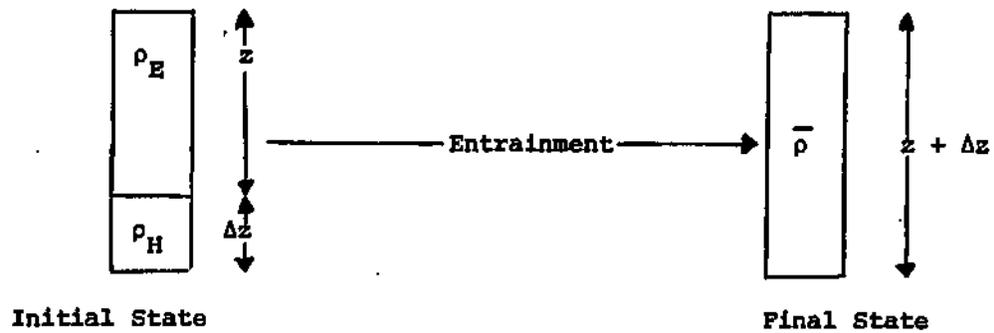
Buoyant Potential Energy Per Unit Area :

$$PE_i = g[\rho_E \, z \, (\Delta z + z/2) + \rho_H \, \overline{\Delta z}^2/2]$$

$$PE_f = g[\overline{\rho} \, \overline{(z + \Delta z)}^2 / 2]$$

Mass Balance :

$$\overline{\rho}(z + \Delta z) = \rho_E z + \rho_H \Delta z$$

Potential Energy Change With Mixing :

$$\Delta PE = PE_f - PE_i = .5 \, g \, z \, (\rho_H - \rho_E) \Delta z$$

Entrainment Rate :

$$U_e = \Delta z / \Delta t$$

$$\therefore \Delta PE/\Delta t = .5 \, g \, z \, (\rho_H - \rho_E) \, \Delta z / \Delta t = .5 \, g \, z \, (\rho_H - \rho_E) \, U_e$$

Model :

$$E = U_e/U_* = k / R = k \, U_*^2 \, \rho_E / [(\rho_H - \rho_E) \, g \, z]$$

$$\therefore \Delta PE/\Delta t = .5 \, k \, U_*^3 \, \rho_E$$

Rate of Kinetic Energy Dissipation :

$$\Delta KE/\Delta t = \rho_E \, U_*^3$$

Mechanical Energy Conversion :

$$\Delta PE/\Delta t = .5 \, k \, \Delta KE/\Delta t$$

4.2.2 Mechanical Energy Balance Models (continued)

Stefan and Ford have extended the above two-layered models to a multi-layered one and have coupled the mechanical and thermal energy balance equations to permit simulation of temperature dynamics in dimictic lakes. Figure 4.2-3 depicts the energy transformations simulated by this model. All of the mechanical energy input from the wind is assumed to be converted to turbulence. The downward movement of the lower edge of the mixed surface layer is simulated using a mechanical energy balance criterion. At each time step, the model evaluates the change in potential energy which would occur if the next depth increment (of approximate thickness 0.25 m) below the mixed surface layer were to be entrained. This is compared to the rate of kinetic energy dissipation in the surface layer. If the ratio of kinetic energy input to potential energy change is greater than one, entrainment occurs; if the ratio is less than one, the kinetic energy is assumed to be dissipated as heat and no change in the depth of the mixed layer occurs. The thermal energy fluxes are modelled primarily using formulations suggested by Dake and Harleman[8].

The model achieved general qualitative agreement with data from Lake Calhoun and Halstead Bay, Lake Minnetonka, Minnesota. In the simulation, heat energy and mechanical energy were input successively at time steps of either one day or three hours (in accordance with the availability of meteorologic data). The authors noted that the computed temperature profiles were rather sensitive to the time step

Figure 4.2-3

**Mechanical and Thermal Energy Transformations in a Lake According to
Stefan and Ford[29]**

Wind Energy

Solar and Atmospheric
Thermal Energy

Air-Water Interface

Mean Flow Kinetic Energy

(circulations, waves, seiches)

Turbulent Kinetic Energy

viscous
dissipation

Stability Criterion ————→ Internal Heat Energy

entrainment

Buoyant Potential Energy
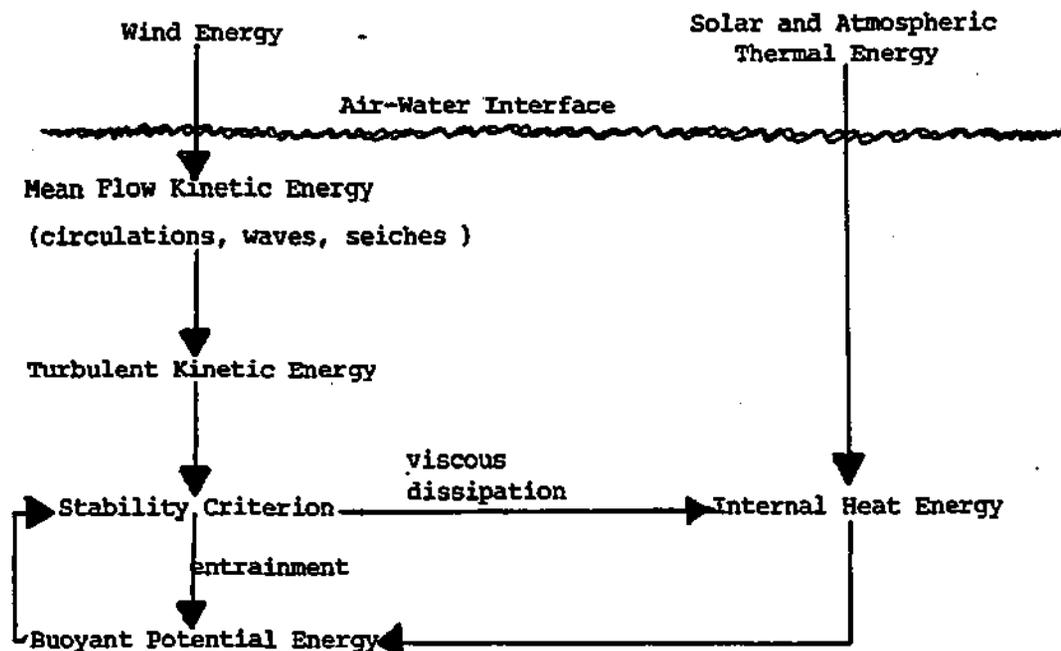
4.2.2 Mechanical Energy Balance Models (continued)

employed, as well as the assumed radiation attenuation coefficient.
The latter conclusion was also reached by Parker et al. in their
evaluation of the MIT deep reservoir model, discussed in Section
4.2.1. Since the attenuation coefficient is a partial function of
algal density, Stefan and Ford suggested that this may be an important
means by which the biota may influence the temperature distributions
and vertical mixing rates in lakes and reservoirs. Light-limited
algal populations would benefit from such effects.

4.2.3 Statistical Studies

A statistical study on vertical mixing data from several
morphologically-different lakes and oceanic areas in the northern
temperate zone was performed by Blanton[6]. The measure of mixing
rate employed was the mean rate of entrainment into the epilimnion,
as determined by the average rate of migration of the thermocline
during the stratified period. The study excluded periods of rapid
heating in the spring and convective mixing in the autumn. Entrain-
ment was correlated with the mean stability across the thermocline,
as computed from the temperature profiles at the beginning and end
of the study period for each lake. The measure of stability employed
was the mean density gradient. Because the mean stability was cal-
culated using only data from the beginning and end of the stratified

4.2.3 Statistical Studies (continued)

period, the estimates may be somewhat biased. Stability tends to go
through a maximum during the stratified period, so that estimates
of mean stability calculated from data at the ends could tend to be
too low.

Blanton presented his data as a plot of stability versus entrain-
ment rate and summarized it with the following relationship:

$$X = a Y^b \qquad\qquad (4.2\text{-}3)$$
$$a = 8.95 \times 10^{-7}$$
$$b = -0.521$$
$$r = -0.84$$

where,

$$Y = \frac{1}{A_0} \frac{\Delta V_E}{\Delta t} = \text{entrainment rate (m/sec)}$$

$$X = \frac{g}{\rho} \frac{\Delta \rho}{\Delta z} = \text{mean stability (sec}^{-2})$$

The statistical basis for the above relationship was not stated in
Blanton's article. In the above form, the dependent variable is
stability and the independent variable is entrainment rate. From
a cause-effect standpoint, it would seem more logical to reverse

4.2.3 Statistical Studies (continued)

these roles. Accordingly, Blanton's data have been replotted in
Figure 4.2-4 as entrainment rate versus stability on logarithmic
scale. The least squares relationship derived from linear regression
on log transformed data is:

$$Y = a X^b \qquad\qquad (4.2-4)$$

$$a = 1.71 \times 10^{-11}$$

$$b = -1.63$$

$$r = -0.91$$

The correlation coefficient, -0.91, is somewhat higher in absolute
value than that computed by Blanton, -0.84, suggesting that Blanton
may have computed his coefficient on a linear scale.

In relating entrainment rate to stability, equation (4) is
analogous to the relationship of E versus R [equation (1)], used
to represent the experimental data of Kato and Phillips and of Turner
et al.[34,35]. The 95% confidence interval for the slope of the Y
versus X relationship is -1.63 ± 0.34, somewhat higher in absolute
value, but not significantly different from the range of slopes in
the E versus R relationship observed in the laboratory experiments,
-1.0 to -1.5.

Figure 4.2-4
Plot of Blanton's [6] Data on Entrainment Rate vs.
Stability for Temperate Lakes
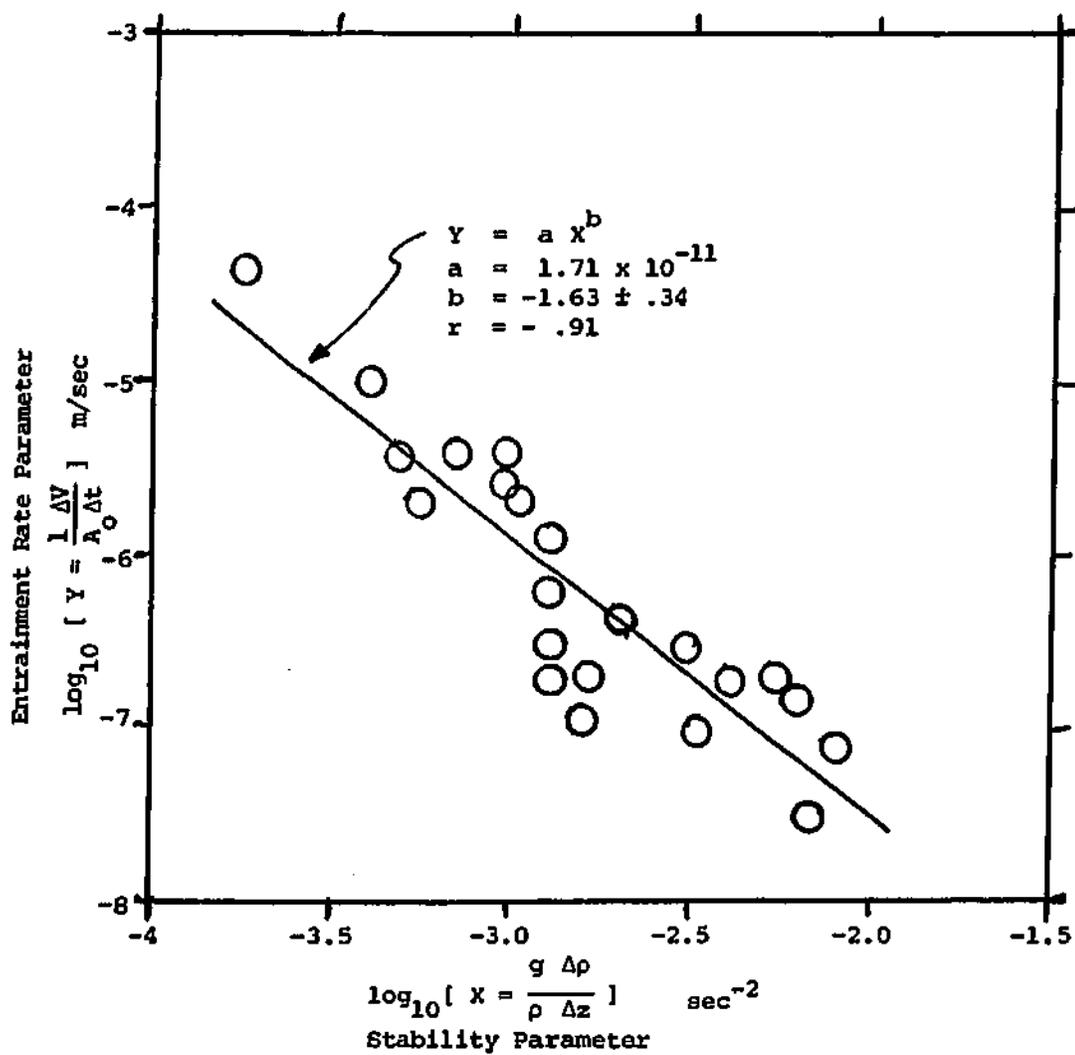
4.2.3 Statistical Studies (continued)

There are two factors which might have contributed to the
relative steepness of the slope in Blanton's data. First, one would
expect a negative correlation between wind velocities and mean
stability. Had the Richardson number been used as a measure of
stability, the slope would have been flattened. Blanton noted that
it would have been preferable to have expressed stability in the
form of a Richardson number. However, the necessary wind data were
not available. The second factor which might have contributed is
the possible bias in Blanton's estimate of mean stability. If one
assumes that the peaking in stability during the stratified period
would tend to be more pronounced in lakes with higher stability,
elimination of this bias in the mean stability estimate would also
tend to flatten the slope. Even without these considerations,
Blanton's data from a wide range of lakes and oceanic regions are
not inconsistent with the experimental results discussed previously.

Blanton also noted a strong correlation between entrainment
rate and mean depth (Figure 4.2-5). Snodgrass[28] summarized Blanton's
data with the following regression equation:

$$Y = 8.65 \times 10^{-9} \, \bar{z}^{1.08} \qquad (4.2-5)$$

$$r = 0.887$$

Figure 4.2-5

Blanton's [6] Data on Entrainment Rate Verses Mean Depth
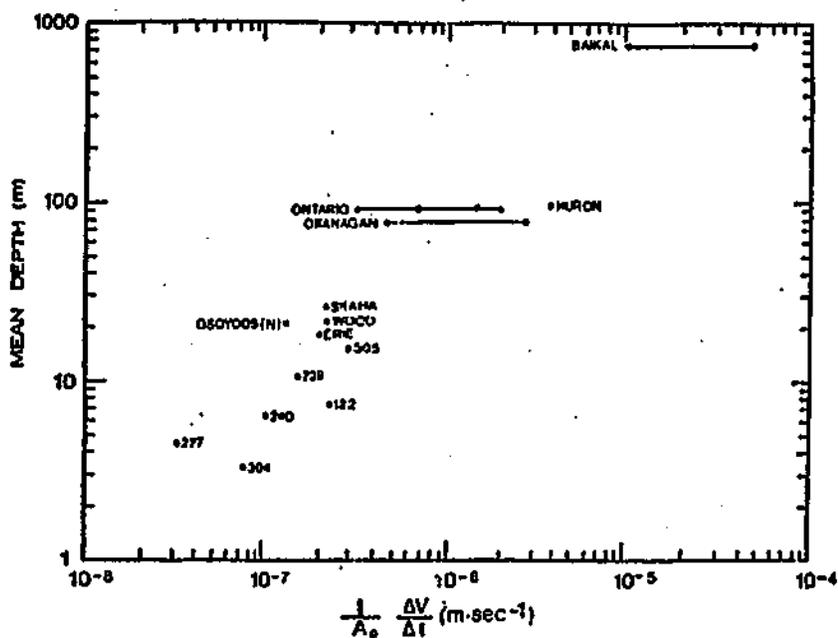
Figure 4.2-6

Snodgrass's [28] Data on Vertical Transport Coefficient
Verses Mean Depth
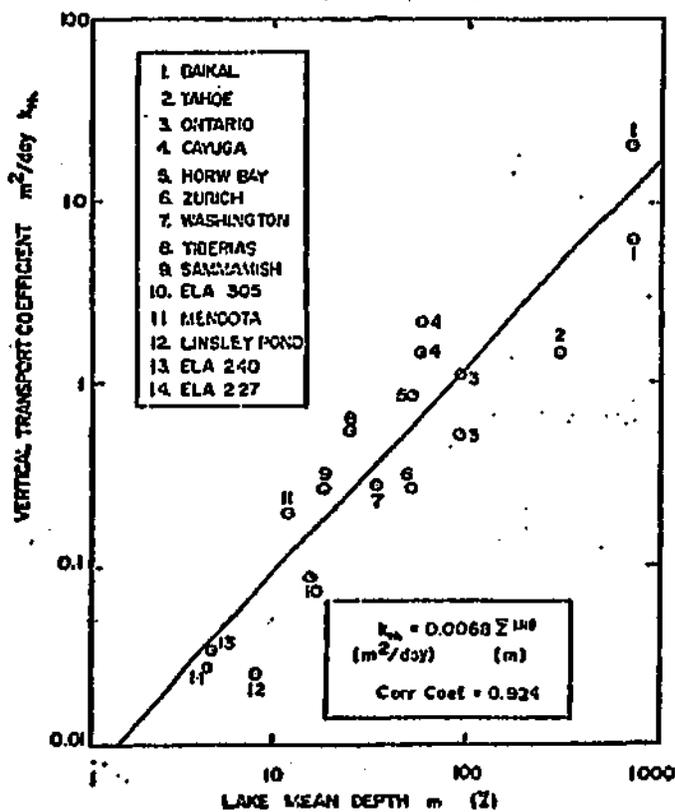
4.2.3 Statistical Studies (continued)

where,

$$Y = \frac{1}{A_o} \frac{\Delta V_E}{\Delta t} = \text{entrainment rate (m/sec)}$$

$$\bar{Z} = \text{lake mean depth (m)}$$

Snodgrass, using a different set of lakes, also found that the relationship between the vertical transport coefficient across the thermocline, expressed as a diffusivity, and mean depth (Figure 4.2-6) could be summarized by :

$$K_{TH} = 0.00682 \; \bar{Z}^{1.12} \qquad\qquad (4.2-6)$$

$$r = 0.924$$

where,

$$K_{TH} = \text{vertical transport coefficient (m}^2\text{/day)}$$

The apparent influence of mean depth on vertical transport rate, expressed either as entrainment or as diffusivity, was explained by Blanton with reference to the mechanical energy theory of Kato and Phillips. Blanton's hypothesis was that deeper lakes have

4.2.3 Statistical Studies (continued)

larger basins, greater available wind fetches, and, hence, greater kinetic energy input rates per unit area. This explanation is examined in more detail below.

In each set of data, the mixing rate variable (Y or K_{TH}) had a range of three orders of magnitude. A similar range of kinetic energy input rate per unit area would be necessary in order for the mechanical energy theory to account for this variation. The rate of kinetic energy input varies as the cube of the surface friction velocity (Figure 4.2-2). Friction velocity is proportional to wind velocity to the 1.25th power, over a moderate range of wind velocities, 1 - 15 m/sec[44]. Hence, kinetic energy input rate varies roughly as the 3.75th power of wind velocity, and about a 6.3-fold range in wind velocity would be required to explain the thousand-fold range in mixing rate. This range may be reasonable, although specific data on wind speed as a function of lake mean depth would be necessary in order to substantiate Blanton's argument. One factor ignored by Blanton is circulation induced by Coriolis forces, which would also tend to be more important in larger lakes[15]. This would serve as an additional source of kinetic energy and would thus reduce the 6.3-fold variation in wind velocities requires in order for the observed variation in mixing rates to be explained by variation in kinetic energy input rates.

4.2.3 Statistical Studies (continued)

Another factor which might contribute to the observed corre-
lation between mixing rate and depth is the apparent sensitivity of
temperature profiles to radiation extinction coefficients, noted in
modelling studies by Parker et al. and by Stefan and Ford. In many
lakes, the rate of attenuation of radiation with depth is largely
a function of algal density[20]. Vollenweider[40] has demonstrated the
negative correlation between lake trophic state and mean depth. These
facts suggest that lower algal densities and lower extinction coeffi-
cients would be typical of lakes with greater mean depths. A lower
extinction coefficient would permit a more even distribution of
radiation absorption with depth, giving rise to a more uniform
temperature distribution and greater mixing rate. Thus, the apparent
correlation between mixing rate and mean depth may be explained, in
part, by biologic, as well as energetic effects.

4.3 Model Development

Essential aspects of the model which has been developed for
simulating vertical mixing in Onondaga Lake are summarized in
Figures 4.3-1 and 4.3-2 and Tables 4.3-1 to 4.3-4. Figure 4.3-1
defines the system. The water, heat, and chloride fluxes corres-
ponding to the various streams in Figure 4.3-1 are given in Table
4.3-1. "Forcing functions", which drive the model, are given in

Figure 4.3-1

System Diagram for the Onondaga Lake Vertical Stratification Model
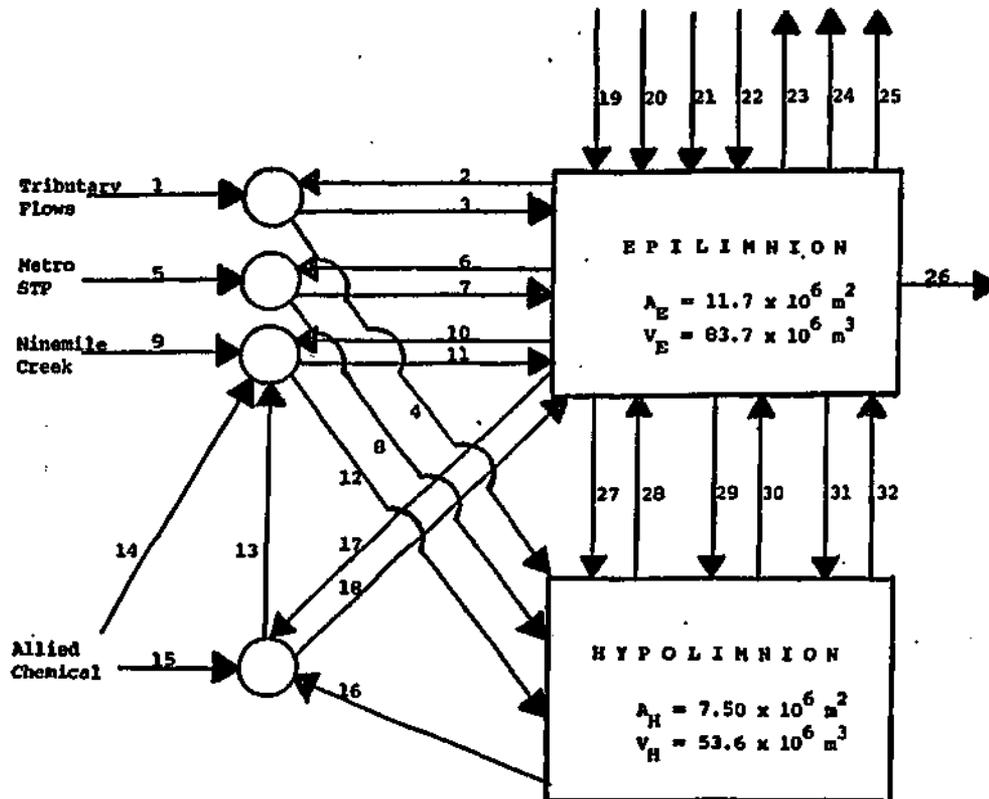
$$V_E \frac{dT_E}{dt} = -H_2 + H_3 - H_6 + H_7 - H_{10} + H_{11} - H_{17} + H_{18} + H_{19} + H_{20} + H_{21}$$
$$+ H_{22} - H_{23} - H_{24} - H_{25} - H_{26} - H_{27} + H_{28} - H_{29} + H_{30} - H_{31} + H_{32}$$

$$V_E \frac{dC_E}{dt} = -M_2 + M_3 - M_6 + M_7 - M_{10} + M_{11} - M_{17} + M_{18} + M_{19} + M_{20} + M_{21}$$
$$+ M_{22} - M_{23} - M_{24} - M_{25} - M_{26} - M_{27} + M_{28} - M_{29} + M_{20} - M_{31} + M_{32}$$

$$V_H \frac{dT_H}{dt} = -H_{16} + H_{12} + H_8 + H_4 + H_{27} - H_{28} + H_{29} - H_{30} + H_{31} - H_{32}$$

$$V_H \frac{dC_H}{dt} = -M_{16} + M_{12} + M_8 + M_4 + M_{27} - M_{28} + M_{29} - M_{30} + M_{31} - M_{32}$$

H_i = heat flux in stream i (KKcal/day)

M_i = chloride flux in stream i (g/day)

T_E, T_H = epilimnion and hypolimnion temperatures ($^\circ$C)

C_E, C_H = epilimnion and hypolimnion chloride concentrations (g/m^3)

Figure 4.3-2

Control Pathways in the Onondaga Lake Vertical Stratification Model
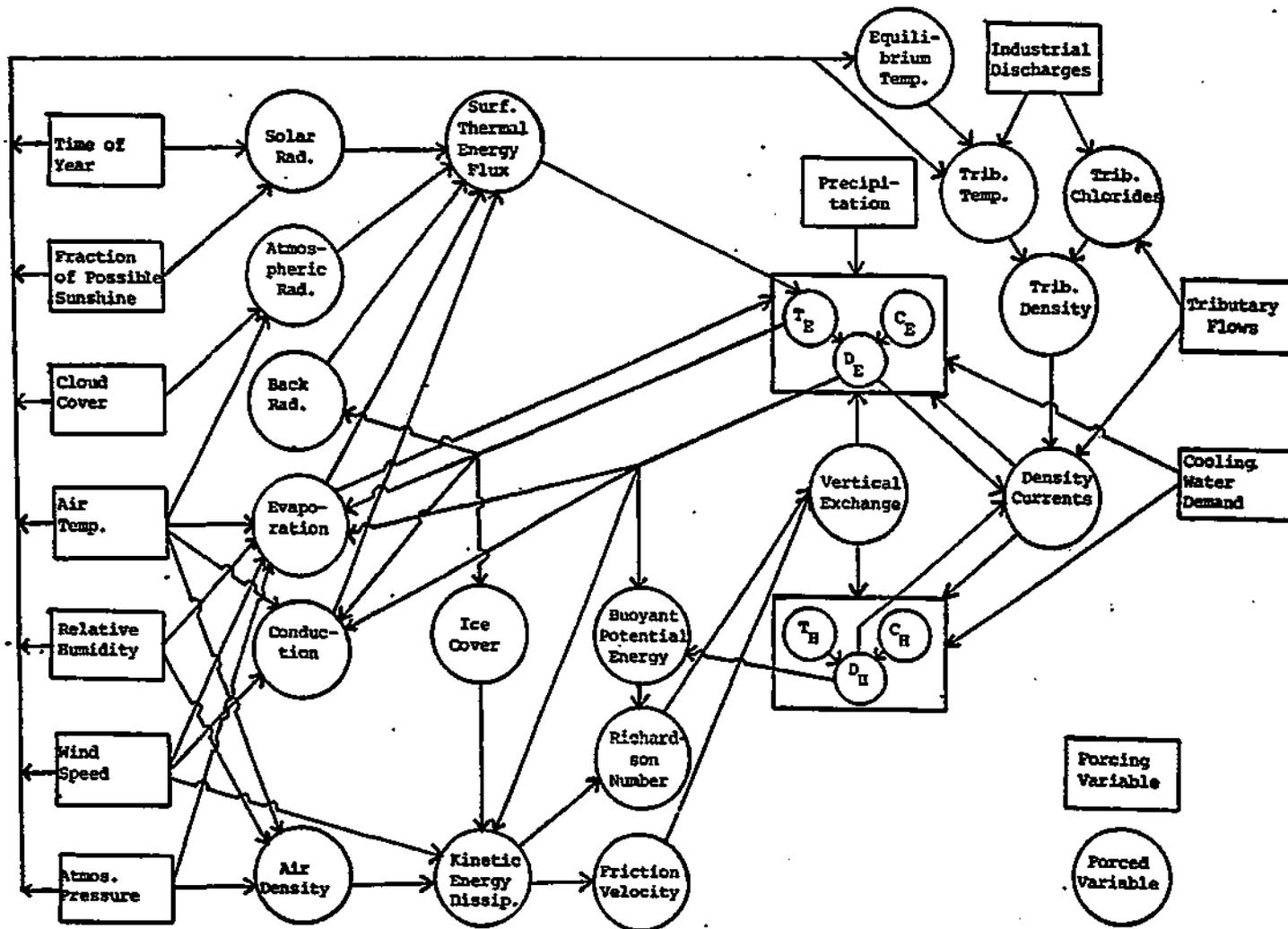
Table 4.3-1

Water, Heat, and Chloride Fluxes

[a] Description		Q_i = Flow (m^3/day)	E_i = Heat flux (KKcal/day)[b]	W_i = Chloride flux (g/day)
1	Tributary Inflow	Q_1	$Q_1 U_{20}$	$Q_1 U_{23}$
2	" Diluent	$Q_1 a_1$	$Q_1 a_1 T_E$	$Q_1 a_1 C_E$
3	" Discharge to Epil.	$Q_1(1+a_1)(1-r_1)$	$Q_1(U_{20}+a_1 T_E)(1-r_1)$	$Q_1(U_{23}+a_1 C_E)(1-r_1)$
4	" Discharge to Hypo.	$Q_1(1+a_1)r_1$	$Q_1(U_{20}+a_1 T_E)r_1$	$Q_1(U_{23}+a_1 C_E)r_1$
5	Metro STP Inflow	Q_2	$Q_2 U_{21}$	$Q_2 U_{24}$
6	" Diluent	$Q_2 a_2$	$Q_2 a_2 T_E$	$Q_2 a_2 C_E$
7	" Discharge to Epil.	$Q_2(1+a_2)(1-r_2)$	$Q_2(U_{21}+a_2 T_E)(1-r_2)$	$Q_2(U_{24}+a_2 C_E)(1-r_2)$
8	" Discharge to Hypo.	$Q_2(1+a_2)r_2$	$Q_2(U_{21}+a_2 T_E)r_2$	$Q_2(U_{24}+a_2 C_E)r_2$
9	Ninemile Inflow	Q_3	$Q_3 U_{22}$	$Q_3 U_{25}$
10	" Diluent	$Q_3 a_3$	$Q_3 a_3 T_E$	$Q_3 a_3 C_E$
11	" Discharge to Epil.	$Q_3(1+a_3)+Q_{13}+Q_{16}(1-r_3)$	$Q_3(U_{22}+a_3 T_E)+H_{13}+H_{16}(1-r_3)$	$Q_3(U_{25}+a_3 C_E)+W_{13}+W_{16}(1-r_3)$
12	" Discharge to Hypo.	$Q_3(1+a_3)+Q_{13}+Q_{16}r_3$	$Q_3(U_{22}+a_3 T_E)+H_{13}+H_{16}r_3$	$Q_3(U_{25}+a_3 C_E)+W_{13}+W_{16}r_3$
13	Allied Heat Plume Discharge	$a_4(Q_{15}+Q_{16}+Q_{17})$	$a_4(H_{15}+H_{16}+H_{17})$	$a_4(W_{15}+W_{16}+W_{17})$
14	" Waste Bed Overflow	0.	$Q_5(Q_{16}+Q_{17})$	U_6
15	" Thermal Load	0.	0.	U_7
16	" Hypo. Intake	$(1-Q_4)F_4$	$(1-Q_4)F_4 T_H$	$(1-Q_4)F_4 C_H$
17	" Epil. Intake	$Q_4 F_4$	$Q_4 F_4 T_E$	$Q_4 F_4 C_E$
18	" East Flume Discharge	$(1-a_4)(Q_{15}+Q_{16}+Q_{17})$	$(1-a_4)(H_{15}+H_{16}+H_{17})$	$(1-a_4)(W_{15}+W_{16}+W_{17})$
19	Precipitation	$Q_8 A_E$	$Q_8 A_E U_{11}$	$Q_{26} C_E$
20	Solar Radiation		$U_9 A_E$	0.
21	Atmospheric Radiation		$e_B(U_{11}+273.1)^6(1+a_9 U_{10}^2)A_E$	0.
22	Conduction in		$a_5 b_5 U_{11} U_{13} U_8 D_E$	0.
23	Conduction out		$a_5 b_5 E_E U_{13} D_E$	0.
24	Surface Emission		$a_7(T_E+273.1)^4 A_E$	0.
25	Evaporation	$a_5 U_{13} F_E$	$a_5 U_{13}(a_{10} T_E+a_{11} T_E+a_{12})A_E b_E$	0.
26	Outflow	$Q_{19}-Q_{25}-Q_9+Q_5+Q_1$	$Q_{26} T_E$	$Q_{26} C_E$
27	Turbulent Exchange (E→H)	$F_{11} A_H$	$F_{11} A_H T_E$	$F_{11} A_H C_E$
28	" (H→E)	$F_{11} A_H$	$F_{11} A_H T_H$	$F_{11} A_H C_H$
29	Molecular Diffusion (E→H)	0.	$a_{13} A_H T_E$	$a_{13} A_H C_E$
30	" (H→E)	0.	$a_{13} A_H T_H$	$a_{13} A_H C_H$
31	Advective Flow (E→H)	F_9	$F_9 T_E$	$F_9 C_E$
32	" (H→E)	F_{10}	$F_{10} T_H$	$F_{10} C_H$

a - for stream definitions, see Figure 4.3-1
b - assuming a volumetric heat capacity of 1 KKcal/m^3-°C

4.3 Model Development (continued)

Table 4.3-2. "System functions", which depend upon values of the state variables, are given in Table 4.3-3. Parameter identifications and values are given in Table 4.3-4. Figure 4.3-2 summarizes the important control pathways. Essential features are discussed in detail below.

4.3.1 System Definition

The model represents the lake as two, completely-mixed compartments of constant volume, corresponding roughly to the epilimnion and the hypolimnion of the lake. Based upon the analysis of density profile data in Chapter 3, the boundary between the two compartments has been set at the average thermocline depth of 9 meters. Because of thermocline migration with season, the compartments do not exactly correspond to the epilimnion and hypolimnion. It was hoped that this highly-aggregated representation of the lake would be adequate for analysis of the outfall design issue, the primary emphasis of which is on potential impacts of the design upon general mixing in the lake. The test of the adequacy of this representation is in its ability to simulate observed temperature and chloride variations with time.

A total of four state variables are integrated in the model:

4.3.1 System Definition (continued)

$$T_E = \text{epilimnion temperature } (^{\circ}C)$$

$$C_E = \text{epilimnion chloride } (g/m^3)$$

$$T_H = \text{hypolimnion temperature } (^{\circ}C)$$

$$C_H = \text{hypolimnion chloride } (g/m^3)$$

Density, an important factor governing the mixing process, is computed as a function of temperature and chloride levels according to the following equations:

$$D = 1 + b_1 XC + b_2 X + b_3 C \qquad (4.3\text{-}1)$$

$$X = 16 - 2 T_{max} T + T^2 \qquad (4.3\text{-}2)$$

$$T_{max} = 4 - 0.211 S \qquad (4.3\text{-}3)$$

$$S = 0.0021 C \qquad (4.3\text{-}4)$$

where,

$$D = \text{water density } (g/cm^3)$$

$$X = \text{function of } T \ (^{\circ}C)^2$$

$$T = \text{temperature } (^{\circ}C)$$

$$C = \text{chloride concentration } (g/m^3)$$

$$T_{max} = \text{temperature of maximum density } (^{\circ}C)$$

4.3.1 System Definition (continued)

$$S = \text{salinity} = \text{total dissolved solids } (kg/m^3)$$

$$b_1 = \text{regression parameter estimate} = 7.87 \times 10^{-11}$$

$$b_2 = \text{regression parameter estimate} = -6.78 \times 10^{-6}$$

$$b_3 = \text{regression parameter estimate} = 1.70 \times 10^{-6}$$

The parameter estimates b_1, b_2, and b_3 have been derived from a linear regression analysis of data on density as a function of temperature and salinity given in Williams[43] ($R^2 = 0.999$)*. The linear model [equation (1)] is designed so that D is maximum at $T = T_{max}$. The relationship between T_{max} and S [equation (3)] is also derived from data in Williams. The relationship between salinity and chloride concentration is typical of Onondaga Lake waters, according to the Onondaga Lake Study[23]. Density differences in the fourth or fifth decimal place can be important in the model calculations. In order to minimize possible effects of round-off errors, actual computations are done using the following linear transformation commonly employed in oceanographic work[43]:

$$D' = 1000 \ (D - 1) \tag{4.3-5}$$

---

* SEE $= 5.4 \times 10^{-8}$, $0 < T < 25$, $0 < S < 10$

4.3.1 System Definition (continued)

As shown in Figure 4.3-1, inflows to the lake have been
aggregated into four sources: Allied Chemical (streams 14 and 15);
Ninemile Creek (stream 9); Metro STP (stream 5); and "Tributary"
flows (stream 1). The latter represents the sum of Onondaga Creek,
Ley Creek, Harbor Brook, Bloody Brook, and local drainage. Stream 14
from Allied Chemical represents the waste bed overflow to Ninemile
Creek, which contains most of the salinity entering the lake. When
the new sewage treatment plant is put into operation, stream 14
will be diverted and mixed with stream 5. Streams 16 and 17 repre-
sent cooling water withdrawals from the hypolimnion and epilimnion,
respectively. Stream 15 is primarily waste heat from Allied Chemical.
This is discharged in streams 13 ("West Plume") and 18 ("East Plume")
to Ninemile Creek and to the lake epilimnion, respectively.

The nodes associated with each inlet stream in Figure 4.3-1
represent entrance mixing zones, in which the entering streams (1,
5, 9) are mixed with diluent epilimnion waters (2, 6, 10). The
mixtures subsequently enter the epilimnion (3, 7, 11) or the hypo-
limnion (4, 8, 12), according to whether their densities are less
or greater than the estimated density at the thermocline, according
to functions described in Section 4.3.3.

The lake surface streams (19 - 25) consist of precipitation (19)
and five thermal energy fluxes, which depend upon meteorologic

4.3.1 System Definition (continued)

conditions and upon lake surface temperature. Most of the functional

forms and parameter values for the latter have been taken from

Harleman and Markofsky[11].

Streams 27 - 32 represent three basic types of internal exchanges

between the epilimnion and hypolimnion. Bulk, density-dependent,

turbulent transport is represented by streams 27 and 28. Mechanisms

involved include eddy diffusivity, seiches, and thermocline erosion.

Streams 29 and 30 represent molecular diffusion across the thermocline.

These are generally on a much small scale than 27 and 28, but have been

distinguished because of their density-independence. The last pair

of streams, 31 and 32, represent advective flows, which are required

in order to satisfy constant volume constraints placed upon the

hypolimnion. Finally stream 26 represents lake outflow, determined

by the epilimnic water balance.

4.3.2 Forcing Functions

Forcing functions are defined as factors which influence, but

are not influenced by the lake system. The functions in Table 4.3-2

constitute the boundary conditions which drive the system and are

comprised chiefly of hydrologic, meteorologic, and industrial variables.

All of these variables have been estimated on monthly-average time

Table 4.3-2

Forcing Function Definitions

Function		Value	Units	Ref
U_1	Tributary Inflow	a	m^3/day	c
U_2	Metro STP Inflow	a	m^3/day	c
U_3	Ninemile Creek Inflow	a	m^3/day	c
U_4	Fraction of Allied Chem Intake from Epilimnion	b	-	7
U_5	Temperature Rise Through AC Power Plant	20	degrees C	7
U_6	Allied Waste Bed Overflow Chloride Load	1.70×10^9	g/day	7,21
U_7	AC East/West Plume Chloride Load	1.23×10^8	g/day	21
U_8	Precipitation	a	m/day	38
U_9	Solar Radiation	b	$KKCal/m^2$-day	b
U_{10}	Cloud Cover	a	-	38
U_{11}	Air Temperature	a	degrees C	38
U_{12}	Fraction of Possible Sunshine	a	-	38
U_{13}	Wind Velocity	a	m/sec	38
U_{14}	Saturation Vapor Pressure at Air Temperature	b	mm Hg	b
U_{15}	Relative Humidity	b	-	b
U_{16}	Dew Point	a	degrees C	38
U_{17}	Atmospheric Pressure	a	mm Hg	38
U_{18}	Equilibrium Water Temperature	b	degrees C	b
U_{19}	Air Density	b	g/cm^3	b
U_{20}	Tributary Temperature	b	degrees C	b
U_{21}	Metro STP Temperature	b	degrees C	b
U_{22}	Ninemile Creek Temperature	b	degrees C	b
U_{23}	Tributary Chlorides	b	g/m^3	b
U_{24}	Metro STP Chlorides	b	g/m^3	b
U_{25}	Ninemile Creek Chlorides	b	g/m^3	b

a - input on a monthly-average basis

b - defined below

c - calculated from hydrologic budget developed in Chapter 3

Table 4.3-2 (continued)

Function Evaluation		Ref.

U_4

Dates	U_4	Dates	U_4	Dates	U_4	7
01/01/68	0.	06/29/71	0.	06/05/73	0.	
↓		↓		↓		
08/06/69	.78	10/14/71	.78	01/21/74	.78	
↓		↓		↓		
12/23/69	.78	07/25/72	.78	06/18/74	0.	
↓		↓		↓		
03/10/70	0.	10/24/72	0.	01/01/75		
↓	.78	↓	.78			

U_9

$$U_9 = 1.70 + 1.21 \cos\theta + 2.82\, U_{12} + .985\, U_{12} \cos\theta$$ 38

$$\theta = 2\pi(t-172)/365$$

$t = $ day of year

[linear regression analysis of monthly-average solar radiation and sunshine data from East Lansing, Mich.; Cleveland, Ohio; Ithaca, N.Y.; Boston, Ma.; Portland, Me.. (N = 283, R = .985, SEE = .260)]

U_{14}

$$U_{14} = \exp[\,20.59 - 5199./(U_{11}+273.1)\,]$$ 10

U_{15}

$$U_{15} = [(112 - .1\, U_{11} + U_{16})/(112 + .9\, U_{11})]^8$$ 19

U_{18}

solution of following transcendental energy balance equation :   39

$$(.211+.042U_{13})(U_{18}-U_{11}) - U_9$$
$$+ 2.59\times10^{-4}(1075-.97U_{18})(1+.224U_{13})[p(U_{18})-U_{15}p(U_{11})] = 0$$

$$p(U_i) = \exp[20.59 - 5199/(U_i + 273.1)]$$ 10

U_{19}

$$U_{19} = .00129[273.1/(U_{11}+273.1)][(U_{17} - .378U_{15}U_{14})/760]$$ 10

		N	R^2	SEE
U_{20}	$U_{20} = 6.14 + 4.19 \cos\theta + .33\, U_{18,0}$	91	.914	1.91
U_{21}	$U_{21} = 11.8 + 2.77 \cos\theta + .23\, U_{18,1}$	91	.815	2.09
U_{22}	$U_{22} = U_{20}$	–	–	–
U_{23}	$\ln(U_{23}) = 6.601 - .621 \ln(U_1/86400)$	115	.434	.48
U_{24}	$\ln(U_{24}) = 5.989 + .140 \ln(U_2/86400)$	127	.003	.41
U_{25}	$U_{25} = U_{23}$	–	–	–

[regression analysis of survey data]   21

$U_{18,i} = $ equilibrium temperature, lag i months

$\theta = 2\pi\,(t-220)/365$; $t = $ day of year

4.3.2 Forcing Functions (continued)

scales for the period of interest (1968-74). They are plotted against time in Appendix E.

Tributary flows U_1, U_2, and U_3 are derived from the hydrologic balance developed in Chapter 3. Tributary temperatures, U_{20}, U_{21}, and U_{22}, are estimated based upon regression models which compute temperature from day-of-year and equilibrium temperature. The latter, U_{18}, is defined as the water temperature which is in equilibrium with ambient meteorologic conditions. It is computed according to a formulation suggested by Velz[39]. The temperature regression models for stream 1 and the Metro STP have been estimated based upon data from the O'Brien and Gere surveys[21]. Chloride levels for these streams, U_{23} and U_{24}, are computed as functions of flow, using regression models also developed from O'Brien and Gere's data. Ninemine Creek temperatures and chloride levels upstream of the Allied Chemical discharges are assumed to be equal to those of stream 1.

A 20°C temperature rise through the Allied Chemical power plant has been assumed, based upon analysis of survey data[21] and personal communications with the Allied Chemical engineering staff[7]. The chloride fluxes from the solvay plant, U_6 and U_7, have also been estimated from survey data[21] on Ninemile Creek and East Flume, respectively. The waste bed overflow component, U_6, has been calculated

4.3.2 Forcing Functions (continued)

from the Ninemile Creek flow and chloride concentration data, with
adjustment for upstream chloride levels. Both the magnitude and
the temporal stability of this flux have been verified by soda ash
production data provided by Allied Chemical[7].

U_4 , the fraction of the Allied Chemical cooling water taken
from the epilimnion, has been estimated as a function of time from
information provided by Clough[7]. The cooling water intake is roughly
regulated to minimize pumping and chlorine demand costs. Hypolimnic
waters are cooler, have more heat capacity, and therefore require
somewhat less flow to dispose of a given quantity of heat. Allied
chlorinates the cooling water to oxidize hydrogen sulfide and minimize
corrosion damage to its equipment. Normally, the plant uses hypolimnic
waters exclusively. When hydrogen sulfide levels become excessive,
an epilimnion intake is opened, to reduce chlorine demand costs.

The remaining forcing functions in Table 4.3-2 are meteorologic
variables, some of which are computed as functions of others. Solar
radiation measurements (U_9) were not available for Syracuse, so
U.S. Weather Service data[38] from stations at roughly the same latitude
were employed to estimate a regression model for radiation as a function
of time-of-year and percent of possible sunshine. The saturation vapor
pressure of air (U_{14}) , relative humidity (U_{15}) , and air density
(U_{19}) are all computed from other reported meteorologic variables,

4.3.2 Forcing Functions (continued)

according to the functions given in Table 4.3-2.

4.3.3 System Functions

System functions depend upon the values of the state variables, and comprise the primary relationships within the model. They are defined and described in Table 4.3-3.

The first three functions represent the fractions of the inflowing waters which sink into the hypolimnion due to density differences. As described above, each stream enters the lake and is initially diluted with epilimnic waters. The density of the diluted stream is compared with the density of the lake at the thermocline, estimated as the average of the computed epilimnion and hypolimnion densities. If the influent stream density is less than the interface density, it is assumed to stay in the epilimnion and the corresponding function value is set to zero. If the influent density is greater, a certain fraction of the flow is assumed to sink into the hypolimnion. The exponential functions employed to compute this fraction are merely used to smooth the transition from zero to the respective maximum values, a_{20}, a_{21}, and a_{22}. This is done to prevent instability in the numerical integration scheme. The value of a_{19} is sufficiently small that the transition occurs

Table 4.3-3

System Function Definitions

Function		Units
F_1	Fraction of Tributary Flow Entering Hypolimnion	-
F_2	Fraction of Metro STP Flow Entering Hypolimnion	-
F_3	Fraction of Ninemile Creek Flow Entering Hypolimnion	-
F_4	Allied Chemical Total Cooling Water Flow	m^3/day
F_5	Average Inlet Temperature in AC Cooling Water	degrees C
F_6	Evaporation Driving Force	mm Hg
F_7	Partial Pressure of Water at Surface Temperature	mm Hg
F_8	Hypolimnic Flow Surplus	m^3/day
F_9	Advective Flow [Epilimnion → Hypolimnion]	m^3/day
F_{10}	Advective Flow [Hypolimnion → Epilimnion]	m^3/day
F_{11}	Vertical Exchange Rate	m/day
F_{12}	Richardson Number	-
F_{13}	Friction Velocity	m/sec
F_{14}	Ice Cover Function	-

Table 4.3-3 (continued)

Function	Ref

$F_1 = 0.$, $D_3 \leq D_I = (D_E + D_H)/2.$

$\quad = a_{20} \exp[a_{19}/(D_I - D_3)] \simeq a_{20}$, $D_3 > D_I$

$F_2 = 0.$, $D_7 \leq D_I$

$\quad = a_{21} \exp[a_{19}/(D_I - D_7)] \simeq a_{21}$, $D_7 > D_I$

$F_3 = 0.$, $D_{11} \leq D_I$

$\quad = a_{22} \exp[a_{19}/(D_I - D_{11})] \simeq a_{22}$, $D_{11} > D_I$

$F_4 = a_{24} + a_{25} F_5$ 7

$F_5 = U_4 T_E + (1 - U_4) T_H$

$F_6 = F_7 - U_{15} U_{14}$ 11

$F_7 = \exp[20.59 - 5199/(T_E + 273.1)]$ 10

$F_8 = Q_4 + Q_8 + Q_{12} - Q_{16}$ [Hypolimnic Water Balance]

$F_9 = 0.$, $F_8 \geq 0.$

$\quad = -F_8$, $F_8 < 0.$

$F_{10} = F_8$, $F_8 > 0.$

$\quad = 0.$, $F_8 \leq 0.$

$F_{11} = 86400 \, a_{14} \, F_{13}/[F_{12}^{a_{15}} + (a_{14}/a_{16})]$, $F_{12} > 0.$ 18

$\quad = 86400 \, a_{16} \, F_{13}$, $F_{12} \leq 0.$

$F_{12} = 9.8 \, a_{17} \, (D_H - D_E)/(F_{13}^2 \, D_E)$ 18

$F_{13} = .0224 \, U_{13}^{1.25} \, F_{14} \, (U_{19}/D_E)^{.5}$ 4,44

$F_{14} = 1.$, $T_E \geq 0.$

$\quad = a_{23} + (1 - a_{23}) \exp(3 \, T_E) \simeq a_{23}$, $T_E < 0.$

4.3.3 System Functions (continued)

rapidly. Examination of the lake and tributary density data revealed that the major tributary stream (1) and the Metro STP stream (5) were always less dense than the epilimnic waters, due to the salinity of the latter. Thus, under past and present conditions, these streams would not be expected to sink, and the corresponding parameter values a_{20} and a_{21} have been set to zero. Likewise, in these cases, initial dilution would be of no consequence, so the corresponding dilution ratios a_1 and a_2 have also been set to zero. Survey data indicated that Ninemile Creek was considerably more dense than the epilimnion and the hypolimnion of the lake during most seasons. Accordingly, the density current phenomenon would be important in this case. Estimation of the parameters for this stream will be discussed in a subsequent section.

The rate of cooling water withdrawal by Allied Chemical, F_4, is assumed to be a weak function of the average inlet temperature, F_5. The linear relationship has been estimated from monthly-average flow data provided by Allied Chemical[7] and corresponding lake temperature data[21]. Flows decrease with decreasing inlet temperature because of increased heat disposal capacity per unit of flow.

Evaporation from the water surface is governed in part by F_6, the evaporation driving force, which depends upon F_7, the partial pressure of water vapor at the lake surface temperature and upon the

4.3.3 System Functions (continued)

partial pressure of water vapor in the air, computed from air temperature and relative humidity.

The hypolimnic flow surplus, F_8 , is required to compute the advective flows to and from the hypolimnion, F_9 and F_{10} , respectively. These are necessary to satisfy the constant volume constraint placed upon the hypolimnion.

The primary vertical exchange function in the model is F_{11} , which is computed from the Richardson number (F_{12}) and the friction velocity (F_{13}). This functional form was suggested in part by the experimental work of Kato and Phillips, as discussed in Section 4.2.2. Their model essentially couples the vertical mixing process with the mechanical energy balance of the lake. The Richardson number, F_{12} , is defined as the ratio of the buoyant potential energy deficit in the system due to density stratification to the rate of kinetic energy dissipation due to wind action. The friction velocity, F_{13} , is computed from wind speed (U_{13}), air density (U_{19}), surface water density (D_8), and the ice cover function (F_{14}), using functions given by Banks[4] and Wu[44]. The ice cover function, F_{14} , describes the effect of ice cover upon wind-induced mixing. When the computed surface temperature is greater than zero, F_{14} is set to one. For surface temperatures less than zero, F_{14} rapidly approaches a minimum value, a_{23} . Again, the exponential function has been employed only

4.3.3 System Functions (continued)

as a means of smoothing transitions and preventing instabilities in the numerical integration. The southern portion of Onondaga Lake generally remains free of ice cover, due to the discharge of heated effluents in this area[23]. This is the reason for the non-zero value of a_{23}.

With the exception of the second term in the denominator, a_{14}/a_{16}, the functional form for the exchange rate, F_{11}, corresponds exactly to the model employed by Kato and Phillips (see Figure 4.2-1). The added term effectively places an upper limit on the vertical mixing rate as the Richardson number approaches zero. This is necessary to prevent numerical instability in the calculations. The upper limit, determined by a_{16}, is sufficiently high to achieve essentially complete vertical mixing.

4.3.4 Parameter Values

The values and bases for the various parameter estimates are given in Table 4.3-4. Some of the parameters are associated with processes or relationships which are general and have defined in other modelling or experimental studies. Values for these have been taken directly from the literature ($a_6 - a_{11}$, a_{18}). Other parameters relate specifically to this particular lake or to this model. Some of these have been defined by measurements or data which are independent of the

Table 4.3-4

Parameter Definitions and Values

Parameter		Value	Units	Ref
a_1	Initial Dilution Ratio – Tributaries	0.	–	a
a_2	" " " – Metro STP	0.	–	a
a_3	" " " – Ninemile Creek	0.	–	a
a_4	Fraction of Allied Cooling Water Discharged to West Flume	0.273	–	23,7
a_5	Surface Mass Transfer Parameter	0.204	sec/(day-mmHg)	b
a_6	Conduction Parameter	0.279	mmHg	11
a_7	Back-Radiation Parameter	1.136×10^{-9}	$KKCal/m^2 {}^\circ K^4$-day	11
a_8	Atmospheric Radiation Parameter	1.064×10^{-14}	$KKCal/m^2 {}^\circ K^6$-day	11
a_9	" " "	0.17	–	11
a_{10}	Heat of Vaporization at $0^\circ C$	596.	Kcal/Kg	11
a_{11}	Effect of Temperature on Heat of Vap.	-0.54	Kcal/Kg-$^\circ C$	11
a_{12}	Vertical Heat Exchange via Molec. Dif.	2.74×10^{-3}	m/day	c
a_{13}	Vertical Chl. Exchange via Molec. Dif.	2.74×10^{-5}	m/day	c
a_{14}	Vertical Exchange Parameter	0.770	–	b
a_{15}	" " "	1.001	–	b
a_{16}	" " "	0.020	–	a
a_{17}	Thermocline Depth	9.0	m	21
a_{18}	Heat Capacity of Water	1.0	Kcal/Kg-$^\circ C$	10
a_{19}	Density Current Parameter	0.00001	g/cm^3	a
a_{20}	Max. Density Cur. Fraction – Tributaries	0.0	–	a
a_{21}	" " " " – Metro STP	0.0	–	a
a_{22}	" " " " – Ninemile Ck.	0.267	–	b
a_{23}	Ice Cover Parameter	0.46	–	d
a_{24}	Allied Cooling Water Flow at $0^\circ C$	2.47×10^5	m^3/day	e
a_{25}	Effect of Inlet Temp. on AC CW Flow	8.21×10^3	m^3/day-$^\circ C$	e

a – assumed value.

b – estimated empirically.

c – estimated from molecular diffusivities of heat and sodium chloride, assuming a thermocline thickness of 4 meters[23].

d – estimated from observed ice cover[23]; corresponds to a 90% reduction in wind-induced mixing under ice cover.

e – based upon analysis of flow data supplied by Clough[7] and lake temperature data supplied by Onondaga County[23]

4.3.4 Parameter Values (continued)

model (e.g. a_4, a_{12}, a_{13}, a_{17}, a_{23}, a_{24}, a_{25}). In some cases, model calculations are insensitive to particular parameter values, and reasonable values have been assumed (a_1, a_2, a_3, a_{16}, a_{19}, a_{20}, a_{21}). The remaining four parameters are lake- or model-specific, have important effects on the model calculations, and cannot be estimated accurately based upon independent evidence. These parameters (a_5, a_{14}, a_{15}, a_{22}) have been estimated by fitting model simulations to observed temperature and chloride data, using Bard's algorithms for parameter estimation in nonlinear dynamic systems. Demonstration and evaluation of the utility of these algorithms in ecosystem modelling is a primary emphasis of the work described in this chapter. The specifics of the parameter estimation exercises will be described in Section 4.5.

4.4 Implementation

A Fortran IV program has been written to perform the calculations specified by the model. The structure of the program is summarized in Figure 4.4-1. The roles and sizes of each of the 17 subroutines are given in Table 4.4-1. Programming and computations have been done on an IBM 370-168.

The system of equations is integrated numerically using a fourth-order, Runge-Kutta procedure with variable step size, contained in the IBM Scientific Subroutine Package[16]. This subroutine (RKGS) automatically adjusts integration step size for control of integration errors. The criterion used for step size adjustment is:

Table 4.4-1

Model Subroutines

Name	Function	Number of Statements	Reference
MAIN	initializes and controls program execution	129	
BOXES	computes compartment surface areas and volumes from morphometric data	27	
SETUP	reads observed temperature and chloride data from disc file and computes spatial averages	32	
CAV	computes weighting factors for spatial averaging of profile data	28	
RKGS	controls fourth-order Runge-Kutta numerical integration	115	16
DERIV	controls time-derivative computations	34	
OUTP	called after each Runge-Kutta step; traces calculations if specified	12	
PFUNC	updates forcing function vector	180	
PUP	updates parameter vector for sensitivity testing	38	
DERY	computes derivatives of state variables for given parameter vector	91	
INLET	distributes inflows and computes water balances	91	
RESLT	prints, plots, stores, and analyzes output at specified times and at end of program execution	287	
RESAN	analyzes and plots residuals	156	
HIST	prints histograms	54	
PLT	prints line plots	75	
TALLY	summarizes residual statistics	37	16
DENSY	function to compute density from temperature and chloride concentration	6	
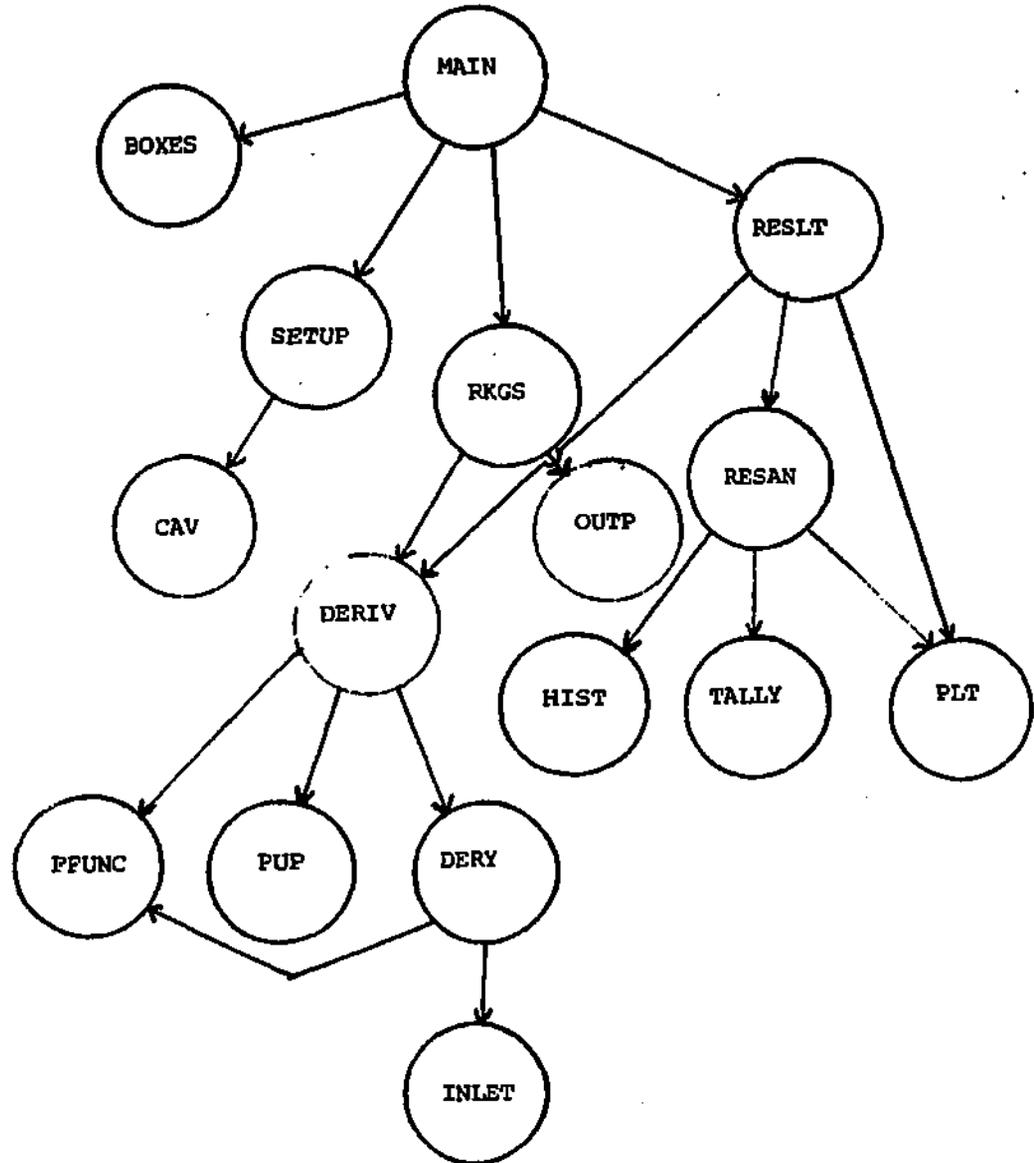
Figure 4.4 - 1

Schematic of Model Subroutines *

* Subroutines are called from left to right and top to bottom

4.4 Implementation (continued)

$$\delta = \frac{1}{15} \sum_{i=1}^{4} w_i \left| y_i^{(1)} - y_i^{(2)} \right| \leq \delta_{max} = 0.004 \qquad (4.4\text{-}1)$$

where,

w_i = weighting factor for state variable i, $i = 1,4$ ($\Sigma \, w_i = 1.0$)

$y_i^{(1)}$ = state variable value one time step ahead

$y_i^{(2)}$ = state variable value two time steps ahead

If the parameter δ is greater than 0.004, step size is halved and the computations are repeated. A reasonable value for δ_{max} was selected by trial-and-error. The weighting factors, w_i , are approximately inversely proportional to the annual ranges of the respective state variables, so that each variable contributes equally to the total integration error. A maximum step size of 0.025 years (9.13 days) has been specified.

The Runge-Kutta method requires evaluation of the derivative vector four times per each integration step. It can be relatively expensive compared to second-order, predictor-corrector methods[16]. The advantages of the former are that it is self-starting and relatively stable. The accuracy of this method is particularly important to successful implementation of the parameter estimation algorithms, as described in the next section. Adjustment of step size according to the above procedure can save considerable computation time. With the above error bound

4.4 Implementation (continued)

parameters, step size varies between about 4.5 days, during relatively
stable periods, to less than 0.5 days, during unstable periods (e.g.,
overturn). The yearly-average step size is about 3 days. Exclusive
of loading, the program consumes approximately 1.7 CPU seconds per
year of simulation on the IBM 370-168 for integration of one set of
state variable equations.

The program permits sensitivity testing by integrating more than
one set of state variable equations simultaneously. One additional
set is added for each parameter or forcing function sensitivity being
evaluated. For each set of state equations added, the value of the
corresponding parameter is increased by a given percent (typically 3%).
This permits calculations of sensitivity coefficients via finite-
differences as functions of time. The sensitivity coefficients are
essentially normalized first derivatives of the state variables with
respect to the parameter values.

A linear interpolation scheme has been employed to input the
values of time-variable forcing functions. As noted in Section 4.3.2,
these functions have been defined on monthly-average time scales.
Step changes in these variables would introduce instabilities in the
numerical integration routine. Accordingly, transitions from one
month to the next have been smoothed by adding a linear segment
extending from the average value for a given month to that of the
next in a time interval of one-half a month. Thus, the first quarter
of each month is the last part of a linear segment extending from the

4.4 Implementation (continued)

previous month's average to the current month's average, the middle half is constant at the average value, and the last quarter is the first part of a linear segment extending to the next month's average.

4.5 Parameter Estimation

A total of four parameters have been estimated by implementing Bard's algorithms for parameter estimation in nonlinear dynamic systems[2] . The general parameter estimation problem has been outlined in Chapter 1. Sections 4.5.1 and 4.5.2 discuss some of the specific options and alterations which have been implemented in this application. Section 4.5.3 describes the data base employed for parameter estimation purposes. Results are presented and interpreted in Section 4.5.4 and 4.5.5, respectively. Finally, the residuals are analyzed in 4.5.6.

4.5.1 Methods

As discussed in Chapter 1, the basic problem is to find a set of parameter values which maximizes an objective function computed from the residuals, subject to constraints imposed by the modal equations. As also noted in that chapter, the general problem can also include the affects of prior distributions of parameter estimates on the objective function, as well as constraints on parameter values. In this particular application, these additional options have not been exercised. The parameter solutions obtained are sufficiently far from any implied constraints (e.g.,positivity) that such constraints

4.5.1 Methods (continued)

are of no consequence to the solution. It was not found necessary to include constraints in order to guide convergence. Accordingly, the only constraints actually imposed are those dictated by the model equations.

The particular objective function employed depends upon the assumed form of the covariance matrix of residuals. As distributed, a total of six options are available in the Bard program:

$$\begin{bmatrix} \text{diagonal} \\ \text{non-diagonal} \end{bmatrix} \times \begin{bmatrix} \text{known} \\ \text{known to within a multiplicative constant} \\ \text{unknown} \end{bmatrix}$$

The most general form of the covariance matrix (non-diagonal x unknown) has been assumed. The corresponding form for the constant portion of the maximum likelihood objective function is :

$$\phi = -\frac{n}{2} \ln[\det(\frac{A}{n})] \tag{4.5-1}$$

$$a_{ij} = \sum_{m=1}^{n} e_{im} e_{jm} \tag{4.5-2}$$

where,

ϕ = objective function

n = number of observations

A = moment matrix of residuals

a_{ij} = element of matrix A; $i,j = 1,k$

e_{im} = residual for variable i and observation m; $i = 1,k$, $m = 1,n$

k = number of observed variables = 4

4.5.1 Methods (continued)

The Bard program provides two optional algorithms for guiding the convergence of an initially-assumed parameter vector to the optimal value: the Gauss-Newton method[1] and the Davidon-Fletcher-Powell method[9]. The former has been implemented here, based upon preliminary comparisons of the effectiveness of these algorithms on test systems[41] and upon recommendations in the Bard program manual[3].

In order to guide the convergence of the parameter vector to the solution, an estimate of the derivative of the objective function with respect to each of the parameter values is required. This is computed within the Bard program from the derivatives of the state variables with respect to the parameters. As distributed, the Bard program is designed for use with analytic derivatives, i.e., sensitivity equations. The discontinuities associated with the excessive number of conditional statements in the model render analytic differentiation impractical in this application. Accordingly, substantial modifications have been made in parts of the Bard program to permit derivative computation via finite-difference methods. For maximum efficiency, a finite-difference scheme should select parameter perturbation sizes which balance truncation error (increasing with increasing step size) against round-off error (decreasing with increasing step size). An algorithm suggested by Bard[3] has been employed to achieve this. This finite difference scheme enhances the flexibility of the model and the parameter estimation algorithms by facilitating coding changes made in model development and implementation.

4-56

4.5.1 Methods (continued)

Parts of the Bard program have also been modified in order to facilitate estimation of the 95% confidence regions for the parameter estimates. The confidence region is defined as that portion of parameter space which satisfies the following criterion:

$$\phi(\underline{\theta}^*) - \phi(\underline{\theta}) \leq \varepsilon \qquad (4.5\text{-}3)$$

where,

$$\phi(\underline{\theta}) = \text{objective function value for parameter vector } \underline{\theta}$$

$$\phi(\underline{\theta}^*) = \text{objective function value for parameter vector } \underline{\theta}^* = \text{solution}$$

$$\varepsilon = \text{indifference interval}$$

The region is estimated based upon a quadratic approximation of the response surface in the vicinity of the solution. A Taylor series expansion around the solution gives:

$$\phi(\underline{\theta}^* + \underline{\delta\theta}) \simeq \phi(\underline{\theta}^*) + \underline{q}^{*T} \underline{\delta\theta} + \frac{1}{2} \underline{\delta\theta}^T \underline{\underline{H}}^* \underline{\delta\theta} \qquad (4.5\text{-}4)$$

4.5.1 Methods (continued)

where,

$$\delta\theta = \theta - \theta*$$

$g*^T$ = transpose of vector of first partial derivatives of the objective function with respect to the parameters at the solution

$H*$ = matrix of second partial derivatives of the objective function with respect to the parameters at the solution

At the solution, the first derivatives are near zero:

$$g*^T \simeq 0 \qquad (4.5-5)$$

Thus,

$$\Phi(\theta* + \delta\theta) - \Phi(\theta*) = \Delta\Phi \simeq \frac{1}{2}\delta\theta^T H* \delta\theta \qquad (4.5-6)$$

The covariance matrix of the parameter values, V_θ, is given by[3]:

$$V_\theta \propto H*^{-1} \qquad (4.5-7)$$

4.5.1 Methods (continued)

Thus, from equations (6) and (7), the indifference region is given by:

$$\Delta\phi \ \simeq \ \frac{1}{2} \, \underline{\delta\theta}^T \, \underline{\underline{V_\theta}}^{-1} \, \underline{\delta\theta} \qquad\qquad (4.5\text{-}8)$$

Bard[3] has also shown that, for normally-distributed errors and if the estimate of V_θ is assumed correct, the quantity $| \, \underline{\delta\theta}^T \, \underline{\underline{V_\theta}}^{-1} \, \underline{\delta\theta} \, |$ is distributed as χ^2 with ℓ (= number of parameters) degrees of freedom. Thus, according to equations (3) and (8), the size of the indifference interval is given by:

$$| \, \Delta\phi \, | \ \simeq \ \frac{1}{2} \, | \, \underline{\delta\theta}^T \, \underline{\underline{V_\theta}}^{-1} \, \underline{\delta\theta} \, | \ \leq \ \epsilon \ = \ \frac{\chi^2}{2} \qquad\qquad (4.5\text{-}9)$$

This definition of the ϵ-difference region is essentially equivalent to a likelihood ratio test applied to the log-likelihood function, ϕ [12]. In order to translate this into a statement about actual parameter values, a principal component analysis of the parameter covariance matrix is first employed. The analysis, contained in the original form of Bard's program, computes linear functions of the parameter values which are independent of each other:

4.5.1 Methods (continued)

$$\rho_i = \sum_{j=1}^{\ell} b_{ij} \, \delta\theta_j \, , \qquad i = 1, \ell \qquad (4.5\text{-}10)$$

When defined in terms of $\delta\theta$, the expected value of each principal component is 0. and the standard deviation is σ_j. Expressed in terms of the principal components, the ε-difference region is given by:

$$\underline{\rho}^T \, \underline{\underline{V}}_\rho^{-1} \, \underline{\rho} \le \chi^2 \qquad (4.5\text{-}11)$$

The advantage of this representation is that the off-diagonal elements of $\underline{\underline{V}}_\rho^{-1}$ are zero, because the principal components are independent of each other. Accordingly, equation (11) can be evaluated as:

$$\sum_{i=1}^{\ell} \frac{\rho_i^2}{\sigma_i^2} \le \chi^2 \qquad (4.5\text{-}12)$$

The end points of each principal axis ρ_j are given by setting $\rho_i = 0$ for $i \ne j$ and solving ρ_j in equation (10):

4.5.1 Methods (continued)

$$\rho_j^2 = \chi^2 \sigma_j^2 \qquad (4.5\text{-}13)$$

$$\rho_j = \pm \chi \sigma_j \qquad j = 1, \ell \qquad (4.5\text{-}14)$$

$$\rho_i = 0, \; i \neq j \qquad (4.5\text{-}15)$$

For each principal component, the end points are determined in terms of the original θ coordinates by inverting and solving equation (10) for $\delta\theta_j$ in terms of ρ_j .

4.5.2 Implementation

Parameter estimation subroutines are listed and described in Table 4.5-1. Figure 4.5-1 depicts the calling sequences for the various subroutines in the program version designed for parameter estimation purposes. Substantial modifications to the Bard subroutine DER have been made to permit: (1) use of Runge-Kutta integration; (2) derivative (sensitivity coefficient) calculation by finite-difference methods; (3) output of the residual sums of squares for each observed variable after each evaluation of the objective function. Likewise, modifications to the Bard subroutine OUT, called after the solution has been reached, have been made to permit:

Table 4.5 - 1

Parameter Estimation Subroutines

Name	Function	Number of Statements	Reference
MAIN	initializes and controls program execution;creates interface between model and Bard subroutines	123	
NLMAX	controls parameter convergence via Gauss-Newton method	222	2
MAP	maps values of objective function for specified parameter values	20	
ACCUM	computes value of objective function for given parameter vector	175	2
EIG	computes scaled eigenvalues and eigenvectors of a given matrix	82	2
BOUND	computes constraint penalty functions and their derivatives	6	2[a]
PRIOR	computes effect of prior distribution of parameters on objective function	3	2[a]
DER	controls integration of model equations between specified time limits; computes derivatives of state variables with respect to parameters	140	2[b]
RUN	sets initial conditions, if unknown	3	2[a]
XTOY	relates state variables and their parameteric derivatives to observed variables	11	2[a]
OUT	provides detailed output after solution has been reached; analyzes response surface in vicinity of solution	184	2[b]
MINV	inverts a matrix	89	16
SIMQ	solves a set of simultaneous equations for end points of principal component axes of parameter estimates	52	16

a - not of consequence in this application
b - original subroutine substantially modified; see text.
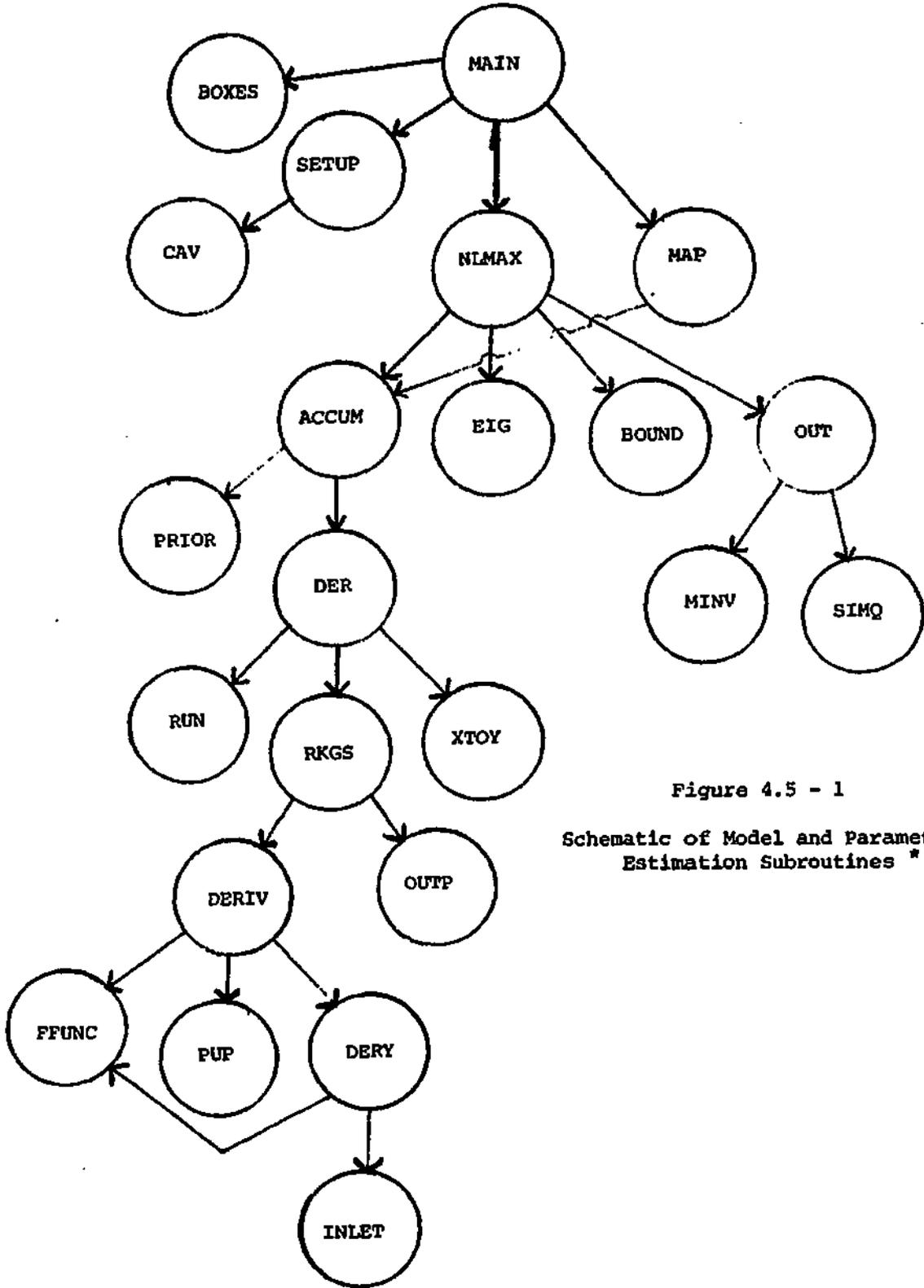
Figure 4.5 - 1

Schematic of Model and Parameter
Estimation Subroutines *

* Subroutines are called from left to right and top to bottom

4.5.2 Implementation (continued)

(1) computation of residual and parameter correlation matrices;
(2) computation of residual serial correlation coefficients; (3)
implementation of a multivariate test for significant bias in the
residuals; (4) estimation of the 95% confidence region for the
parameter vector. In addition, the subroutine MAP has been added
to permit direct sampling of the response surface for a given set
of parameter values. Any modifications of the Bard subroutines
have been designed for general application, i.e., they are not
specific for use with this particular model.

4.5.3 Lake Data

Temperature and chloride profile data from the O'Brien and Gere
surveys[23,21] have served as a basis for parameter estimates. From
April, 1968 through December, 1974, a total of 136 complete profiles
at 3-meter depth increments were available. These data have been
spatially aggregated to provide epilimnion- and hypolimnion-average
values for each sampling date. Within each level, averaging has been
done by integrating concentration times surface area with respect to
depth (using the trapezoidal rule), and dividing the result by total
compartment (epilimnion or hypolimnion) volume. Accordingly, the
weighting factors applied to each seven-membered profile vector
(0 - 18 meters) to compute the epilimnic average (0 - 9 meters) are:

Table 4.5-2

Temperature and Chloride Data Used for Parameter Estimation

Obs.	Time[a]	T_R	T_H	C_R	C_H	Obs.	Time[a]	T_R	T_H	C_R	C_H
1	1.243	10.6	7.6	1464.4	1929.1	69	3.693	20.8	15.6	1588.7	1818.1
2	1.332	11.3	8.7	1460.9	1393.7	70	3.769	20.8	13.3	1431.6	1707.0
3	1.373	13.8	9.8	1564.3	2104.6	71	3.707	17.5	13.9	1927.4	2400.0
4	1.427	18.5	12.2	1557.7	2126.6	72	3.789	14.7	13.4	1681.1	2056.0
5	1.449	20.6	12.9	1524.1	2143.0	73	3.895	10.5	10.6	1619.2	1906.9
6	1.466	20.1	12.3	1582.0	2146.6	74	3.923	7.6	7.7	1783.1	1948.5
7	1.493	18.5	13.1	1532.0	2067.2	75	3.947	1.3	1.6	1421.2	1662.1
8	1.523	21.1	13.0	1292.9	2000.8	76	4.035	1.5	2.0	1462.0	1753.1
9	1.542	22.7	13.6	1337.3	2090.4	77	4.099	2.0	2.9	1465.6	1737.8
10	1.547	23.1	12.0	1470.2	1937.8	78	4.132	1.6	2.2	1362.2	1921.3
11	1.564	22.1	13.1	1605.4	2105.2	79	4.203	2.1	2.2	1238.9	1868.2
12	1.619	23.4	12.5	1659.7	2162.4	80	4.247	3.3	2.4	901.1	1621.3
13	1.658	22.2	13.6	1565.9	2189.4	81	4.208	6.2	4.5	861.6	1329.4
14	1.646	20.8	14.2	1805.9	2303.1	82	4.355	11.1	7.0	970.4	1391.2
15	1.751	19.1	14.4	1332.7	1665.7	83	4.595	18.9	7.9	1461.6	1622.9
16	1.770	16.3	13.7	1822.2	2146.0	84	4.648	22.2	12.1	1553.4	1769.4
17	1.808	16.1	14.3	1843.2	2462.1	85	5.058	1.0	1.2	1863.4	1942.4
18	1.836	12.2	12.8	1367.6	1924.6	86	5.167	0.7	1.0	1650.4	2326.9
19	1.847	12.2	12.6	1626.9	2136.4	87	5.247	2.4	1.3	1196.4	2237.5
20	1.923	6.3	7.5	1374.9	1750.0	88	5.266	2.4	1.3	1111.3	2122.8
21	1.978	2.1	2.7	1711.4	1346.1	89	5.321	7.7	3.2	1091.9	1806.9
22	2.038	0.6	2.0	1456.0	1619.6	90	5.397	16.7	9.0	1179.1	1741.5
23	2.063	1.0	1.8	1168.1	1731.0	91	5.436	16.9	8.2	1051.7	1767.7
24	2.104	0.5	2.7	1494.8	1616.5	92	5.512	18.9	8.8	647.1	1588.1
25	2.123	1.0	2.0	1409.2	1603.3	93	5.551	21.6	7.9	966.8	1772.5
26	2.137	1.7	2.7	1482.6	1763.3	94	5.589	21.2	7.6	1177.9	1804.1
27	2.175	1.6	2.6	1456.3	1633.5	95	5.627	19.4	12.1	1261.0	1817.8
28	2.238	3.3	3.0	1362.6	1487.4	96	5.666	21.2	14.5	1461.8	1949.4
29	2.258	3.8	3.0	1323.0	1934.0	97	5.704	22.0	10.8	1490.2	1979.2
30	2.271	5.4	3.6	1228.0	1607.6	98	5.742	16.3	11.0	1512.4	2010.9
31	2.270	7.8	5.3	1471.3	1572.0	99	5.704	14.1	9.9	1674.8	1974.8
32	2.310	9.3	5.8	1093.4	1761.5	100	5.819	10.5	8.9	1706.6	2117.2
33	2.329	10.2	6.2	1278.3	1685.8	101	5.860	9.1	9.4	1405.5	1941.6
34	2.348	11.9	6.5	1150.0	1688.1	102	5.953	3.4	5.3	1382.5	1497.1
35	2.367	12.2	7.9	1662.6	1681.8	103	6.071	5.8	6.1	1135.7	1366.2
36	2.386	15.4	8.0	1277.4	1643.7	104	6.184	3.0	3.2	1224.4	1490.4
37	2.403	15.1	7.8	963.9	1654.1	105	6.222	3.3	3.2	1149.4	1435.0
38	2.425	17.0	9.1	1081.3	1534.1	106	6.279	5.0	5.0	1017.9	1146.5
39	2.444	17.6	8.2	1479.1	1935.3	107	6.318	9.8	7.4	1008.1	1410.7
40	2.463	18.3	8.1	1411.4	1793.3	108	6.356	11.7	10.3	1168.5	1421.1
41	2.482	16.5	9.2	1233.7	1771.3	109	6.395	11.9	9.9	1114.0	1474.6
42	2.501	20.4	9.5	1537.4	1674.5	110	6.452	20.1	11.2	1044.4	1483.9
43	2.521	21.3	9.4	1487.0	1742.7	111	6.525	21.4	12.5	1276.6	1534.2
44	2.546	23.6	10.3	1395.8	1649.4	112	6.567	22.6	13.2	1374.4	1577.8
45	2.655	23.3	12.1	1677.7	1852.7	113	6.605	24.0	15.1	1481.7	1613.7
46	2.674	24.1	13.0	1774.5	1331.7	114	6.644	21.6	15.1	1527.6	1714.6
47	2.715	20.9	13.7	1674.1	1722.3	115	6.682	23.5	16.1	1563.1	1782.9
48	2.732	19.9	12.8	1790.3	1896.2	116	6.721	16.4	14.7	1554.5	1641.8
49	2.770	17.1	13.0	1847.3	1940.3	117	6.739	16.8	14.6	1515.7	1722.9
50	2.811	12.9	13.0	2004.7	2062.2	118	6.797	13.0	12.7	1707.8	1726.9
51	2.849	10.8	10.9	1908.8	2045.7	119	6.932	7.4	6.4	1691.3	1731.5
52	2.890	8.9	9.3	2025.1	2054.3	120	7.161	3.0	3.0	1633.0	1593.1
53	2.920	5.2	5.9	1891.8	2310.1	121	7.238	0.3	0.5	1420.4	1425.8
54	2.964	3.1	5.1	1876.2	2193.1	122	7.277	3.4	3.7	1300.8	1742.7
55	3.044	1.9	3.7	1532.7	2217.6	123	7.315	7.0	5.7	1069.0	1254.5
56	3.079	1.7	1.8	1503.7	2171.3	124	7.353	9.3	10.0	1117.2	1141.3
57	3.118	2.6	4.4	1371.1	2292.7	125	7.392	14.7	12.7	997.4	1106.5
58	3.153	2.7	4.3	1635.8	2245.0	126	7.430	17.7	12.4	1022.2	1174.4
59	3.320	9.8	6.7	1641.1	2072.0	127	7.468	18.7	14.3	1106.7	1192.6
60	3.366	18.9	12.4	1695.9	2184.5	128	7.526	21.3	14.9	1117.5	1413.7
61	3.429	19.4	10.2	1463.2	1821.6	129	7.564	20.0	14.2	1389.7	1576.9
62	3.463	21.4	11.0	1616.8	1961.8	130	7.603	21.7	14.6	1632.7	1739.6
63	3.462	19.8	9.6	1642.2	1971.0	131	7.641	21.6	15.8	1779.3	1714.2
64	3.515	21.3	11.3	1163.0	1596.4	132	7.679	18.4	14.1	1698.3	1773.5
65	3.540	20.7	10.6	1108.4	1567.2	133	7.718	17.4	15.4	1864.3	2033.8
66	3.575	22.9	11.7	1186.4	1567.2	134	7.756	13.2	13.5	1945.3	1951.8
67	3.614	23.7	10.7	1643.9	2039.2	135	7.795	12.5	12.0	1818.1	1814.1
68	3.674	20.4	12.4	1645.4	1901.7	136	7.833	10.7	10.9	1767.5	1893.3

a - Time in Years from January 1, 1967.

4.5.3 Lake Data (continued)

0.210, 0.351, 0.305, 0.134, 0., 0., and 0. The corresponding set of
weights for the computation of the hypolimnic average (9 - 18 meters)
are: 0., 0., 0., 0.210, 0.364, 0.286, and 0.140. The data are listed
in Table 4.5-2. In estimating the parameters, the first observation
has been employed as an initial condition, leaving a total of 135
observations as a basis for parameter estimation.

4.5.4 Results

Initial estimates of the parameters are required in order to
start the formal parameter estimation routine. Independent evidence
suggested values for each of the four parameters. a_{14} and a_{15}
represent the intercept and slope of the exchange rate function,
F_{11} , and correspond to the intercept and slope on the plot of Kato
and Phillips's experimental data on exchange rate versus Richardson
Number (Figure 4.2-1). The data of Kato and Phillips suggested that
values of 2.5 and 1.0 would be appropriate for a_{14} and a_{15} ,
respectively. Because of the logarithmic nature of this relation-
ship, the logarithm of a_{14} was actually estimated. This effectively
increased the linearity of the response surface in the region of the
solution, a desirable aspect from the point of view of estimating the
parameter covariance matrix. The surface heat and mass transfer
parameter, a_5 , is specified as 0.18 by Harleman and Markofsky[11].

4.5.4 Results (continued)

This parameter is multiplied by wind speed to compute the transfer coefficient for evaporation and conduction. It has been included as an unknown here partially because of uncertainty as to the adequacy of wind speed data measured at Hancock Airport as representative of conditions over the lake. Thus, the final estimate for this parameter may include a correction factor for average wind speed, as well as any deviations of the actual parameter from the suggested value.

Initial simulations indicated that a value of 1.0 for parameter a_{22} , the maximum fraction of Ninemile Creek flow entering the hypolimnion, resulted in consistent over-prediction of mid-summer hypolimnic temperatures, even if the turbulent exchange parameter, a_{14} , was set to zero. The over-prediction became more severe with increasing values of the initial dilution ratio, a_3 . This suggested that, despite its relative density, all of Ninemile Creek does not flow into the hypolimnion. Apparently, much of it is dispersed in the epilimnion before it has a chance to sink. If the fraction entering the hypolimnion is assumed to be zero, the chloride gradients are consistently under-predicted. Thus, it has been assumed that some portion of the flow enters the hypolimnion without initial dilution. The value for parameter a_{22} has accordingly been assumed to lie in the range of zero to one.

4.5.4 Results (continued)

The results of the final parameter estimation run are given in Table 4.5-3. Principal characteristics of temperature and chloride residuals obtained from this run are summarized in Table 4.5-4. The initial parameter guesses employed in this case correspond to a solution previously obtained with parameter a_{16} equal to 0.01. This parameter determines the maximum rate of mixing under unstable (negative density gradient) conditions. Examination of residuals from this fit suggested that a higher maximum rate would be desirable. With some testing it was found that a doubling of a_{16} would enhance the rate without appreciably affecting the stability of the calculations. Accordingly, a_{16} was set to 0.02 and the parameter estimation routine was re-started at the solution obtained for $a_{16} = 0.01$. As shown in Table 4.5-2, the intercept and slope parameters, a_{14} and a_{15}, were reduced somewhat, but the best estimates of the other two parameters changed only slightly. Convergence to the solution was achieved with 7 iterations (derivative evaluations) and 16 objective function evaluations, or a total of $16 + 4 \times 7$ or 44 equivalent function evaluations. This consumed a total of 8.15 minutes of CPU time on the IBM 370-168. The performance of the algorithm will be discussed in greater detail in Section 4.7.

Strictly speaking, it cannot be guaranteed that the solution obtained above is a global one, because of the inherent nature of the nonlinear estimation problem. Generally, the robustness of the solution

Table 4.5-.3

Parameter Estimates and Confidence Regions Based upon All Data [e]

	Parameter a_{14}[b]	a_{15}	a_{22}	a_5	Objective Function
Estimate[f]	-0.2617	1.0012	0.2668	0.2062	-1542.65 = φ*
Standard Deviation	0.2872	0.0367	0.0122	0.0095	
Correlation Matrix					
a_{14}	1.0				
a_{15}	0.989	1.0			
a_{22}	0.249	0.144	1.0		
a_5	0.224	0.205	-0.108	1.0	σ_j Standard Deviation
Principal Component Coefficients					
ρ_1	2.36	18.10	17.42	24.80	1.458
ρ_2	-0.12	0.62	-59.61	72.43	1.053
ρ_3	-0.59	-7.58	53.43	72.36	0.872
ρ_4	-2.49	18.89	6.71	2.92	0.066
End Points of Principal Axes [c]					$\phi_j - \phi^*$[d]
$\rho_1 = -\chi \sigma_1$	-1.135	0.892	0.255	0.196	-3.23
$\rho_1 = +\chi \sigma_1$	0.612	1.111	0.278	0.216	-5.04
$\rho_2 = -\chi \sigma_2$	-0.230	0.998	0.295	0.185	-4.75
$\rho_2 = +\chi \sigma_2$	-0.293	1.004	0.238	0.227	-2.76
$\rho_3 = -\chi \sigma_3$	-0.131	1.029	0.246	0.189	-5.34
$\rho_3 = +\chi \sigma_3$	-0.392	0.974	0.288	0.224	-3.27
$\rho_4 = -\chi \sigma_4$	-0.220	0.996	0.267	0.206	-4.45
$\rho_4 = +\chi \sigma_4$	-0.303	1.006	0.267	0.206	-4.25

b natural logarithm transform

c correspond to approximate bounds of 95% confidence region (see eq. 4.5-13 to 15)

$\chi = \sqrt{\chi^2_{.05}}$ with 4 degrees of freedom = 3.08

d sampled objective function value at end point of principal axis - value at solution ; expected value, based upon quadratic approximation of response surface (equation 4.5-6), = -4.75 = $-\chi^2_{.05}/2.0$

e Total of 135 profiles between April, 1968 and December, 1974

f initial guesses : .5096, 1.091, .2632, .2110; ϕ_o = -1552.17

Table 4.5-4

Characteristics of Temperature and Chloride Residuals

Variable	T_E	T_H	C_E	C_H
Units	$^\circ C$	$^\circ C$	g/m^3	g/m^3
Mean	0.280	0.289	-33.5	-15.8
Standard Error[a]	1.676	1.873	198.9	185.3
Serial Correlation Coefficient	0.453	0.603	0.511	0.517
Correlation Matrix				
T_E	1.0			
T_H	0.370	1.0		
C_E	-0.103	0.070	1.0	
C_H	0.076	0.132	0.461	1.0
O S S [b]	0.788×10^4	0.260×10^4	0.994×10^7	0.102×10^8
R S S [c]	0.377×10^3	0.470×10^3	0.530×10^7	0.460×10^7
R^2 [d]	0.952	0.819	0.467	0.549

a - Standard Error = (Residual Sum of Squares/ 134)$^{0.5}$

b - O S S = Observed sum of squares of deviations from means.

c - R S S = Residual sum of squares of deviations from zero.

d - R^2 = 1 - (R S S / O S S)

4.5.4 Results (continued)

would be tested by starting the estimation routine at a variety of
different initial parameter values and testing whether they all
converge to the same point. However, the dimension of the parameter
vector and the cost of execution renders this approach impractical
in this application. The adequacy of the solution is indicated
first by the apparent feasibility of the parameter estimates, compared
to the expected values and ranges discussed above. Secondly, the
parameters are all relatively well-defined, with coefficients of
variation ranging from 0.036 to 0.287, or corresponding t values
ranging from 27.3 to 3.5*. This indicates that this maximum of the
objective function is a relatively distinct region in parameter-
space. Using the mapping option added to the Bard program, the
objective function has been evaluated at the end points of the
principal component axes. These evaluations (Table 4.5-3) indicate
that the response surface is fairly well-behaved in the vicinity of
the solution, since the computed values of the objective function
at the end points are not greatly different from those estimated
based upon a quadratic approximation of the response surface
(equation 4.5-6), which predicts that the solution should be 4.75,
as compared with computed values ranging from 3.23 to 5.34. This
indicates that response of the function to changes in the parameter

---

* a_{14} was estimated on a log scale; the standard deviation of the
logarithm of a_{14} approximately equals the coefficient of variation
of the parameter on a linear scale.

4.5.4 Results (continued)

values near the solution is not highly non-linear. The solution is not a "peak" in the objective function, but rather a high point on a ridge, because of the high correlation between parameters a_{14} and a_{15} ($r = 0.989$).

As discussed in Chapter 1, parameter stability is a desirable property of a valid model. In order to develop further evidence for the verification of the model, the stability of the parameter estimates has been examined by dividing the data set into six groups and estimating optimal parameters for each group separately. The six groups correspond to years, with the exception that data for 1970 and 1971 have been combined, because only nine observations were available for the latter. To ease parameter estimation difficulties and reduce costs, only two of the four parameters have bsen estimated for each group: a_{14} and a_{22}. The remaining two, a_{15} and a_{5}, have been constrained to the values obtained above. As will be discussed in the next section, the value of a_{15} has a theoretical basis and conforms to the value derived by Kato and Phillips. It is also highly-correlated with a_{14}. a_{5} corresonds roughly to the value suggested by Harleman and Markofsky[11]. As will be ehown, this parameter is not as critical to analysis of the outfall design issue as are the other parameters, because it does not strongly influence vertical exchange rates. The other two parameters are more empirical and do not have external bases for

4.5.4 Results (continued)

their values. Accordingly, these have served as fool for parameter stability studies.

Sampling error in the covariance matrix of the residuals can become important as sample size is reduced. In order to minimize such effects and permit a focus on parameter variations alone, the covariance matrix of the residuals for the individual data groups has been assumed to be equal to that obtained from all the data combined (Table 4.5-4). In this case, the constant portion of the log-likelihood objective function employed in the Bard program is:

$$\Phi = -\frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} a_{ij} \, [v]_{ij}^{-1} \qquad (4.5\text{-}16)$$

where,

a_{ij} is defined in equation 4.5-2

$[v]_{ij}^{-1} =$ element of inverse of specified covariance matrix

This amounts to weighted least squares, with known weights determined from the inverse of the specified covariance matrix.

4.5.4 Results (continued)

The assumed initial conditions for each period have been determined by the last observation in the previous period, with the exception of the 1968 and 1972 groups, in which cases the first observations in the respective years have served as initial conditions. The solution obtained for all years together has also been used as a starting point for the parameter vector in the estimation routine. Results of the analysis are given in Table 4.5-5. Approximate 95% confidence regions for the parameter estimates are illustrated in Figure 4.5-2.

Optimization of the parameter estimates for the individual years has increased the objective functions by values ranging from 0.59 for 1973 to 6.50 for 1968. Based upon the test discussed in Section 4.5.1, the ϵ-indifference region for the objective function which defines the parameter confidence region is approximately given by $\chi^2/2$, for normally-distributed errors. For two parameters, the 95% confidence region is thus defined by $\Delta\Phi = 3.0$, if the estimate of $\underline{V}_\beta$ is assumed correct. Observed increases in the objective functions for the individual years were all less than this value, with the exception of 1968. Thus, for five out of six groups, the optimal parameters were not significantly different from the best overall estimates.

1968 was the only year in which the simulations were begun in April, after the onset of stratification. It is possible that the differences in the optimal parameter estimates for that year could

Table 4.5-5

Best Parameter Estimates for Individual Years

Years	1968		1969		1970-71		1972		1973		1974	
N	20		33		30		17		17		17	
Φ_0 a	-41.75		-47.83		-69.42		-41.19		-29.06		-33.93	
Φ_* b	-35.25		-46.93		-68.42		-39.17		-28.47		-32.05	
$\Phi_* - \Phi_0$ c	+6.50 *		+0.90		+1.00		+2.02		+0.59		+1.88	
Parameters d	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Mean	-0.690	0.299	-0.341	0.235	-0.269	0.242	-0.414	0.298	-0.206	0.266	-0.140	0.242
Standard Dev.	0.237	0.030	0.093	0.023	0.064	0.020	0.110	0.032	0.077	0.038	0.103	0.039
End Points of Principal Axes [e]												
	-1.181	0.231	-0.559	0.181	-0.339	0.264	-0.677	0.220	-0.270	0.297	-0.257	0.287
	-0.099	0.368	-0.123	0.290	-0.199	0.221	-0.150	0.376	-0.140	0.235	-0.021	0.197
	-0.433	0.273	-0.278	0.220	-0.408	0.199	-0.360	0.283	-0.303	0.180	-0.363	0.156
	-0.847	0.325	-0.404	0.251	-0.130	0.286	-0.465	0.313	-0.028	0.351	0.083	0.327

a - Φ_0 = objective function value at starting point = best estimate for all years [θ_1 = -0.262, θ_2 = 0.267]

b - Φ_* = objective function value at solution for given year

c - * indicates that $\Phi_* - \Phi_0$ is significantly different from 0 at the 95% confidence level if normally-distributed errors are assumed

d - θ_1 = ln (a_{14}), θ_2 = a_{22}, a_{15} = 1.0012, a_5 = 0.2062

e - corresponds to approximate bounds of 95 % confidence region for parameter estimates, assuming normally-distributed errors

Figure 4.5-2

Approximate 95% Confidence Regions for Estimates of Parameters a_{14} and a_{22}

Obtained for Various Years
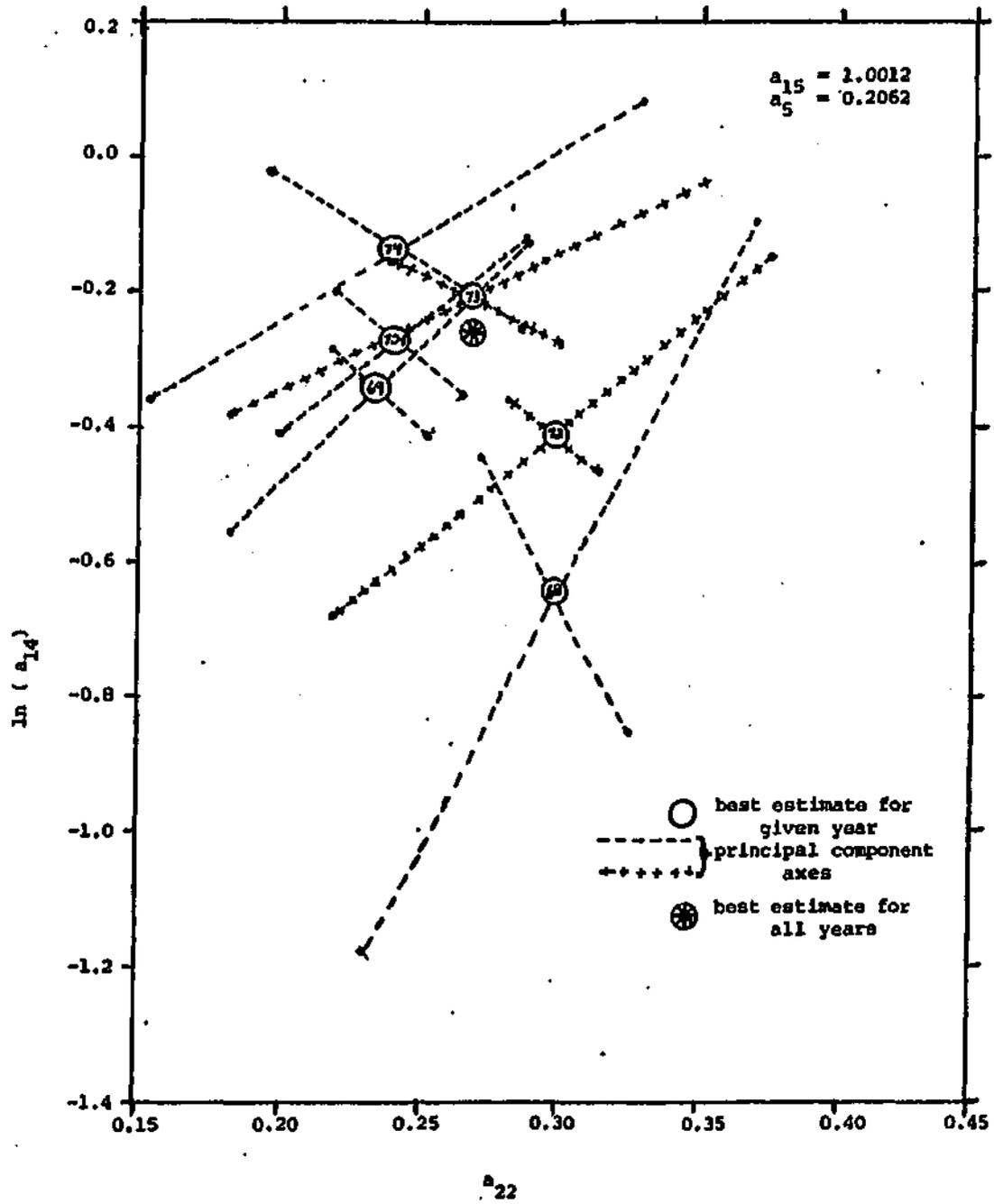
4.5.4 Results (continued)

be attributed to the initial conditions. Errors in the initial
conditions at this point could have profound effects on model
simulations for the entire stratification period. Some of the
seasonal biases in model performance, discussed in Section 4.5.6,
could also be involved. In addition, the sensitivity patterns to
parameter and forcing function values appear to be different in
1968 than in other years, as will be discussed in Section 4.6.1.

4.5.5 Interpretation of Parameter Estimates

Estimates of the parameters dealing with the vertical exchange
function, F_{11} , can be interpreted from a mechanical energy balance
point of view. The basis of the mixing model is the relationship
between kinetic energy input at the surface due to wind shear stress
and the increase in buoyant potential energy due to mixing. From
the definition of the surface friction velocity[4], energy input to
the lake is given by[29]:

$$\frac{d \overline{KE}}{dt} = D_E F_{13}^3 A_E \tag{4.5-17}$$

4.5.5 Interpretation of Parameter Estimates (continued)

where,

$$F_{13} = \text{surface friction velocity (m/sec)}$$
(see Table 4.3-3)

For vertical exchange at a rate F_{11} between the hypolimnion and the epilimnion, the rate of change of buoyant potential energy is given by:

$$\frac{d\,\overline{PE}}{dt} = g\,F_{11}\,A_H\,\overline{\Delta Z}_G\,(D_H - D_E) \tag{4.5-18}$$

where,

$$g = \text{acceleration of gravity} = 9.8 \text{ m/sec}^2$$

$$\overline{\Delta Z}_G = \text{difference in center of gravity between the epilimnion and hypolimnion} = 8.5 \text{ meters*}$$

In units of meters per second, the exchange rate function at moderate to high Richardson Numbers is given by**:

---

* Determined from lake surface area versus depth plot (Chapter 3).

** The second term in the denominator of the exchange rate function in Table 4.3-3 is negligible during stratified periods.

4.5.5 Interpretation of Parameter Estimates (continued)

$$P_{11} = \frac{a_{14} P_{13}}{\dfrac{a_{15}}{F_{12}}} \qquad (4.5\text{-}19)$$

$$F_{12} = \text{Richardson Number} = \frac{g\, a_{17}\, (D_H - D_E)}{D_E\, F_{13}^2} \qquad (4.5\text{-}20)$$

Substituting in the appropriate parameter values, and assuming $a_{15} = 1.0012 \simeq 1$:

$$F_{11} = \frac{0.086\, D_E\, F_{13}^3}{g\, (D_H - D_E)} \qquad (4.5\text{-}21)$$

Combining with equation (18):

$$\frac{d\, \overline{PE}}{dt} = 0.73\, D_E\, F_{13}^3\, A_H \qquad (4.5\text{-}22)$$

This result can be compared with the expression for the rate of kinetic energy input [equation (17)]:

4.5.5 Interpretation of Parameter Estimates (continued)

$$\frac{d \overline{PE}}{dt} = 0.73 \frac{A_H}{A_E} \frac{d \overline{KE}}{dt} = 0.47 \frac{d \overline{KE}}{dt} \qquad (4.5-23)$$

Thus, the parameter estimates indicate that 47% of the kinetic energy supplied to the system by the wind is used to increase the potential energy of the system by mixing the hypolimnic and epilimnic waters. The remainder is presumably dissipated as heat of friction.

Implicit in the estimate of a_{14} and, therefore, in the estimate of 47% conversion, is a correction factor for average wind speed over the lake relative to average wind speed at Hancock Airport. Thus, while the model calculations for this lake are not sensitive to any differences in average wind speeds, use of the same parameter values in modelling other lakes may not be appropriate, unless adequate adjustements are made. The estimate of the slope, a_{15}, is independent of such errors and would be expected to be more generally applicable.

The formulation employed by Kato and Phillips to represent their experimental data has been shown to be useful in modelling the mixing process in a real lake. The best estimate of the slope, a_{15}, is essentially the same as that employed by Kato and Phillips. The value of 1 gives rise to the mechanical energy balance arguments presented

4.5.5 Interpretation of Parameter Estimates (continued)

above. The intercepts, a_{14} , are appreciably different (0.77 versus
2.5). As discussed in Section 4.2, Kato and Phillips's value of 2.5
indicates that the rate of change of potential energy in their system
was 125% of the rate of kinetic energy input. It was suggested that
this could be attributed to possible errors in their kinetic energy
input rates. The same general type of problem may exist in this case,
with the possible errors in average wind speed discussed above. One
important difference between the two applications is that Kato and
Phillips applied the model to represent entrainment (i.e., one-way
transfer from the bottom to the top layer, resulting in a moving
interface), whereas the function has been applied here to represent
two-way exchange between the bottom and top layers with a fixed inter-
face. Apparent differences in the a_{14} parameter estimates derived
from the two studies might also be attributed to this factor.

4.5.6 Analysis of Residuals

Essential properties of the residuals are summarized in Table 4.5-6. Comparisons of model predictions have been made with observations of temperature, chloride, and density for the epilimnion, hypolimnion, and vertical gradient (hypolimnion-epilimnion). Time series plots of observations and prediction are displayed in Figures 4.5-3 to 4.5-5. Residuals are plotted against time in Figures 4.5-6 to 4.5-8. Observations are plotted against predictions in Figure 4.5-9. Histograms and normal probability plots are depicted in Figures 4.5-10 and 4.5-11, respectively.

t-tests (line b, Table 4.5-6) indicate statistically-significant bias in the model predictions for six out of the nine components. However, the bias is generally small compared to the size of the residuals, since less than six percent of the total residual variance can be attributed to deviations of the mean from zero (line d). Effective R^2s range from 0.26, for chloride gradient, to 0.95 for surface temperature (line e). Both the Durbin-Watson statistics (line f) and the non-parametric runs tests (line g) indicate significant auto-correlation in the time series of residuals in all cases. This serial correlation is also evident in residual time series plots (Figures 4.5-6 to 4.5-8).

The Kolmogorov-Smirnov test has been applied to test for normality. If the calculated residual means and standard deviations are employed in the test, non-normality is indicated at the 95% confidence level

Table 4.5-6
Results of Residuals Analysis

Variable

		T_E	T_H	$T_H - T_E$	C_E	C_H	$C_H - C_E$	D_E	D_H	$D_H - D_E$
Observed Variable	Mean	13.25	9.03	-4.21	1448.	1811.	364.	1.39	2.68	1.30
	Standard Dev.	7.64	4.39	4.38	271.	.275.	232.	1.00	0.52	0.85
Estimated Variable	Mean	12.97	8.75	-4.22	1481.	1796.	315.	1.42	2.62	1.20
	Standard Dev.	8.21	5.30	4.67	294.	260.	202.	1.09	0.49	0.95
Residual	Mean	0.28	0.29	0.01	-33.	16.	49.	-0.03	0.06	0.10
	Standard Dev. a	1.65	1.84	1.99	195.	184.	193.	0.43	0.35	0.47
T-test for Bias	b	1.97*	1.83*	0.06	-1.98*	0.99	2.95*	0.87	2.08*	2.37*
Standard Error	c	1.67	1.87	1.99	198.	185.	199.	0.43	0.36	0.48
(Mean/ Std. Error)^2	d	.028	.024	<.001	.028	.007	.060	.006	.031	.040
R^2	e	.95	.82	.79	.47	.55	.26	.82	.53	.69
Durbin-Watson Statistic	f	1.07*	0.78*	0.83*	0.96*	0.97*	1.09*	1.10*	0.84*	0.99*
Runs Test Parameter	g	-5.43*	-6.47*	-6.28*	-4.74*	-4.16*	-4.46*	-4.79*	-4.27*	-4.25*
Prob(normality)	h	.855	.349	.045*	.463	.657	.387	.446	.111	.982
Prob(normality)	i	.104	.013*	.041*	.016*	.160	.019*	.107	.001*	.057
Yearly F Statistic	j	3.65*	1.66	1.72	5.29*	4.78*	4.48*	4.98*	5.55*	5.24*
Variance Fraction	k	.146	.072	.074	.199	.183	.174	.189	.206	.197
Seasonal F Statistic	l	8.87*	13.48*	13.32*	2.94*	3.24*	5.11*	2.72*	1.25	5.08*
Variance Fraction	m	.390	.492	.490	.174	.189	.269	.163	.083	.268

a - Standard Deviation around mean.
b - t = mean/standard error of mean ; degrees of freedom = 135.
c - Standard Error around zero = square root of mean squared residual.
d - fraction of residual variance attributed to bias.
e - R^2 = 1 - (Residual Sum of Squares)/(Observed Sum of Squares of Deviations around Mean).
f - autocorrelation present in time series at the 95% level if $d < d_L$ = 1.59 (ref. 17)
g - Runs test parameter = (observed number of runs - expected number)/standard deviation of expected number; distributed as N(0,1.) (ref. 2).
h - Kolmogorov-Smirnov Test, assuming mean and standard deviation given in lines 5 and 6 (ref. 12)
i - Kolmogorov-Smirnov Test, assuming mean = 0., standard deviation = standard error (line 8) (ref. 12)
j - Analysis of Variance Statistic for Yearly Stratification ; degrees of freedom = 6,128.
k - Fraction of total residual variance explained by yearly groupings = between-group sum of squares/ total sum of squares.
l - Analysis of Variance Statistic for Seasonal Stratification; degrees of freedom = 9,125.
m - Fraction of total residual variance explained by seasonal groupings = between-group sum of squares/ total sum of squares.
* - Null Hypothesis rejected at the 95% confidence level.

Figure 4.5-3
Observed(*) and Estimated(-) Temperatures Versus Time
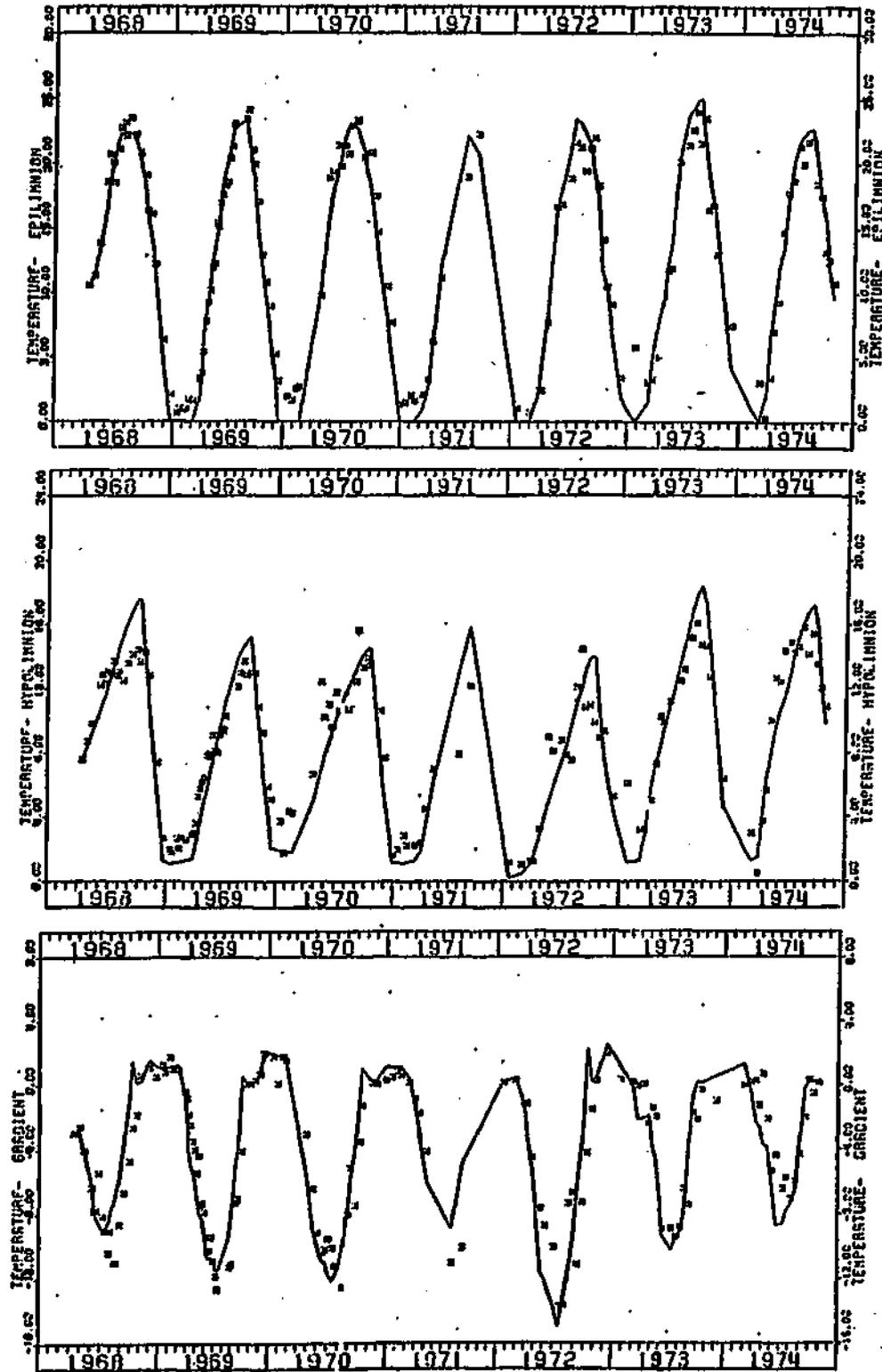
Figure 4.5-4

Observed(*) and Estimated(-) Chlorides Versus Time
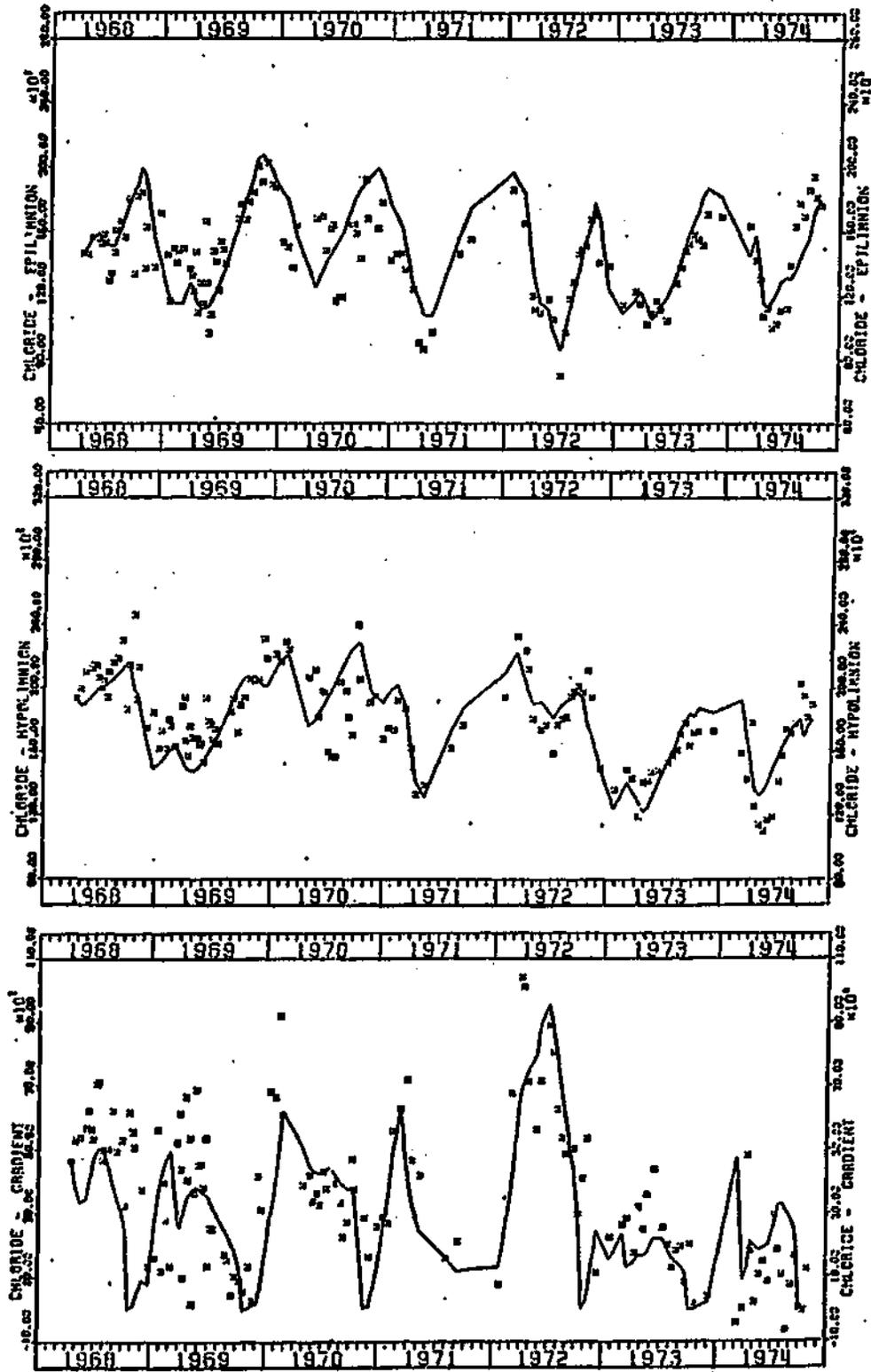
Figure 4.5-5

Observed(*) and Estimated(-) Densities Versus Time
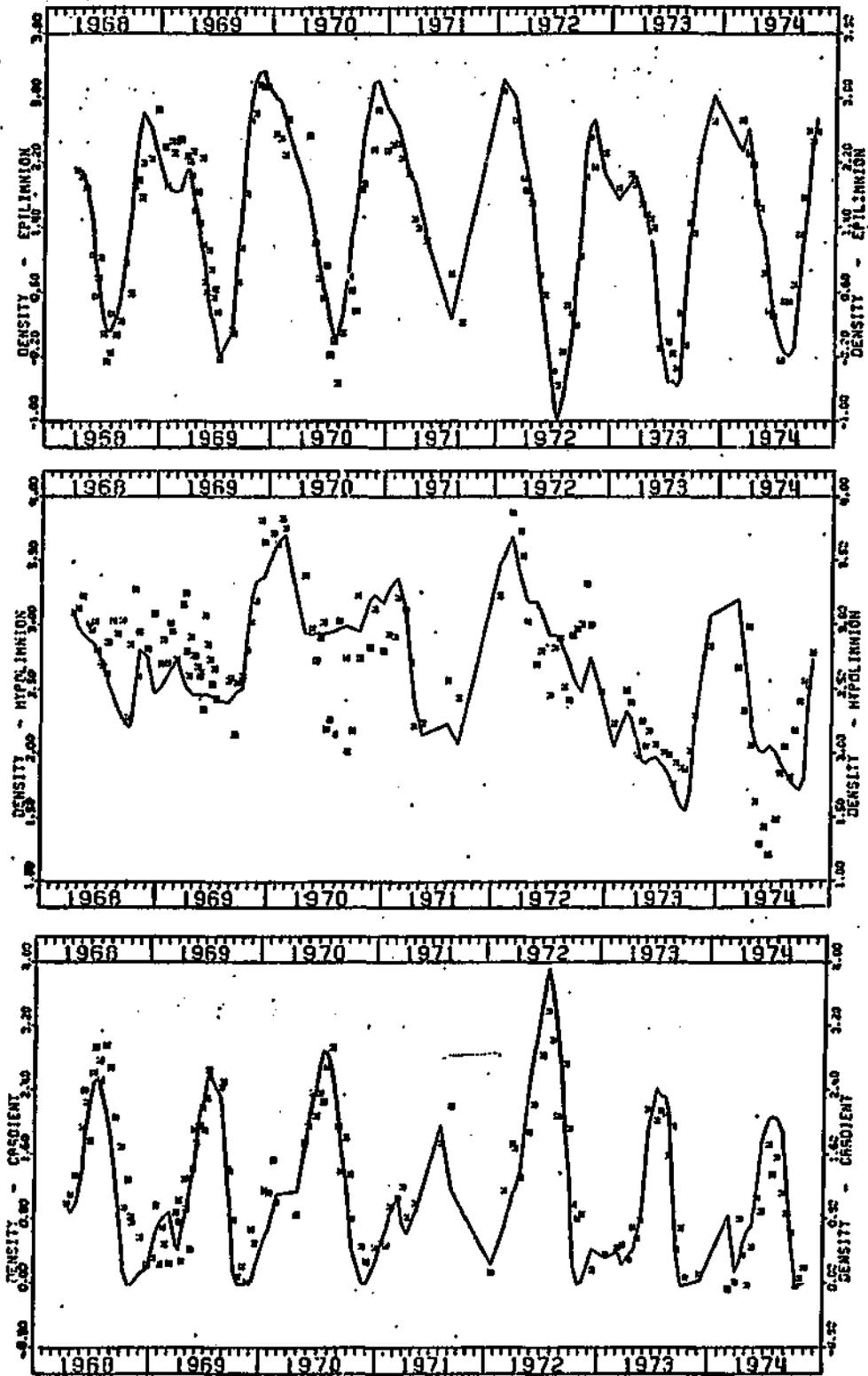
Figure 4.5-6
Temperature Residuals Versus Time
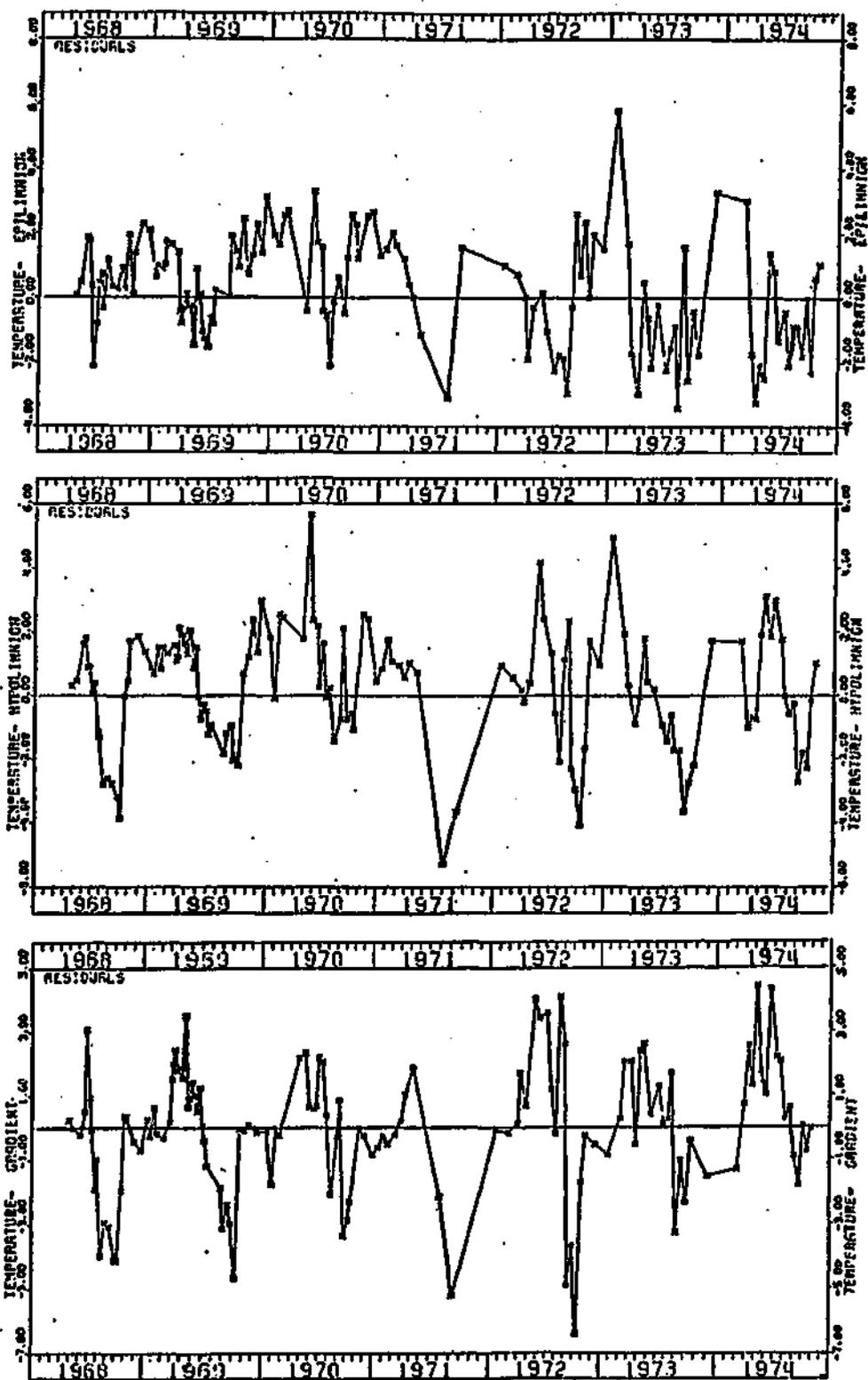
Figure 4.5-7

Chloride Residuals Versus Time
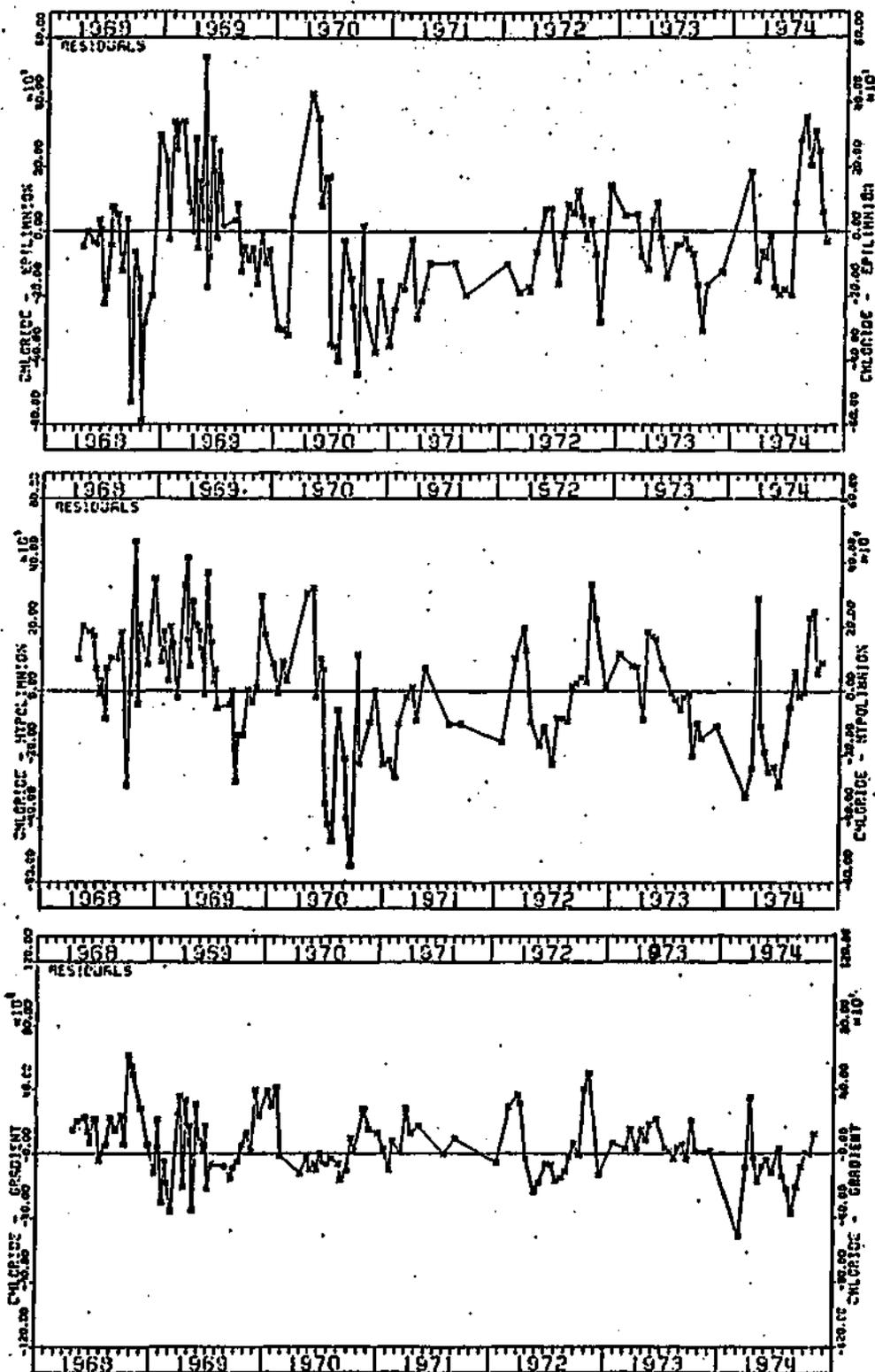
Figure 4.5-8

Density Residuals Versus Time
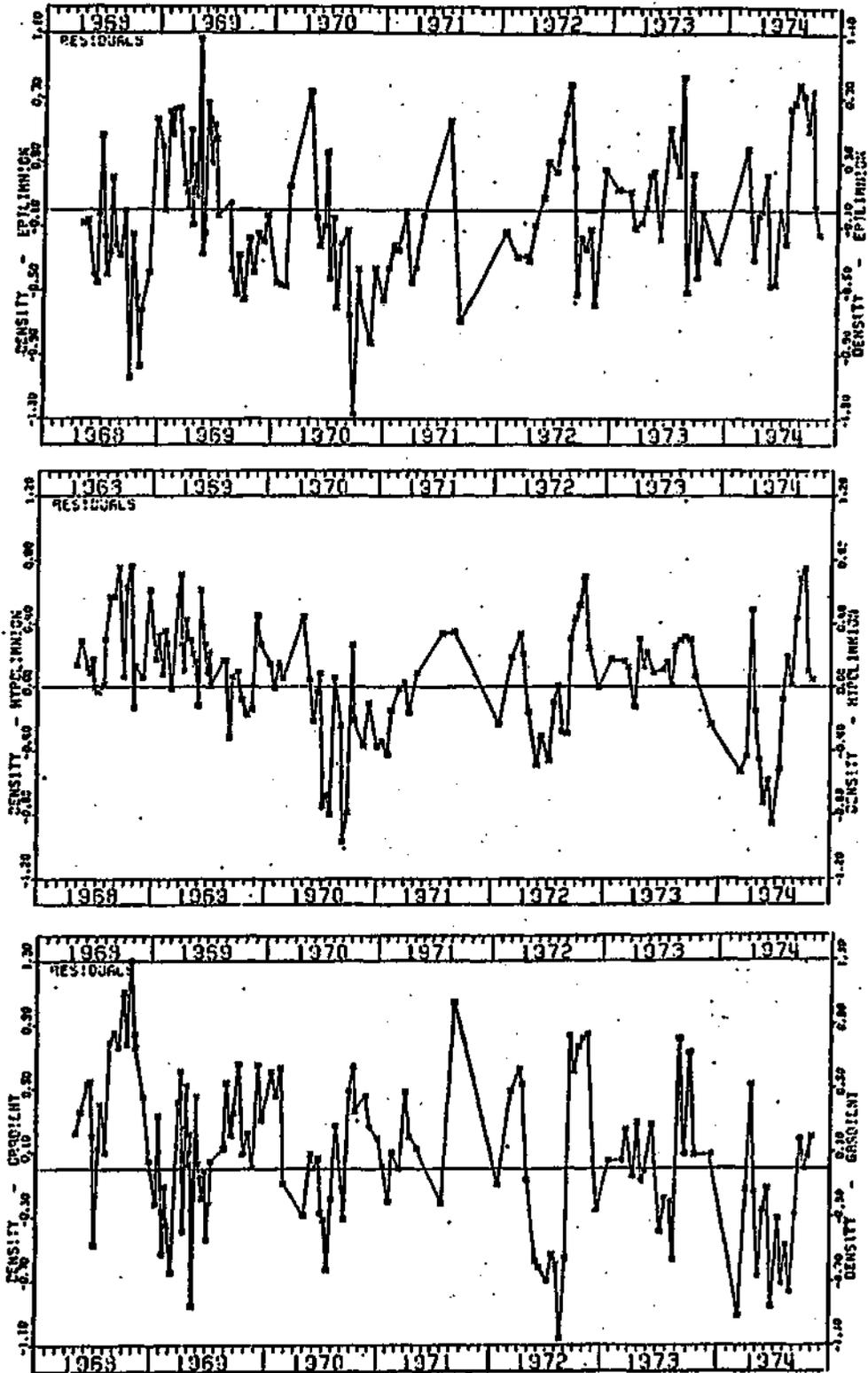
Figure 4.5-9

Observations Verses Predictions
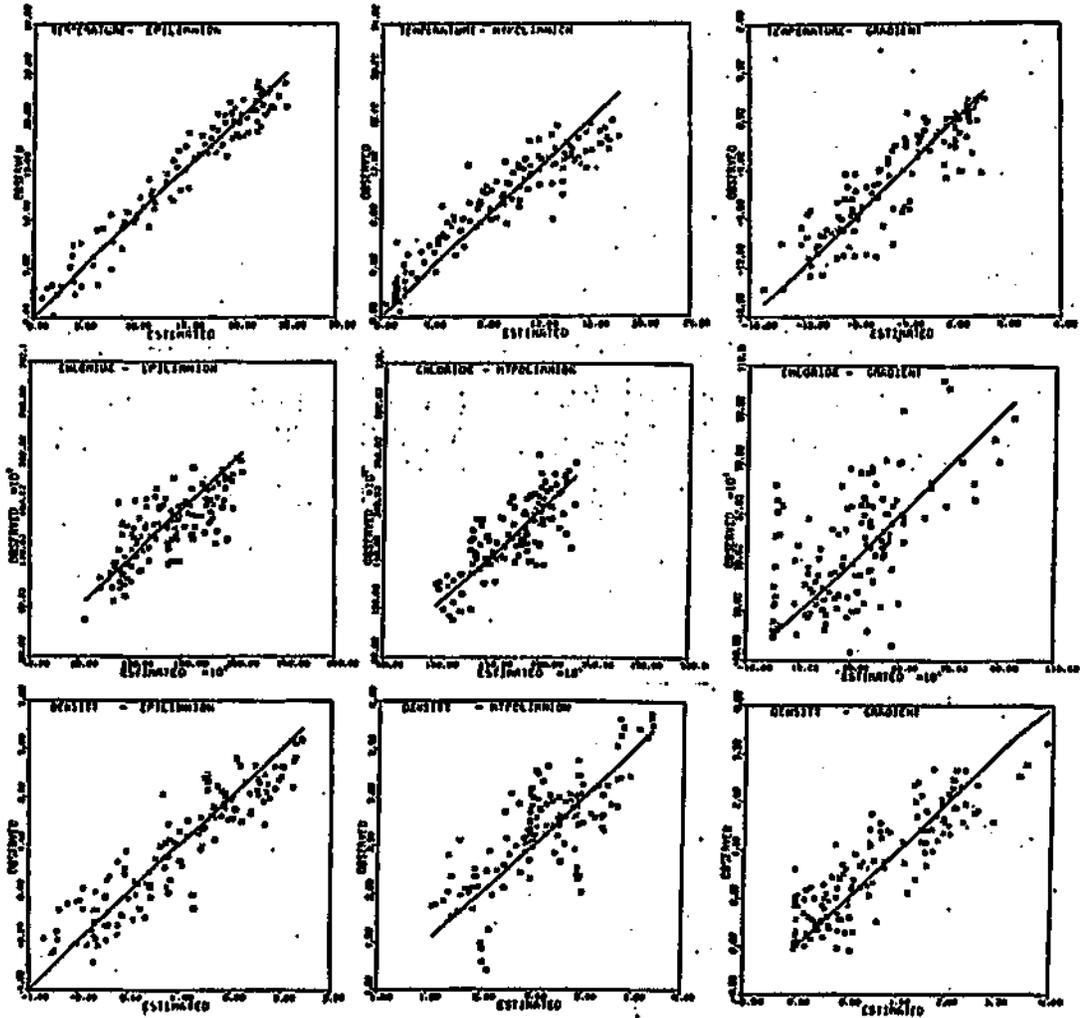
Figure 4.5-10

Histograms of Residuals
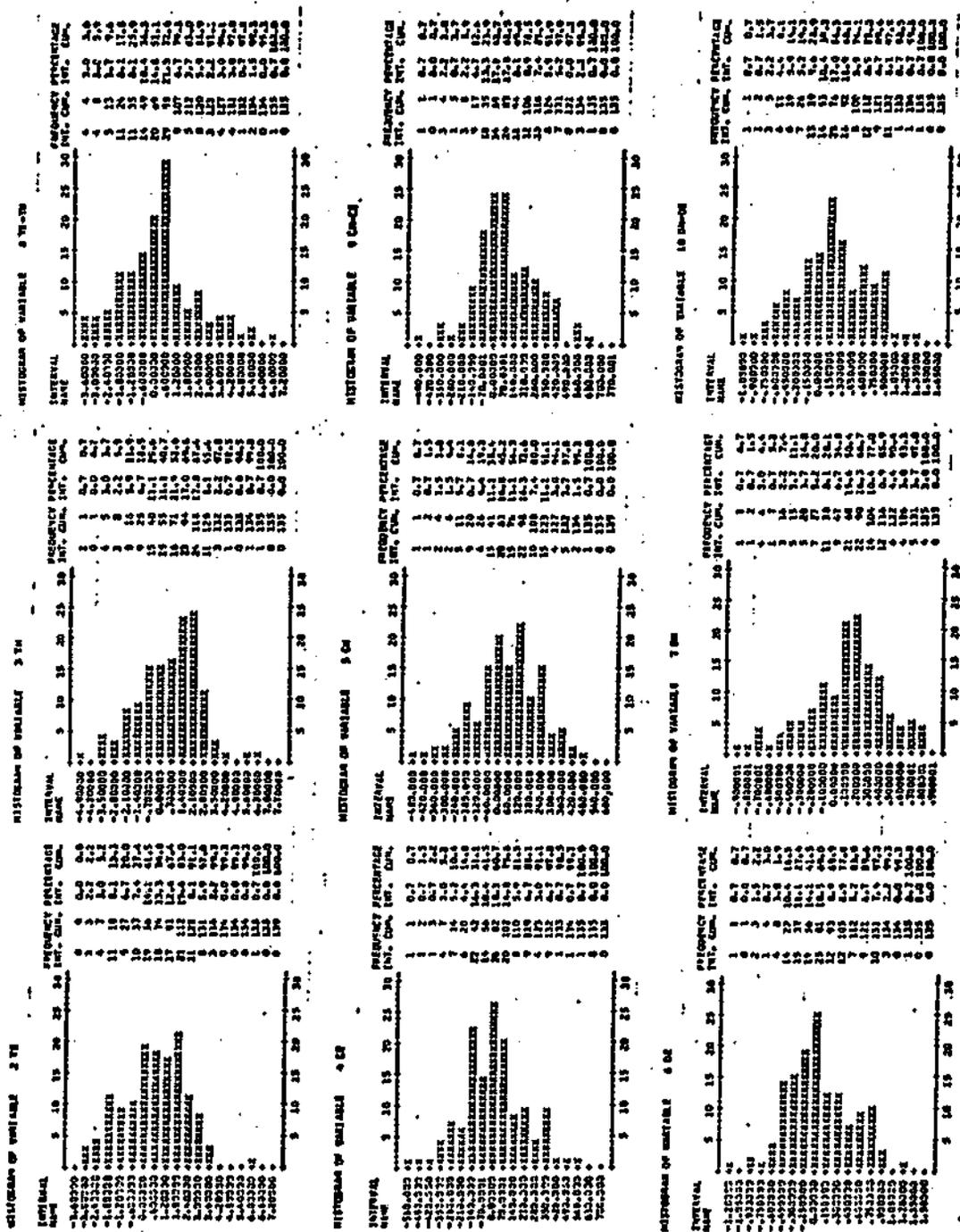
Figure 4.5-11

Normal Probability Plots of Residuals
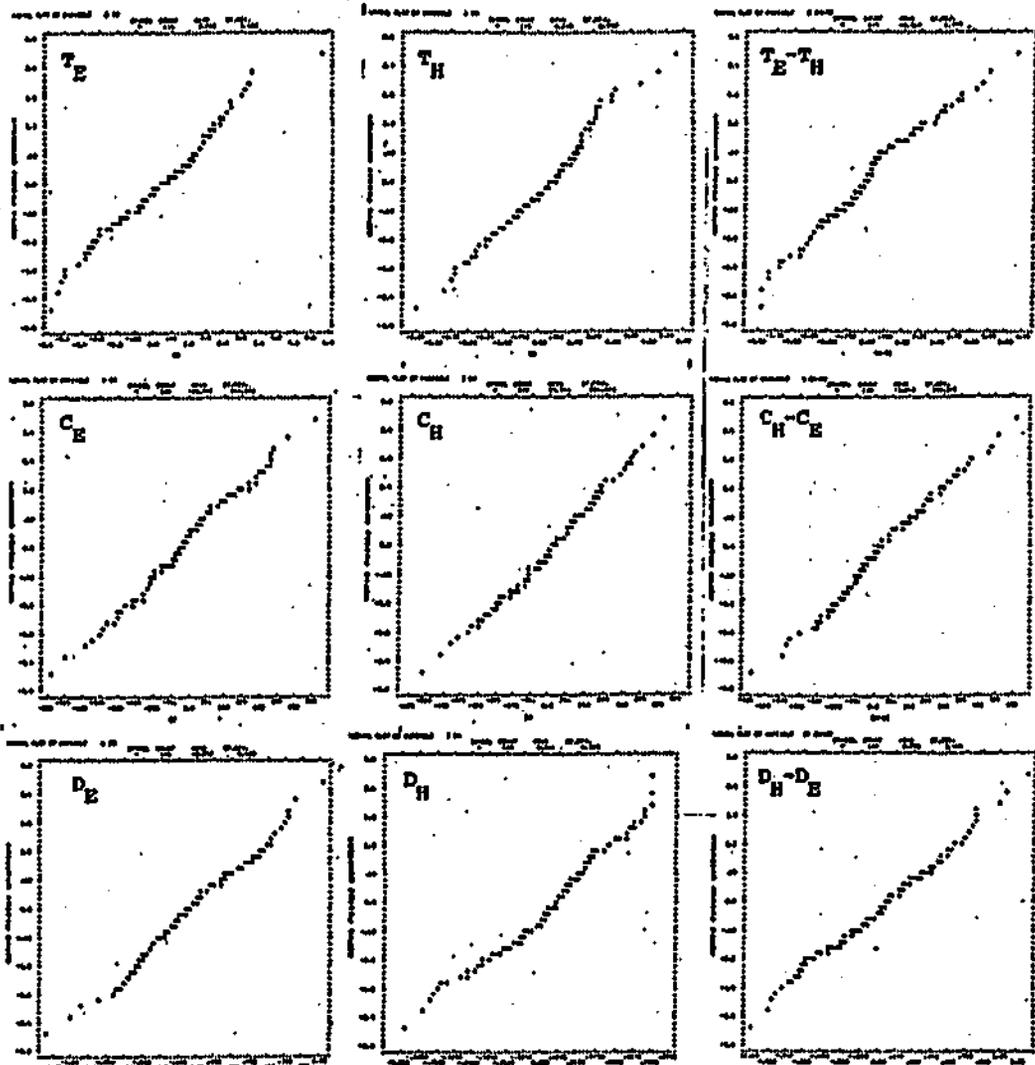
4.5.6 <u>Analysis of Residuals</u> (continued)

only in the case of temperature gradient (line h). If the residuals are assumed to have means of zero and standard deviations corresponding to the computed standard errors in line c, the hypothesis of normality is rejected in cases of five out of the nine variables (line i). It should be noted that, since the mean and standard deviation have been calculated directly from the sample, the power of the Kilmogorov-Smirnov test to detect non-normality is somewhat hindered. As a result, the computed probabilities of normality are biased upwards. Histograms (Figure 4.5-10) and normal probability plots (Figure 4.5-11) support the conclusion that the temperature gradient residuals are the least normal of any of the variables examined. Their distribution appears to be skewed toward positive values. This could be a consequence of the approximate lower limit of zero in this variable, since observed hypolimnion temperatures rarely exceed those of the epilimnion. Some of the probability plots tend to curve downwards at the low tails and upwards at the high tails, indicating a tendency for the tails to be lighter than normal, or for the distributions to be on the uniform side of normal.

The residuals have also been examined for seasonal and yearly patterns. A total of ten seasonal groups have been formed by aggregating at increments of 0.1 years. For each type of grouping, an analysis of variance has been done to determine the statistical signficance of variations of the means of the residuals between groups. The confidence levels for the statistics computed from these analyses

4.5.6 Analysis of Residuals (continued)

are only approximate because of the serial correlation within the groups. These analyses have been done chiefly to characterize general features of the residual patterns. It should also be noted that the yearly and seasonal groupings are not completely independent of each other, since sampling strategy shifted in the later years of the lake survey. During that period, sampling crews were apparently less eager to venture out onto the lake ice, so that cold-season samples are relatively scarce in the later years (Table 4.5-2). Hence, apparent yearly effects may be partially seasonal effects and vice-versa.

Seasonal- and yearly-mean residuals for each variable are plotted in Figures 4.5-12 and 4.5-13, respectively. F statistics in these figures and in Table 4.5-6 indicate significant variations between yearly groups in seven out of the nine variables and between seasonal groups in eight out of the nine variables. The variance fractions indicate that a maximum of 20% of the residual variance (in the case of D_H) can be attributed to year-to-year variations, while up to 50% can be attributed to seasonal variations (T_H , $T_H - T_E$).

Generally, seasonal effects on temperature residuals appear to be the strongest of the relationships examined and also appear to be quite periodic. Time series plots of temperature residuals support this (Figure 4.5-6). Epilimnion temperature is generally over-predicted by about 2°C during the cold seasons and under-predicted by about 1°C during the warm seasons. Periodicity in the hypolimnion and gradient

Figure 4.5-12

Mean Residuals Versus Time of Year

Figure 4.5-13

Mean Residuals Versus Year
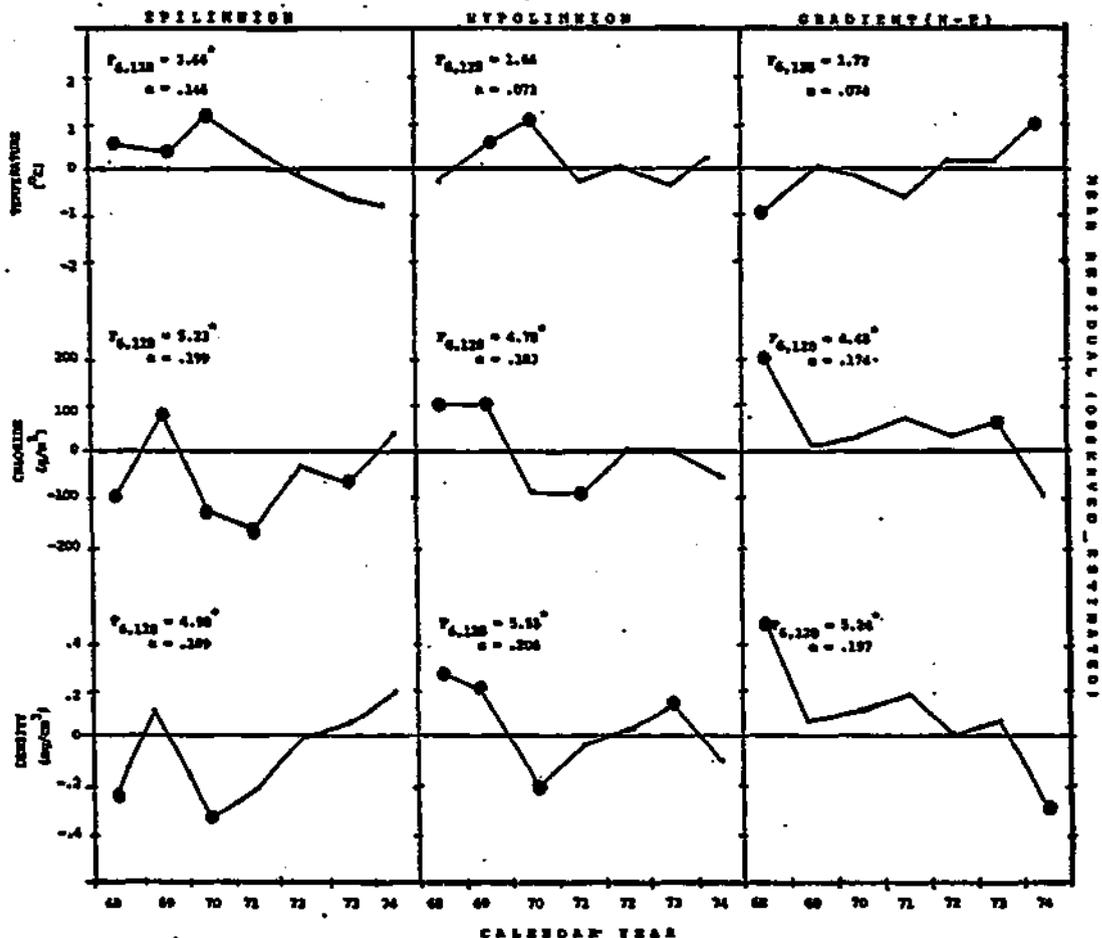
4.5.6 Analysis of Residuals (continued)

residuals may reflect the observed periodicity in the epilimnion

residuals, with an apparent lag time of about 0.3 years. As will be

shown, the epilimnion temperatures are sensitive only to a_5, the surface

mass transfer parameter, and not sensitive to any of the other estimated

parameters dealing with exchange between the hypolimnion and epilimnion.

Average epilimnic temperatures are fixed by the overall heat balance

on the lake.

The periodicity in the epilimnion temperature residuals can be

attributed to two factors. First, during cold seasons, a minimum

temperature of $0^{\circ}C$ has been arbitrarily specified for the epilimnion.

In actual fact, observed temperatures rarely go below $2^{\circ}C$, even under

ice cover. It is apparent that if this minimum temperature were

specified at $2^{\circ}C$ instead of $0^{\circ}C$, much of the periodicity in the

residuals would be removed. Secondly, the function employed for

calculation of solar radiation (U_9 in Table 4.3-2) may have been

inadequate. This function includes only a single cosine curve.

A recent article by Thompson[36] outlines a general method for estimating

solar radiation from sky cover for locations in the United States. In

this scheme, the function used to compute clear-sky radiation involves

five periodic components. Comparisons of the predictions of the two

methods using Syracuse meteorologic data indicate that the method

employed in the model tends to under-predict radiation in cold seasons

and to over-predict it in warm seasons. An examination of the residuals

obtained in estimating U_9 (see Table 4.3-2) has revealed similar

4.5.6 <u>Analysis of Residuals</u> (continued)

patterns, despite a high R^2 of 0.970. Thus, some of the seasonal periodicity in the temperature residuals may have been due to improper specification of the boundary conditions. Despite these problems, the model explains 95% of the variance in the epilimnion temperature data. The apparent decreasing trend in the yearly-average epilimnion temperature residuals (Figure 4.5-15) may be due to decreasing sampling frequencies during cold seasons, as discussed above. Because of these problems, the estimate of the surface mass transfer coefficient, a_5 , may be somewhat biased.

Yearly effects on epilimnic chloride residuals appear to be substantial (Figure 4.5-13). Again, the surface conditions are not very sensitive to the mixing mechanisms or parameter values. In fact, average surface chloride concentrations are essentially fixed by the hydrologic and salt-influx boundary conditions, due to mass balance constraints. Thus, variations in the mean yearly surface chloride residuals can be attributed to errors in specifications of the hydrology or salt loadings.

Patterns in the density residuals reflect those in the temperature and chloride residuals. An under-prediction of the density gradient is evident during cold seasons (Figure 4.5-12), particularly during the seasons 0.7 and 0.8. While some of this may be due to the temperature problems discussed above, model errors could also be involved. One aspect of the model which could be important here is the

4.5.6 <u>Analysis of Residuals</u> (continued)

assumption of constant epilimnion and hypolimnion volumes. Erosion
of the thermocline effectively increases epilimnion depth and volume
in the late seasons prior to fall overturn. This effectively increases
the volume over which the kinetic energy due to wind shear stress is
dissipated. With an increase in depth (a_{17}), the Richardson Number
(F_{12}, Table 4.3-3) would increase, the vertical exchange rate (F_{11})
would decrease, and the computed density gradient would, in turn,
increase. Thus, if effects of a migrating thermocline were incorporated
into the model, the tendency to under-predict late summer and early
fall density gradients might be reduced.

Serial correlation in the residuals can be attributed to periodic
or auto-correlated errors in the specification of boundary conditions,
to the effects of factors not considered in the model, and to aspects
of the measurement process. Bard[3] notes that serial correlation in
the residuals is more generally the rule than the exception in cases of
dynamic models. These results are not grounds for rejection of the
model. The primary effect of the serial dependence of the errors is
upon the confidence regions of the estimated parameter values.

If the model were linear, and the residuals were uncorrelated,
unbiased, and normally-distributed, the confidence regions calculated
for the parameter estimates would apply exactly. In this case, however,
the model is nonlinear, and the residuals are auto-correlated, somewhat
biased, and, in some cases, not normally distributed. A standard

4.5.6 Analysis of Residuals (continued)

approach to determining confidence regions for parameters under such conditions would involve Monte-Carlo simulation. Such an approach is precluded in this case by the expense of implementing the parameter estimation routine. Thus, in applying the model and in assessing errors in model projections, the approximate nature of the estimate of the parameter covariance matrix must be considered.

4.6 Model Applications

4.6.1 Sensitivity Analysis

Before applying the model to assess the potential effects of the sewage outfall design upon general aspects of vertical mixing in Onondaga Lake, an analysis has been performed to estimate the sensitivity of the model simulations to some of the parameters and forcing functions. The parameters studied have included a_{14} (the numerator of the vertical exchange rate function, F_{11}), a_5 (the surface mass transfer parameter for evaporation and conduction), and a_{22} (the maximum fraction of Ninemile Creek flow entering the hypolimnion). Sensitivities to U_6 (the chloride flux from Allied Chemical's waste beds) and F_4 (the Allied cooling water flow) have also been evaluated. The sensitivity to a_{14} can also be used as a measure of the sensitivity to the average rate of wind-induced kinetic energy input to the lake. In the numerator of F_{11}, the exchange rate function, a_{14} and F_{13}^3 occur as a product. F_{13}^3, the cube of the friction velocity, is proportional to the rate of kinetic energy input, according to equation 4.5-17. The purpose of the analysis is to assess the relative importance of various mechanisms in controlling lake mixing under past and present conditions.

4.6.1 Sensitivity Analysis (continued)

The period of record (1968-74) has been simulated using the best
estimates of the parameters developed in Section 4.5 and recording
results at 0.05-year intervals. Normalized sensitivity coefficients
have been computed as a function of time for each factor studied, using
the finite-difference methods discussed in Section 4.4. Sensitivity
coefficients are essentially normalized first derivatives of the state
variables with respect to the factor values:

$$S_{i,j,t} = P_i \frac{\partial y_{j,t}}{\partial P_i} \qquad (4.6-1)$$

where,

$$S_{i,j,t} = \text{sensitivity coefficient for state variable } y_j \text{ at time } t \text{ and factor } P_i$$

$$P_1 = a_{14}$$
$$P_2 = a_5$$
$$P_3 = a_{22}$$
$$P_4 = U_6$$
$$P_5 = F_4$$

The coefficients can be interpreted as the marginal changes in the
state variables which would result from fractional changes in the
factors. The normalization adjusts for differences in factor scales.
Sensitivity coefficients for temperature, chloride and density are
plotted in Figures 4.6-1 to 4.6-3 respectively. The numbers labelling
the curves in each plot correspond to the subscripts of the P_i
variables defined above.

Figure 4.6-1

Temperature Sensitivity Coefficients Versus Time
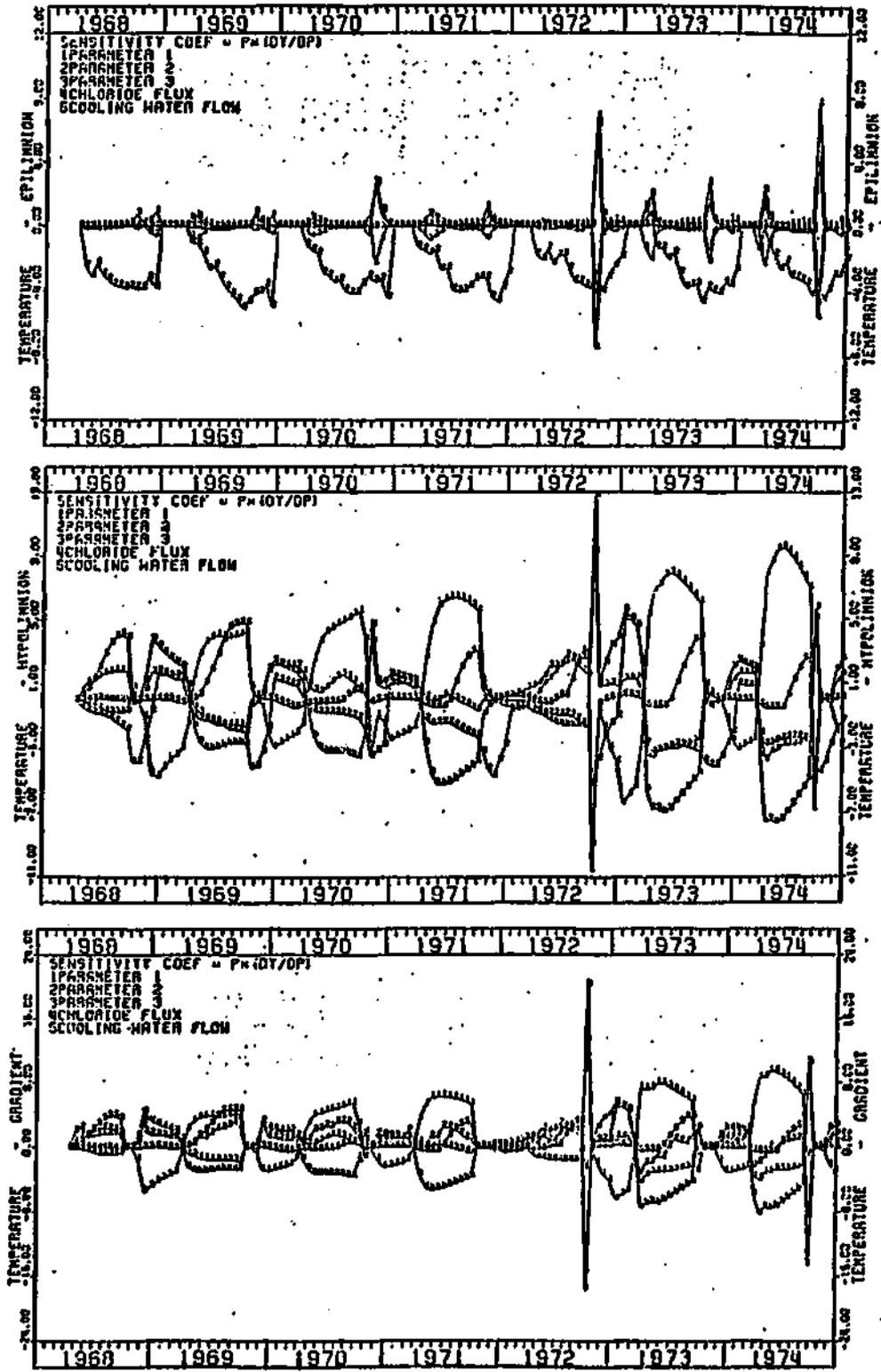
Figure 4.6-2

Chloride Sensitivity Coefficients Versus Time
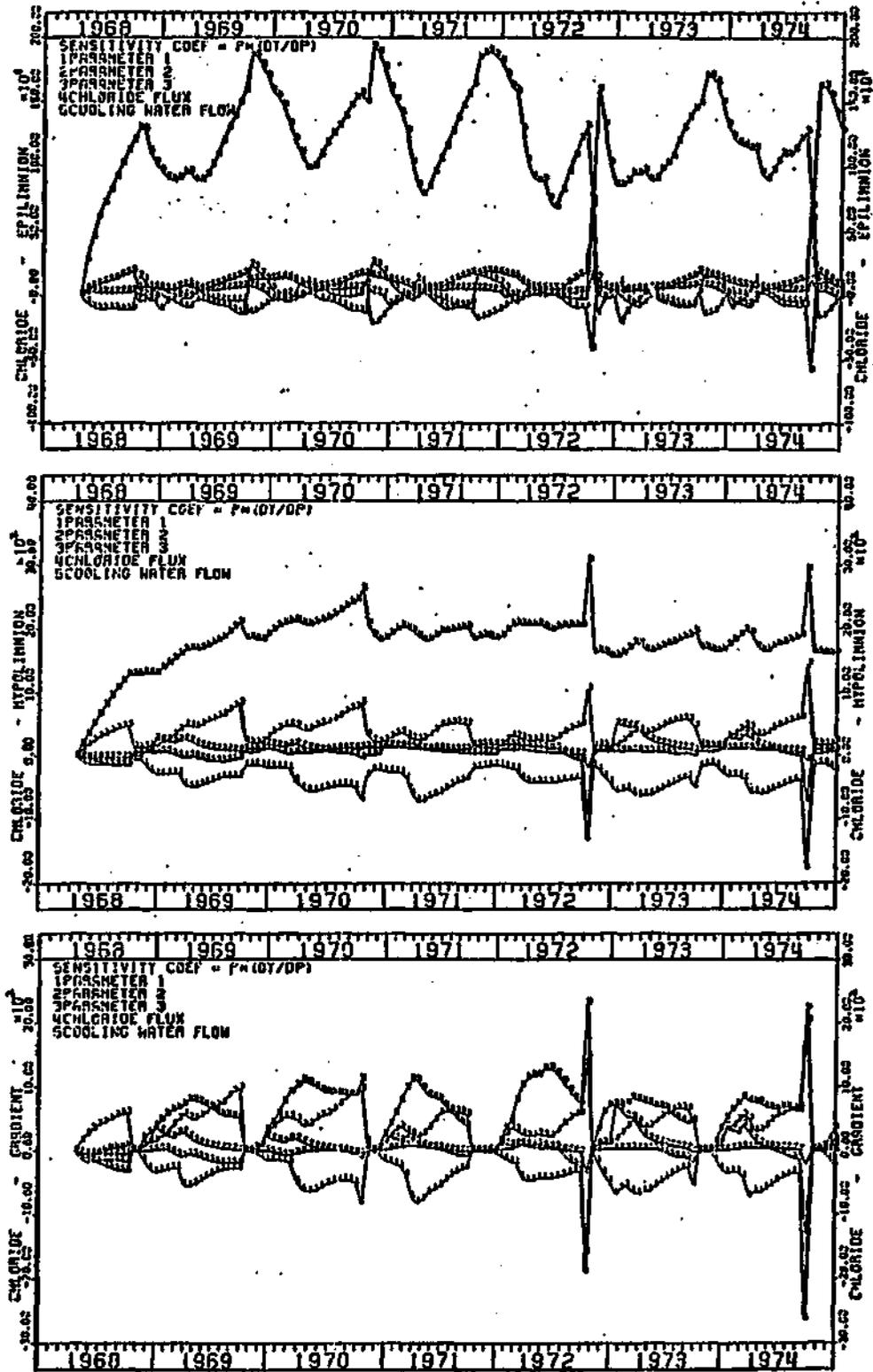
Figure 4.6-3

Density Sensitivity Coefficients Versus Time
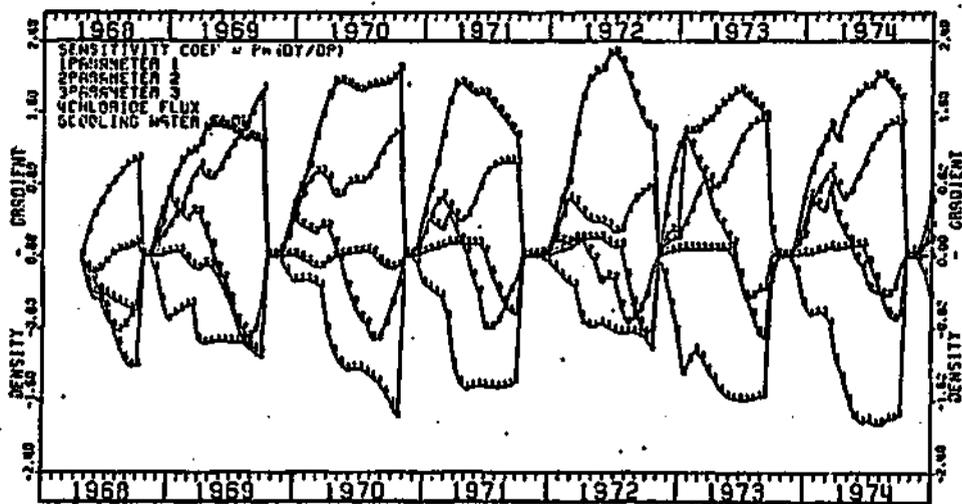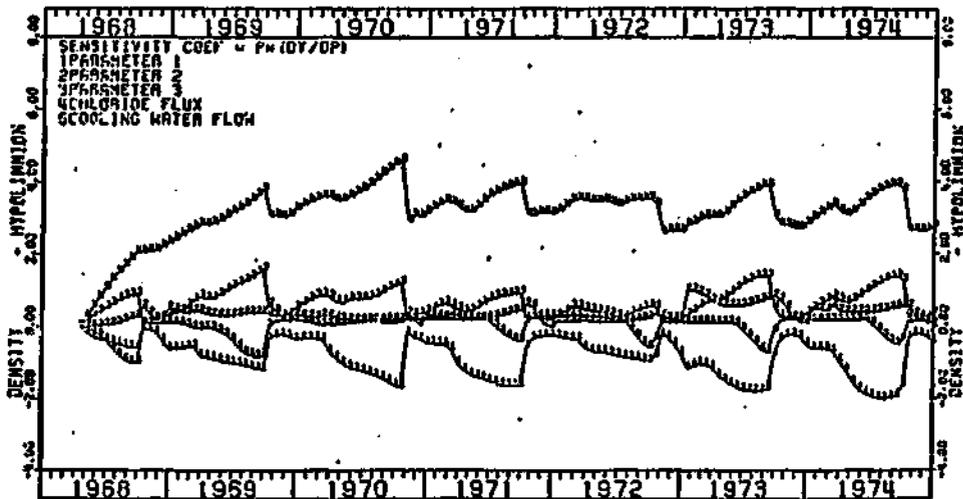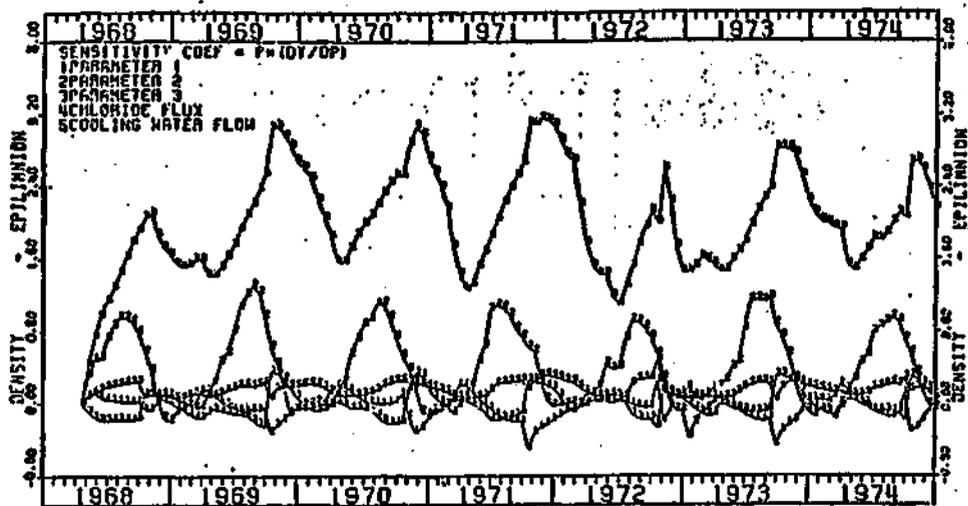
4.6.1 <u>Sensitivity Analysis</u> (continued)

One general aspect of the temperature and chloride sensitivies
is the tendency for "spikes" to occur during overturn periods, parti-
cularly in late 1972 and 1974. This is essentially because changes
in the factors influence the times at which density equivalence is
reached and rapid mixing occurs. Just prior to overturn, both
temperature and chloride gradients may exist, but with opposing
effects on the density gradient, which is approaching zero. When
mixing occurs, temperature and chloride variables change rapidly,
giving rise to the spikes in the sensitivity plots.

Epilimnion temperatures appear to be insensitive to all factors
examined, except for the surface transfer parameter, P_2 . As
vertical stratification develops during each season, sensitivities of
hypolimnion temperature with respect to the dominant factors, P_1 and
P_4 , increase in absolute values and are opposite in sign. Temperature
gradients (hypolimnion - epilimnion) are also chiefly controlled by these
factors. Chloride sensitivities (Figure 4.6-2) indicate the dominance
of P_4 , the Allied Chemical chloride flux, in both the epilimnion and
hypolimnion. Chloride gradients are controlled chiefly by P_1 and P_4 .
Epilimnion density is controlled by P_4 and P_2 , while hypolimnion
density and density gradient both respond most dramatically to changes
in P_4 and P_1 during most years.

The density gradient sensitivities are particularly important and
warrant more detailed analysis. Generally, P_1 and P_4 appear to be
controlling, as in the cases of temperature and chloride gradients.

4.6.1 Sensitivity Analysis (continued)

During peak stratification seasons, p_1 sensitivity coefficients are
on the order of -1.5, while p_4 coefficients are approximately 2.0.
This means that a 1% increase in the chloride flux is estimated to have
about the same effect on peak density gradients in the lake as a
2/1.5 = 1.33% decrease in a_{14} , or, equivalently, a 1.33% decrease
in the average rate of kinetic energy input. Since the latter depends
upon the 3.75 power of wind speed, this would correspond to a 0.35%
decrease in average wind speed over the lake. This helps to quantify
the stratifying effect of the industrial discharge relative to the
natural forces tending to mix the lake. Because of the nonlinear
nature of the model, these figures are valid only for small changes
in the factors.

Cooling water flow, p_5 , has a negative impact on a density
gradient during late summer seasons, except during 1970. In that
year, cooling water was consistently withdrawn from the epilimnion
(see U_4 , Table 4.3-2). Thus, this factor was not as important a mixing
mechanism in the lake during 1970 as it was during other years, in which
the cooling water was withdrawn from the hypolimnion.

In 1968, the sensitivity equations are markedly different from
those observed during other years, in that factors 3 and 5 appear to be
controlling peak density gradients. The reasons for this are unclear,
but may be related to the fact that the simulation was started after
the onset of stratification in the spring. This fact, coupled with
some of the seasonal biases in the model discussed in Section 4.5-6,
may account for the differences. During this period, the sensitivity

4.6.1 Sensitivity Analysis (continued)

coefficients are not in equilibrium with the factors driving the model, but are still heavily influenced by initial conditions. The qualitative differences in the sensitivity coefficients during 1968 may in part account for the differences in optimal parameters estimated for that year in Section 4.5.

The degrees and types of mixing in the lake over the study period are characterized in Figure 4.6-4. In units of $year^{-1}$, the total hypolimnic dilution rate (HDR) is defined as the total outflow rate from the hypolimnion divided by the hypolimnion volume:

$$HDR_T = \frac{Q_{16} + Q_{28} + Q_{30} + Q_{32}}{V_H} \times 365. \qquad (4.6-2)$$

The non-advective portion of this is attributed to the density-dependent exchange between the epilimnion and hypolimnion, F_{11}, and is given by:

$$HDR_{na} = \frac{Q_{28}}{V_H} \times 365. \qquad (4.6-3)$$

These variables are particuarly important as measures of the oxygen resources of the hypolimnion. The logarithms of these two variables are plotted against time in Figure 4.6-4. During fall overturn periods, maxima are reached in the vicinity of 2.6. These upper limits are

Figure 4.6-4

Simulated Total and Non-Advective Hypolimnic Dilution Rates
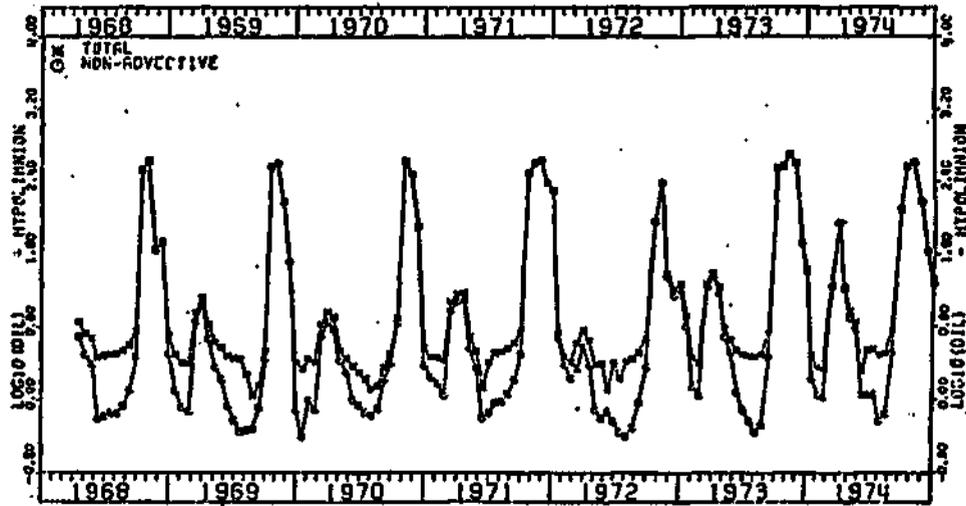
Figure 4.6-5

GRID Display of Dissolved Oxygen
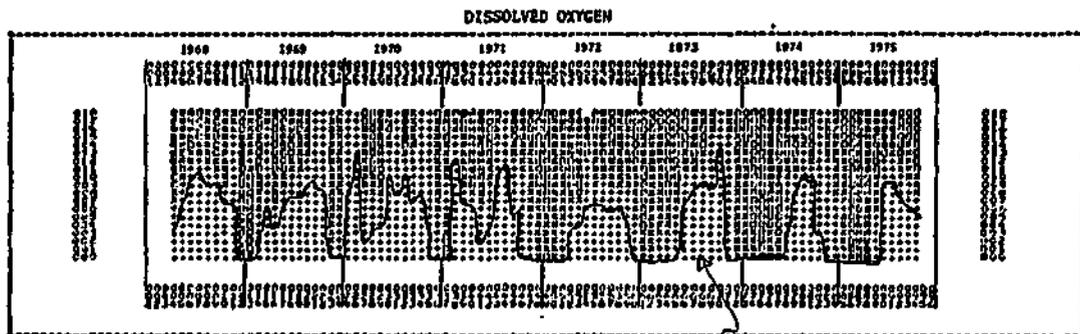
D.O. < 1 mg/liter
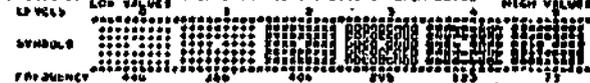
4.6.1 <u>Sensitivity Analysis</u> (continued)

determined by the assumed value of a_{16}, which is somewhat arbitrary.
Accordingly these periods should be interpreted as periods in which
the lake is essentially completely-mixed.

During periods of peak stratification, the advective mixing
mechanisms become important. If cooling water is withdrawn exclusively
from the hypolimnion, the advective component of the HDR amounts to
2.24 year^{-1}, or 0.35 on a logarithmic scale. The non-advective
component generally reaches a minimum in the vicinity of -0.2 on a
log scale, or about one quarter of the advective cooling-water component.
During low flow seasons, Ninemile Creek flow amounts to about 4 m^3/sec.
With a value of 0.267 for a_{22}, this corresponds to an HDR of 0.63 year^{-1},
or -0.2 on a log scale, about equivalent to the non-advective mixing
component. However, this does not contribute to the total HDR, because
it merely satisfies some of the demand imposed by cooling water withdrawal.
Thus, during peak stratification, cooling water use accounts for about
80% of the total hypolimnic displacement, if it is withdrawn exclusively
from the hypolimnion. When the epilimnion intake is opened, it is
estimated that 78% of the water entering the plant comes from the surface
waters. Under these conditions, at peak stratification, the total dis-
placement rate is reduced by about 60%.

The simulations indicate that mixing during spring overturn periods
is not as complete or as rapid as in the fall. This is chiefly due to
the chloride gradients which are built up under ice cover during the

4.6.1 <u>Sensitivity Analysis</u> (continued)

winter. An increasing trend in the intensity of the spring mixing
period is evident. Computed total dilution rates are higher by about
an order of magnitude during the spring of 1974, as compared with the
springs of 1969 and 1970. This correlates well with corresponding
increases in hypolimnic dissolved oxygen levels noted in Chapter 3
and shown again in Figure 4.6-5. Comparisons of patterns in mixing
rate with periods of hypolimnic anaerobiosis indicates that the latter
correspond roughly to periods in which the logarithm of the total
dilution rate is less than about 0.8. The lengths of these periods
from spring overturn to fall overturn, estimated from Figure 4.6-4,
range from about 0.3 years in 1974 to about 0.5 years in 1970. These
factors suggest that the observed enhancement in hypolimnic dissolved
oxygen levels in later years of the survey may have been due to climatologic
influences on lake mixing, as simulated by the model developed here. More
specific evidence for this could be developed by increasing the complexity
of the model to permit direct simulation of dissolved oxygen. This will
be discussed further in Section 4.7.

4.6.2 Case Simulations

A total of thirteen cases have been formulated in order to permit further analysis of the impacts of coding water use, salt discharge, and Metro STP outfall design on general aspects of lake mixing. Case conditions are summarized in Table 4.6-1. The first three examine the effect of cooling water withdrawal under present conditions of salt discharge to Ninemile Creek. Cooling water is withdrawn exclusively from the hypolimnion in Case 1, primarily from the epilimnion in Case 2, and is totally absent in Case 3. The next three cases examine the same cooling water conditions in the absence of the salt discharge. Case 6 can be considered as a baseline condition or undisturbed state. The last seven cases examine the effect of increasing initial dilution ratio for discharge of the combined municipal/industrial effluent into the epilimnion. Case 7, with a dilution ratio of 0., is essentially equivalent to discharge of the effluent into the hypolimnion, due to density current effects. Initial dilution ratios increase by factors of two from a value of 1 for Case 8 to a value of 16 for Case 12. In Case 13, the discharge has been prevented from sinking into the hypolimnion by setting a_{21} equal to zero. This is essentially equivalent to an infinite initial dilution ratio. All of the last seven cases have assumed exclusive cooling water withdrawal from the hypolimnion, which has been Allied Chemical's operating policy since mid-1974[7].

One of the factors which introduces uncertainty into the analysis of the impact of the initial dilution ratio is uncertainty in parameter

Table 4.6-1

Definitions of Cases Studied

	COOLING WATER			SALT				COMMENTS		
Case	Flow	Withdrawl Epil.[c]	Level Hypo.	Load	Stream	Initial Dilution Ratio	Max. Fraction Sinking Into Hypolim.	Cooling Water Withdrawl	Salt	
1	F_4	0.00	1.00	U_6	Q_9	$a_3 = 0.$	$a_{22} = 0.266$	Hypolimnion	Present Conditions	
2	F_4	0.78	0.22	"	"	"	"	Epilimnion	"	"
3	0.	–	–	"	"	"	"	Absent	"	"
4	F_4	0.00	1.00	0.	–	"	$a_{22} = 0.0$	Hypolimnion	Absent	
5	F_4	0.78	0.22	"	–	"	"	Epilimnion	"	
6	0.	–	–	"	–	"	"	Absent	"	
7[d]	F_4	0.00	1.00	U_6	Q_5	$a_2 = 0.$	$a_{21} = 1.0$	Hypolimnion	Metro STP ;	Initial Dil. d = 0.
8	"	"	"	"	"	1.	"	"	"	d = 1
9	"	"	"	"	"	2.	"	"	"	d = 2
10	"	"	"	"	"	4.	"	"	"	d = 4
11	"	"	"	"	"	8.	"	"	"	d = 8
12	"	"	"	"	"	16.	"	"	"	d = 16
13	"	"	"	"	"	–	$a_{21} = 0.0^b$	"	"	d = ∞

a - see Table 4.3-1,-2,-3, and -4 for definitions of Q, U, F, and a variables, respectively.

b - setting a_{21} equal to zero forces all of the influent stream to remain in the epilimnion and is equivalent to an infinite initial dilution ratio, a_2.

c - Fraction of cooling water withdrawn from epilimnion = U_4

d - Essentially equivalent to discharging into the hypolimnion, due to density current effects.

4.6.2 Case Simulations (continued)

a_{21}, the maximum fraction of the diluted effluent which is allowed to sink into the hypolimnion, subject to density constraints (see F_2, Table 4.5-3). A value of 1.0 has been assumed for this parameter in Cases 7 - 12. The analogous coefficient for Ninemile Creek, a_{22}, has been estimated empirically at 0.267. It would be difficult to predict these parameters a priori without more detailed modelling of the hydrodynamics of the outfall site, including specifics of wind velocities (speeds and directions), horizontal currents, and bottom topography. However, Cases 7 and 13 represent extreme conditions, which are insensitive to the value of a_{21}. Case 7 essentially represents disposal in the hypolimnion, while Case 13 represents infinite initial dilution in the epilimnion. Thus, while uncertainty in a_{21} may introduce uncertainty as to the specific path from Case 7 to Case 13 as dilution ratios increase, the end points are not influenced.

Simulation of each of the cases has been done under the hydrologic and meteorologic conditions from October, 1967 through December, 1974. Metro STP flow (U_2) has been assumed constant at the design value of 3.78 m^3/sec. Results have been recorded at 0.05-year increments. The first 15 months of each simulation have not been analyzed because of sensitivity to assumed initial conditions. Thus, the cases have been compared based upon model simulations over the six-year period from 1969 through 1974. In Cases 1 - 3, sensitivity coefficients to each of the four empirically-estimated parameters (a_{14}, a_{15}, a_5, and a_{22}) have also been calculated. In Cases 4 - 13, a_{22} is of no consequence

4.6.2 Case Simulations (continued)

and has not been included in the sensitivity calculations. The primary
incentive for computing sensitivity coefficients is to permit estimation
of the effects of uncertainty in the parameter estimates on errors in
model projections.

Four primary "objective functions" have been used to compare lake
responses to the conditions specified by the various cases: (1) mean
density gradient; (2) maximum annual density gradient; (3) mean hypolimnic
dilution rate (HDR); (4) minimum annual HDR. A final supplementary
criterion is average turnover frequency, defined as the average number
of periods per year of simulation in which the Richardson Number, F_{12},
drops below 100. The Richardson Number is a measure of the resistance
to vertical mixing. The vertical exchange rate, F_{11}, is roughly
inverse to its value, which ranges from zero during unstable periods
to about 5×10^4 during peak density stratification. A value of 100
has been arbitrarily employed as a turnover definition. This corresponds
to an HDR value of 123 year^{-1}, or a complete exchange of hypolimnic waters
in about 3 days.

For the first and third objective functions specified above, a total
of three factors has been assumed to contribute to variations within
each case: season, year, and parameter values. Seasonal variations
in the functions are attributed to annual meteorologic and hydrologic
cycles. Yearly variations represent the effects of variations in average
annual meteorologic and hydrologic conditions. To provide a basis for

4.6.2 <u>Case Simulations</u> (continued)

assessing the severity of the effects of parameter uncertainty on model

projections, the seasonal and yearly variations have been compared with

those possibly attributed to parameter variations. To determine the

latter for each case, estimates of objective function values at the end

points of the parametric principle component axes (Table 4.5-3) have been

derived from the value at the center of the parameter confidence region,

the sensitivity coefficient matrix, and the vector of parameter values

at the end points of each principal component axis. For a given state

variable y_j at time t :

$$y_{j,t,k} \simeq y_{j,t,o} + \sum_{i=1}^{4} \left(\frac{\partial y_{j,t}}{\partial \theta_i}\right)(\theta_{i,k} - \theta_{i,o}) \qquad (4.6-4)$$

where,

$y_{j,t,k}$ = value of state variable j at time t and
parameter vector k

$y_{j,t,o}$ = value of state variable j at time t and
optimal parameter vector $\underline{\theta}_{i,o}$

$\dfrac{\partial y_{j,t}}{\partial \theta_i}$ = sensitivity coefficient of state variable y_j
with respect to parameter θ_i at time t

$\theta_{i,k}$ = value of parameter θ_i at principal component
axis end point k

$\theta_{i,o}$ = optimal value of θ

4.6.2 Case Simulations (continued)

This scheme has been employed to develop a three-dimensional array of values for the density gradient and $\log_{10}$(HDR) for each case. Each array consists of 20 seasons, 6 years, and 9 parameter sets. The ranges of the means within each of the three factor groups have been used as bases to assess the relative effects of the factors on mean density gradient and on $\log_{10}$(HDR). In analysis of yearly extreme values (maximum density gradient and minimum HDR), the extremes for each year and case have been extracted from the general, three-dimensional array to generate a two-dimensional matrix (6 years x 9 parameter sets) for each case.

Results are presented in Table 4.6-2 and in Figures 4.6-6 to 4.6-9. Simulations of Cases 6, 7, and 13 are presented in Figures 4.6-11 to 4.6-13, respectively. These can be compared with the simulation of actual 1968-74 conditions in Figure 4.6-10.

Comparing the results for Cases 1, 2, and 3 with Cases 4, 5, and 6 provides a basis for evaluating the effects of the salt discharge to Ninemile Creek. Average turnover frequencies for Cases 1 - 3 range from 0.50 to 1.00, while 2 turnovers per year are consistently observed in the absence of the salt discharge. The primary effect of the discharge is to inhibit the spring turnover period. The difference in average density gradient between the two groups of cases is 0.68 mg/cm^3, corresponding to 62% of the Case 1 - 3 mean. Thus, if the salt discharge were eliminated, it is indicated that average density gradients in the

Table 4.6-2

Results of Case Simulations

CASE[a]

COMPONENT		1	2	3	4	5	6	7	8	9	10	11	12	13
Turnover Frequency[b]		1.00	0.83	0.50	2.00	2.00	2.00	0.00	0.00	0.50	0.83	1.00	1.83	1.83
Density Gradient[c]	Mean	0.96	1.17	1.21	0.35	0.49	0.44	5.15	2.71	1.84	1.13	0.64	0.41	0.40
Season[e]	Minimum	0.00	0.07	0.15	0.00	0.00	0.00	3.25	0.90	0.23	0.01	0.00	0.00	0.00
	Maximum	2.55	2.88	2.81	1.49	1.89	1.72	7.18	4.31	3.10	2.02	1.24	0.91	1.42
	Range	2.55	2.81	2.67	1.52	1.89	1.72	3.93	3.41	2.87	2.01	1.24	0.91	1.42
Year[e]	Minimum	0.71	0.86	0.86	0.25	0.34	0.30	4.89	2.52	1.70	1.03	0.57	0.30	0.22
	Maximum	1.36	1.63	1.58	0.44	0.62	0.56	5.47	3.03	2.10	1.31	0.75	0.72	0.90
	Range	0.65	0.77	0.73	0.19	0.28	0.26	0.57	0.50	0.40	0.29	0.18	0.42	0.68
Parameter[e]	Minimum	0.90	1.08	1.05	0.32	0.45	0.41	4.98	2.64	1.80	1.10	0.62	0.36	0.37
	Maximum	1.03	1.27	1.37	0.37	0.53	0.47	5.32	2.78	1.88	1.14	0.64	0.45	0.44
	Range	0.13	0.19	0.31	0.05	0.08	0.07	0.34	0.14	0.09	0.04	0.02	0.09	0.07
Annual Maximum Density Gradient[c]	Mean	2.57	2.92	2.87	1.51	1.92	1.74	7.22	4.37	3.16	2.04	1.24	1.02	1.47
Year	Minimum	2.00	2.37	1.32	1.10	1.44	1.30	6.71	4.07	2.94	1.85	1.08	0.61	1.04
	Maximum	3.52	3.81	3.62	1.88	2.34	2.16	8.17	5.11	3.70	2.32	1.37	2.16	2.82
	Range	1.52	1.44	1.30	0.77	0.90	0.86	1.46	1.04	0.76	0.48	0.29	1.56	1.78
Parameter	Minimum	2.45	2.80	2.74	1.42	1.80	1.65	7.02	4.25	3.08	1.99	1.22	0.99	1.36
	Maximum	2.69	3.05	3.00	1.60	2.04	1.84	7.43	4.48	3.24	2.08	1.26	1.06	1.58
	Range	0.24	0.26	0.26	0.18	0.23	0.20	0.42	0.23	0.17	0.10	0.05	0.08	0.22
Log_{10}(Hypolimnic Dilution Rate)[d]	Mean	1.04	0.79	0.70	1.70	1.44	1.45	0.49	0.80	1.03	1.30	1.58	1.73	1.49
Season	Minimum	0.44	0.06	0.04	0.55	-0.11	-0.07	0.41	0.71	0.88	1.10	1.34	1.24	0.56
	Maximum	2.60	2.21	1.91	2.67	2.66	2.67	0.61	1.03	1.84	2.43	2.41	2.56	2.63
	Range	2.16	2.16	1.86	2.12	2.78	2.74	0.20	0.31	0.96	1.33	1.07	1.32	2.07
Year	Minimum	0.86	0.53	0.49	1.54	1.24	1.24	0.46	0.77	0.95	1.23	1.51	1.41	1.14
	Maximum	1.19	1.01	0.97	1.88	1.67	1.67	0.51	0.85	1.12	1.41	1.66	1.84	1.73
	Range	0.33	0.48	0.48	0.33	0.43	0.44	0.05	0.08	0.16	0.18	0.15	0.44	0.59
Parameter	Minimum	0.97	0.70	0.61	1.64	1.38	1.32	0.47	0.79	1.02	1.28	1.57	1.63	1.40
	Maximum	1.10	0.88	0.79	1.77	1.49	1.57	0.50	0.82	1.04	1.32	1.59	1.82	1.58
	Range	0.14	0.18	0.19	0.13	0.11	0.25	0.03	0.03	0.03	0.03	0.02	0.19	0.18
Annual Minimum Log_{10}(HDR)[d]	Mean	0.39	0.04	0.00	0.52	-0.13	-0.08	0.40	0.70	0.87	1.09	1.33	1.21	0.52
Year	Minimum	0.28	-0.02	-0.08	0.48	-0.32	-0.28	0.39	0.69	0.86	1.08	1.33	0.48	0.41
	Maximum	0.44	0.13	0.11	0.60	0.09	0.13	0.41	0.71	0.88	1.09	1.34	1.59	0.56
	Range	0.15	0.15	0.19	0.12	0.40	0.41	0.02	0.02	0.02	0.01	0.02	1.11	0.16
Parameter	Minimum	0.37	0.00	-0.04	0.50	-0.21	-0.17	0.40	0.70	0.87	1.09	1.33	1.11	0.49
	Maximum	0.41	0.08	0.05	0.54	-0.04	0.00	0.41	0.70	0.87	1.09	1.33	1.31	0.54
	Range	0.04	0.07	0.09	0.05	0.17	0.17	0.01	0.01	0.01	0.00	0.00	0.21	0.05

a - for Case definitions, see Table 4.6-1.
b - Turnover Frequency defined as average number of times per year Richardson Number drops below 100.
c - Density gradient = hypolimnion density - epilimnion density (mg/cm³).
d - Hypolimnic Dilution Rate (HDR) = total flow through hypolimnion / hypolimnion volume (year⁻¹).
e - Ranges of within-group means for seasonal, yearly, and parameter factors.

Figure 4.6-6

Simulated Mean Density Gradients for Various Cases
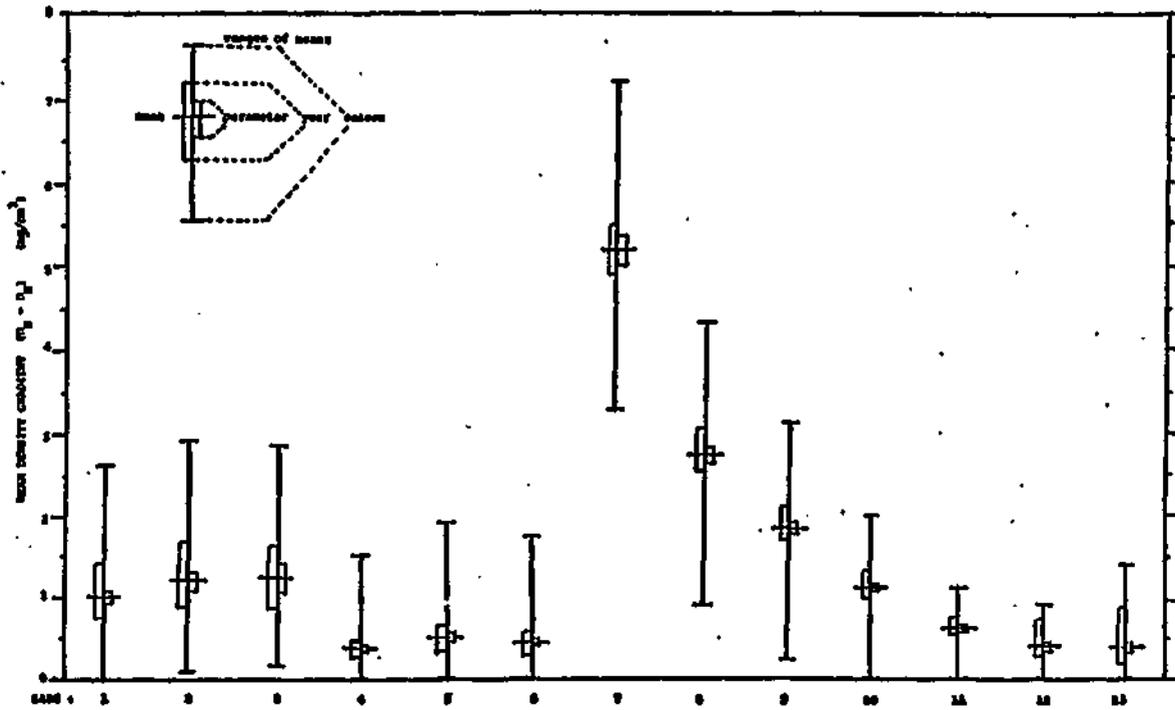
Figure 4.6-7

Simulated Maximum Annual Density Gradients for Various Cases
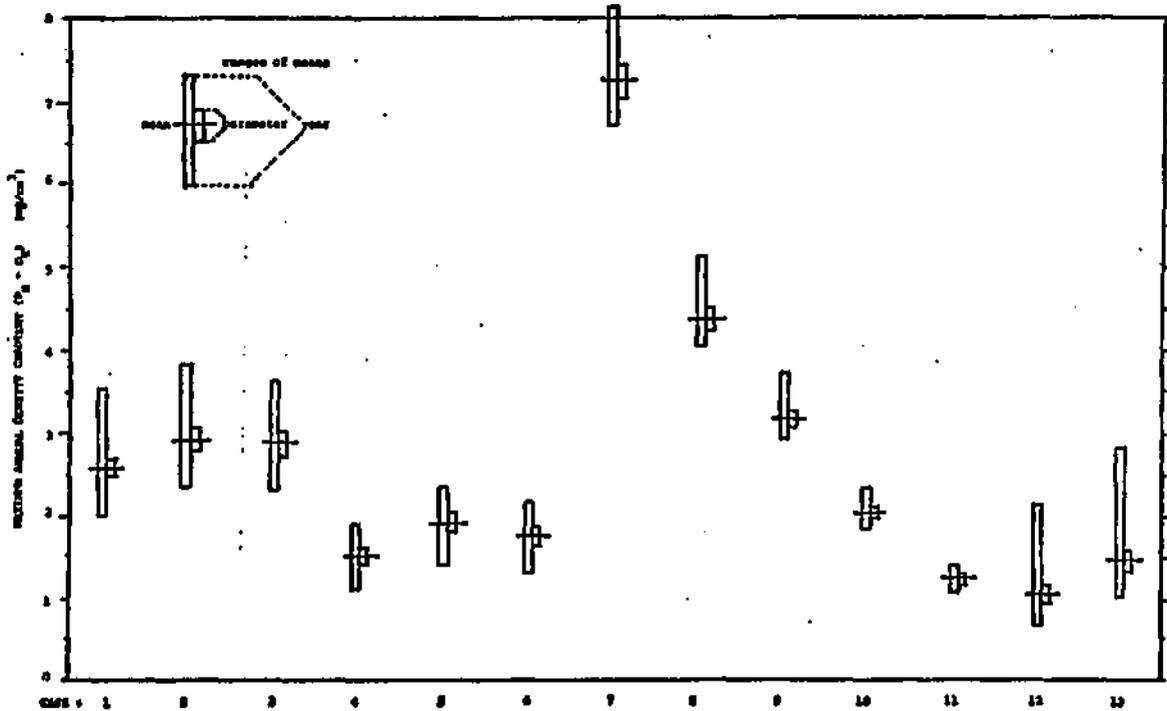
Figure 4.6-8

Simulated Mean Hypolimnic Dilution Rates for Various Cases
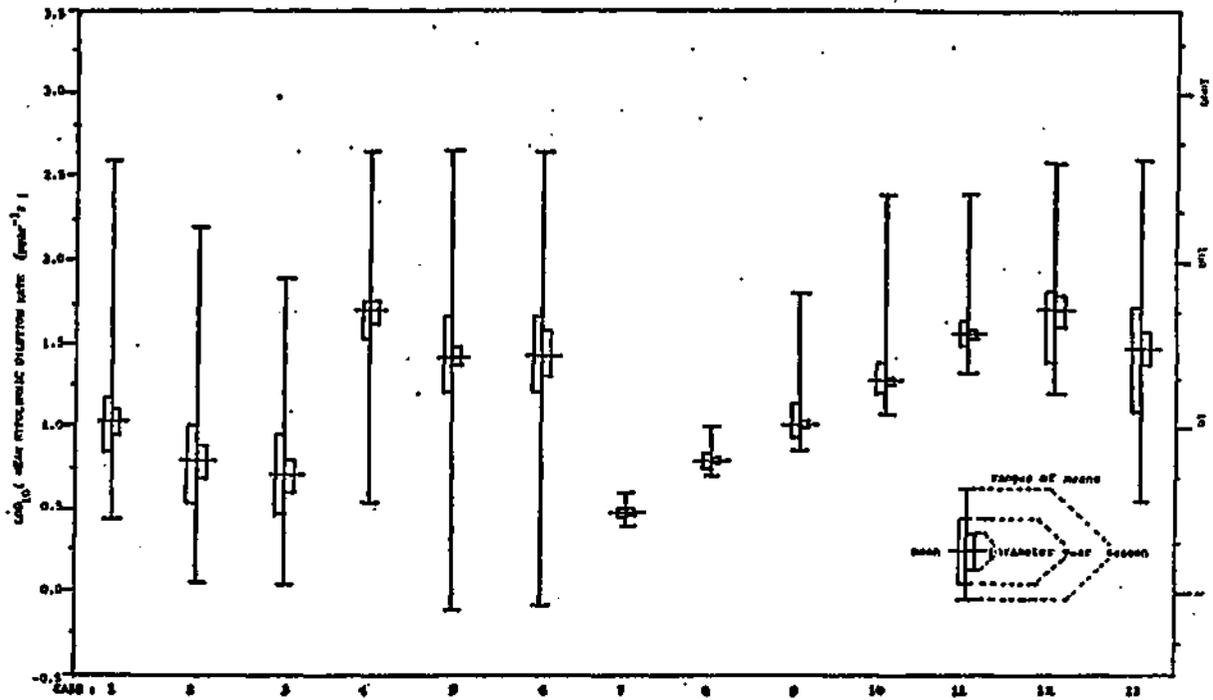
Figure 4.6-9

Simulated Minimum Annual Hypolimnic Dilution Rates for Various Cases
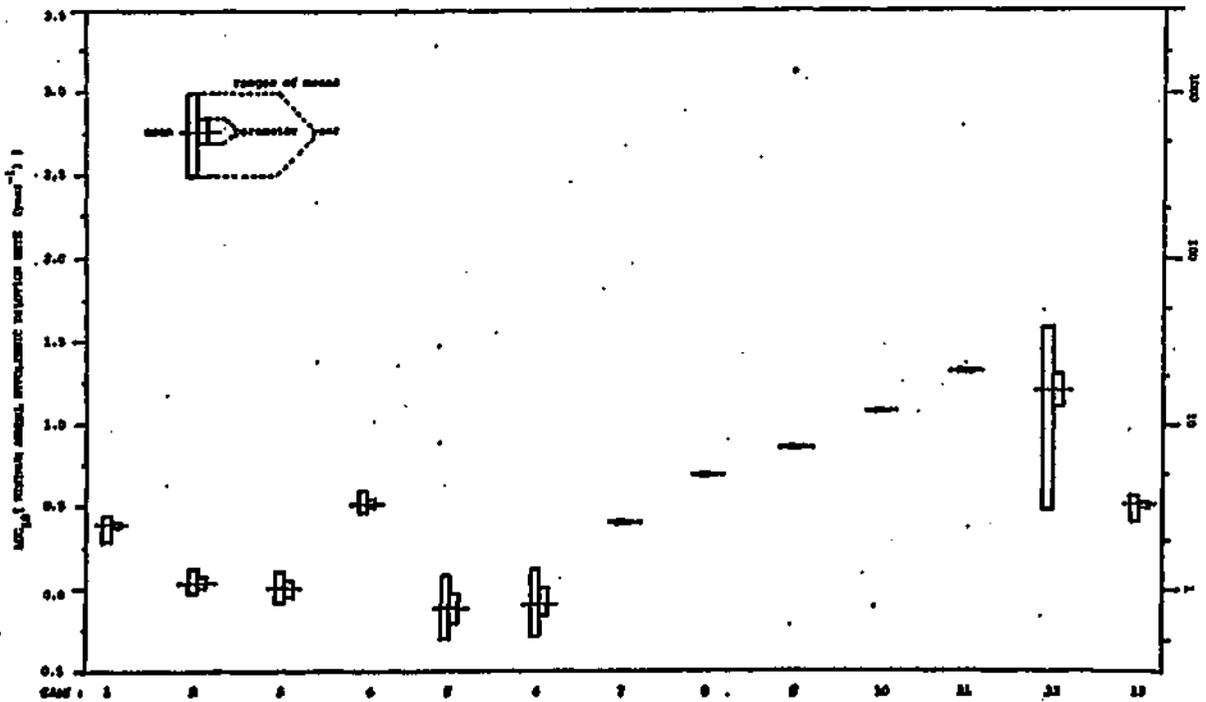
Figure 4.6-10

Simulations of Actual 1968-74 Conditions
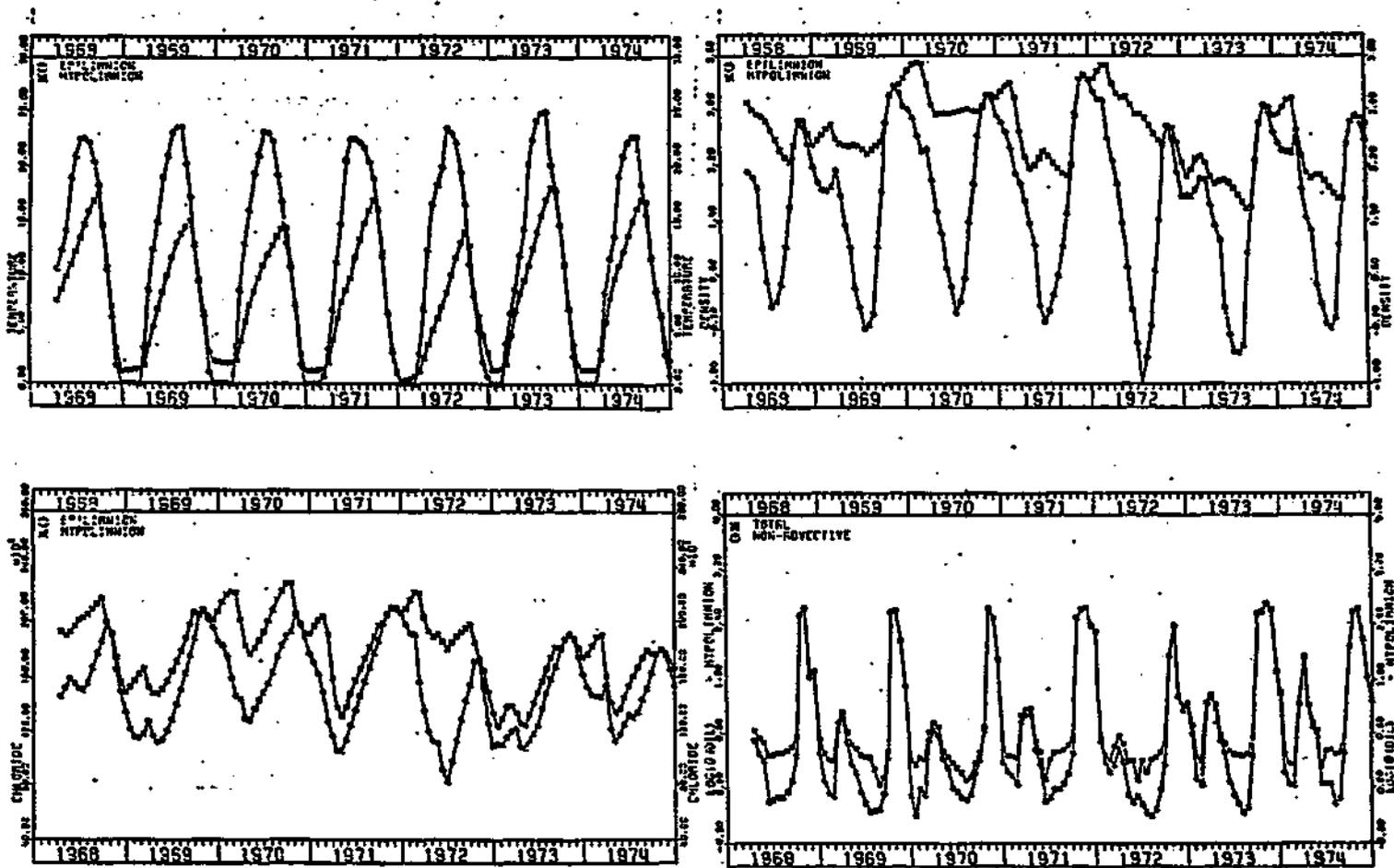
Figure 4.6-11

Simulations of Case 6 Conditions
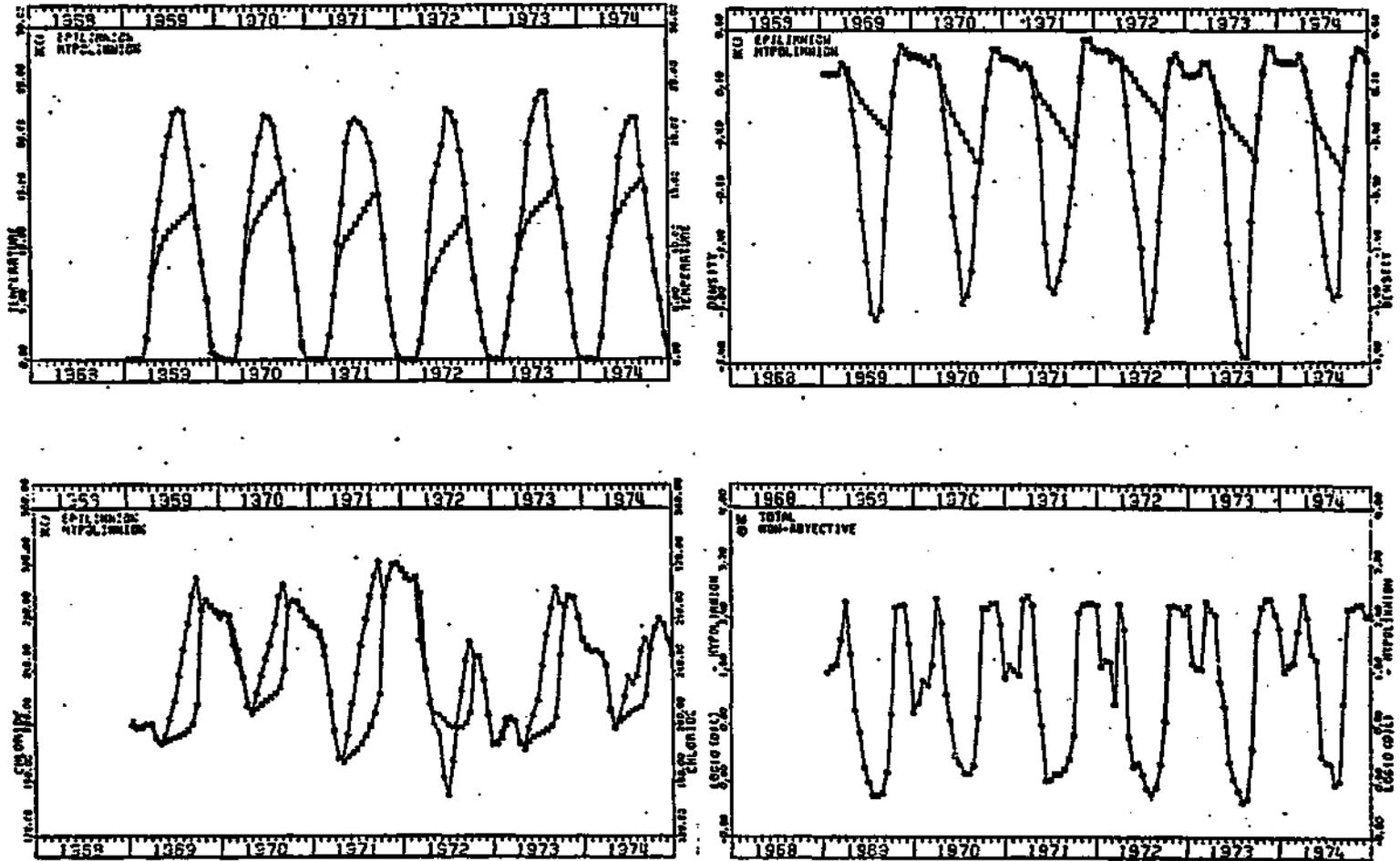
Figure 4.6-12

Simulations of Case 7 Conditions
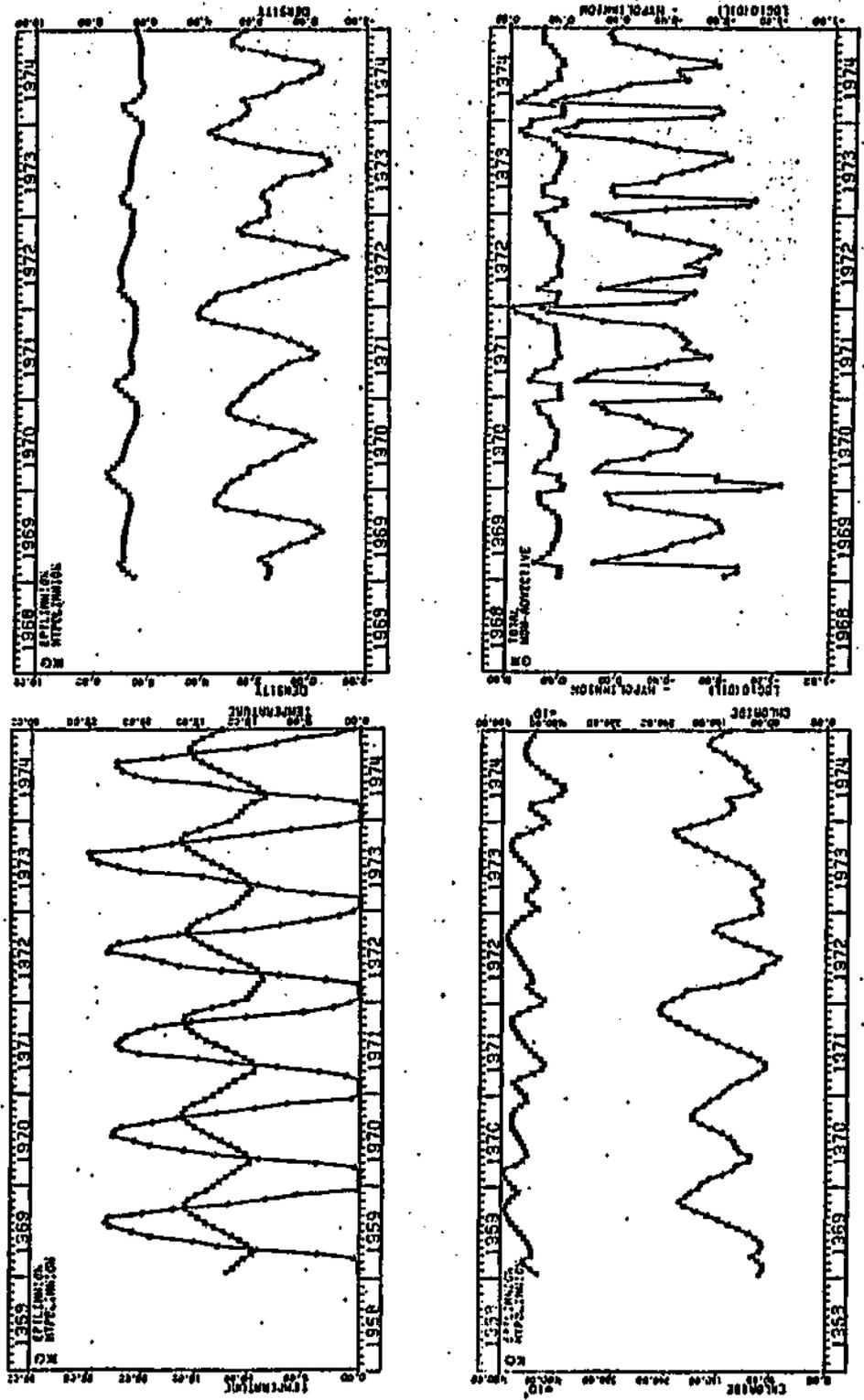
Figure 4.6-13

Simulations of Case 13 Conditions
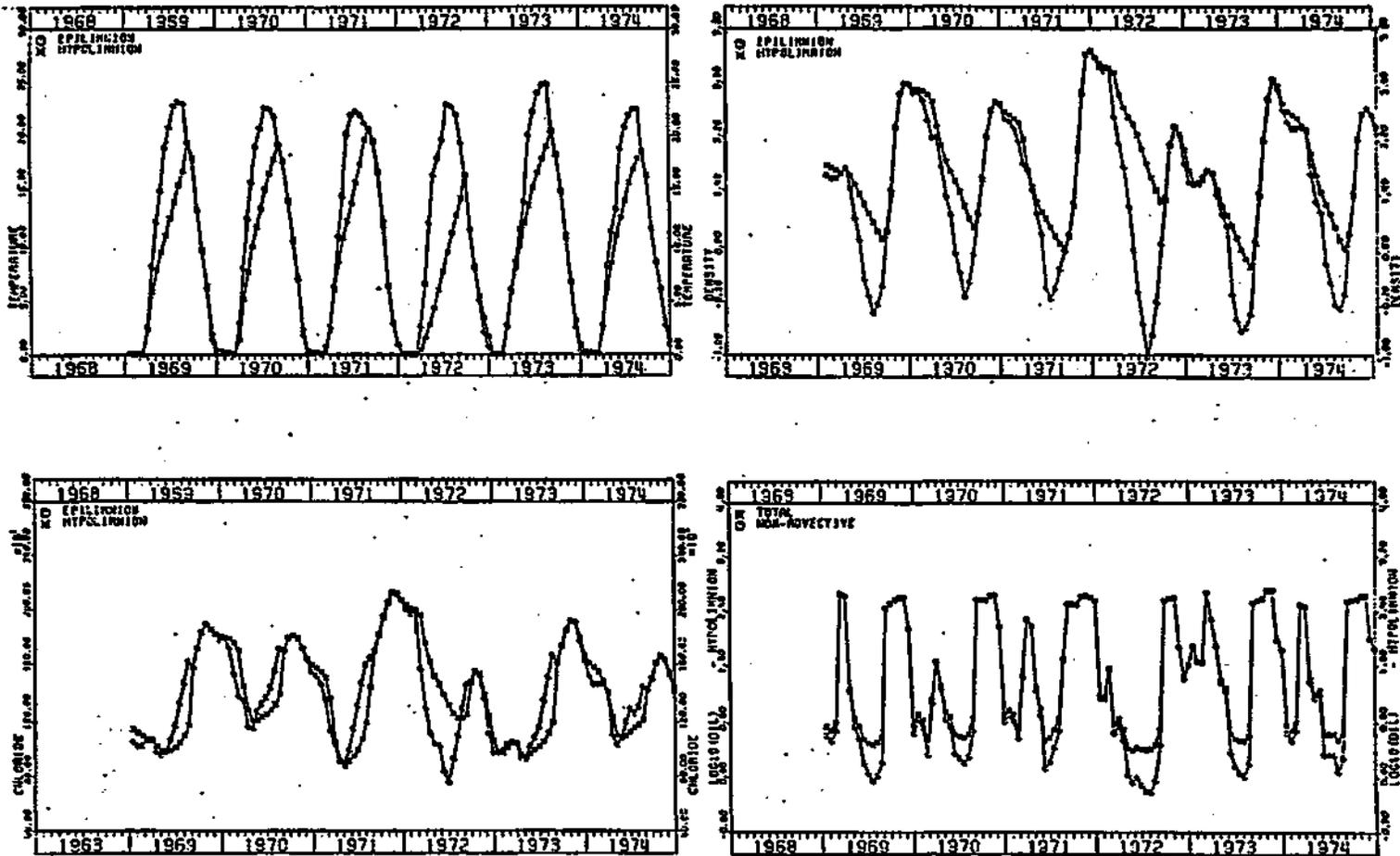
4.6.2 Case Simulations (continued)

lake would decrease by 62%. The corresponding reduction for maximum
annual density gradient is estimated to be 38%. The difference in the
logarithm of the mean HDR between the two groups of cases is about 0.7,
indicating a ratio of five in average displacement rates. The minimum
HDR values are governed chiefly by advective flows due to cooling water
withdrawal and are not as severely impacted by the salt discharge.

Comparing Cases 1, 2, and 3 provides as basis for evaluating the
effects of cooling water use under current salt disposal conditions.
Results indicate that the lake mixing afforded by withdrawal of cooling
water from the hypolimnion and subsequent discharge to the epilimnion
does have a calculable impact upon density gradiente and total HDR
values. Results for Cases 1 and 3 indicate that if the cooling water
use were stopped, turnover frequencies would decrease by 50%, mean and
maximum annual density gradients would decrease by 26% and 12%, respec-
tively, and mean and annual minimum HDR's would decrease by factors of
2 and 2.5, respectively. The differences between Cases 1 and 2 suggest
that observed increases in lake mixing rates over the 1968 to 1974
period may have been due in part to more frequent withdrawal of cooling
water from the hypolimnion in later years (see U_4, Table 4.3-2).
Corresponding effects of cooling water use without the salt discharge
(Cases 4 - 6) are somewhat less marked, because of the increase in
density-dependent mixing, which reduces the relative importance of the
advective mixing afforded by cooling water use. Density gradients
for Case 5, in which cooling water is withdrawn primarily from the

4.6.2 Case Simulations (continued)

epilimnion, are somewhat higher than for Case 6, without cooling water. This is an effect of heat disposal in the epilimnion increasing the density stratification. The average heat disposal rate amounts to about 19% of the average annual solar radiation incident on the lake.

Simulations of Case 7 indicate that discharge of the combined municipal/industrial effluent into the epilimnion without initial dilution, or, equivalently, discharge into the hypolimnion, would result in permanent stratification under the hydrologic and meteorologic conditions of 1968-74 (Figure 4.6-10). At the other extreme, results for Case 13 (Figure 4.6-13 suggest that infinite initial dilution in the epilimnion would result in an average of 1.83 turnovers per year, and in density gradient and HDR values not much different from the "undisturbed state", Case 6. As noted above, the path between these two cases as a function of dilution ratio is uncertain because of uncertainty in parameter a_{21}. Results obtained for a_{21} equal to 1.0 are conservative in that they would tend to over-predict density currents and resultant effects on chloride and density gradients in the lake. The consequences of these results on outfall design will be discussed in Section 4.7.

For all cases and objective functions, estimated variations due to parameter values are small compared with those attributed to season and year and compared with differences in means among the various cases. The parameter variations have been derived from the 95%

4.6.2 Case Simulations (continued)

confidence regions of the parameter estimates. As noted in Section 4.5, the estimate of the size of this region is inexact because of non-linearity in the model and the serial dependence and non-normality of the residuals. Observed yearly ranges are on the order of four times those attributed to the parameters. The former have been derived from a sample of six years, and thus do not represent 95% confidence ranges. If the distribution of yearly means is assumed to be normal, the standard deviation of the yearly variations would be given by 0.395 times the range for a sample size of 6[27], and the 95% confidence range would be about 15.8 times the observed range for six years. Thus, in order for the parameter variations to equal yearly ones, the estimate of the size of the parameter confidence region (expressed as distance along the principal component axee) would have to be low by a factor of 1.58 x 4 = 6.32.

Bard[3.] has derived an expression for the ε-indifference region of the maximum-likelihood objective function for the situation in which the sampling distribution (multivariate distribution of the parameters) cannot be assumed to be normal:

$$\Delta\phi = \varepsilon = \frac{1}{2}\frac{\ell}{1 - \gamma} \tag{4.6-5}$$

where,

4.6.2 Case Simulations (continued)

$$\ell = \text{number of estimated parameters} = 4$$

$$\gamma = \text{confidence level} = 0.95$$

Paralleling the development in Section 4.5 for the normal case, the corresponding distance along the principal component axes, expressed in multiples of the respective principal component standard deviations is given by:

$$z = \sqrt{2\ \Delta\phi} = \sqrt{\frac{\ell}{1-\gamma}} = 8.94 \qquad (4.6\text{-}6)$$

For the normal case, $\Delta\phi$ is given by $\chi^2/2$, and z is given by $\sqrt{\chi}$, or 3.08. Thus, if an attempt is made to account for non-normality, the dimensions of the confidence region increase by a factor of $8.94/3.08 = 2.90$. The same proportionate increase would be expected in the ranges of the density gradient and HDR's attributed to parameter variations, according to equation (4.6-4) Under these conditions, the 95% confidence range attributed to yearly variations would be $6.32/2.90 = 2.18$ times the range attributed to parameter ranges. Generally, the higher this ratio, the more difficult it would be to detect projection errors due to errors in the parameter estimates. The non-linearity of the model and serial dependence of the residuals have not been accounted for, however. This is about as far as the analysis can be taken without resorting to Monte-Carlo type methods to determine actual parameter confidence regions.

4.6.3 Error Analysis

An analysis has been done to estimate some of the various error components of model projections in each case studied. Assuming computation error to be negligible, total projection error can be broken down into three basic components[3] :

$$S_T^2 (t) = S_I^2 (t) + S_P^2 (t) + S_R^2 \qquad (4.6-7)$$

where,

$S_T^2 (t)$ = variance of a projection of a given variable at time t

$S_I^2 (t)$ = variance attributed to independent variable error

$S_P^2 (t)$ = variance attributed to parameter error

S_R^2 = residual variance

In the case of a dynamic model, all of the above are time-variable, with the exception of S_R^2. The first term is given approximately by:

$$S_I^2 (t) \approx \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \left(\frac{\partial y(t)}{\partial x_i}\right)\left(\frac{\partial y(t)}{\partial x_j}\right) \text{Cov} (x_i, x_j) \qquad (4.6-8)$$

4.6.3 Error Analysis (continued)

where,

$y(t)$ = value of predicted variable y at time t

n_x = number of independent variables

x_i = value of independent or forcing variable i

Similarly, the parameter error term is given by:

$$s_p^2(t) \simeq \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \left(\frac{\partial y(t)}{\partial p_i}\right)\left(\frac{\partial y(t)}{\partial p_j}\right) \text{Cov}(p_i, p_j) \qquad (4.6\text{-}9)$$

where,

p_i = value of parameter i

n_p = total number of parameters

The last term is attributed to residual error and is determined by the standard error of estimate of the predicted variable. This component can be further dissected into two components :

$$s_R^2 = s_E^2 + s_M^2 = \overline{SEE}_y^2 \qquad (4.6\text{-}10)$$

4.6.3 Error Analysis (continued)

where,

$$S_E^2 = \text{model error}$$

$$S_M^2 = \text{measurement error}$$

$$\overline{SEE}_Y^2 = \text{square of standard error of estimate of predicted variable}$$

In order to estimate the two components of the residual variance, some independent evidence is generally required. For example, if replicate samples, chloride analyses, and temperature measurements were available, the measurement error components could be estimated and subtracted from the observed residual variance to obtain estimates of the model error components. In the absence of such data, the aggregated form of the residual variance has been employed below. As noted in Section 4.5, residual standard errors are not strictly independent of time or season, so the assumption that they are constant is not completely valid, but it sufficient for the analysis below.

As discussed in Chapter 1, a method of assessing the adequacy of data for parameter estimation purposes is to compare tha error components attributed to parameter uncertainty, S_p^2 (t) with the residual variance, S_R^2 . The former can usually be reduced by increasing the size of the data sample used for parameter estimation, while the latter would not necessarily change under such conditions, provided that the additional data were of the same quality as that

4.6.3 Error Analysis (continued)

originally employed.

The magnitudes of the parameter and residual error components have been compared for each case simulated and for predictions of the epilimnic, hypolimnic, and gradient values of temperature, chloride, and density. The parameter error term is time-variable and its distribution for the entire 6-year simulation has been summarized by the extreme values and the mean. The variance components for each case and variable are summarized in Table 4.6-3. Generally, mean parameter variance amounts to a small fraction of the sum of the two error components. The maximum fraction observed is 0.18, for prediction of temperature gradient in Case 3. Case 3 appears to be the most sensitive to parameter values of the cases studied. For prediction of density gradient the fraction of total variance attributed to parameters ranges from 0.0004 in Case 11 to 0.034 in Case 7. Despite the approximate nature of the estimate of the covariance matrix of parameter values, it appears that the estimate would have to be considerably in error in order for the parameter variance to become a substantial component of the total projection error.

A more complete error analysis would also incorporate the effects of independent variable error. This would involve considerable effort, due to the number of independent variables and the difficulties of estimating their covariance matrix. To some extent, this type of error has already been incorporated into the parameter and residual error terms, since the parameter and residual covariance matrices have

Table 4.6-3

Parametric and Residual Variance Components for Each Case and Predicted Variable

Case	Parameter			Residual Variance	Total Variance	Mean Param / Total
	Min. Var.	Max. Var.	Mean Var.			
			T_H			
1	0.0	0.0426	0.0156	2.7900	2.8056	0.0056
2	0.0	0.0313	0.0148	2.7900	2.8048	0.0053
3	0.0	0.1131	0.0163	2.7900	2.8063	0.0058
4	0.0	0.0435	0.0157	2.7900	2.8057	0.0056
5	0.0	0.0411	0.0149	2.7900	2.8049	0.0053
6	0.0	0.0330	0.0118	2.7900	2.8018	0.0042
7	0.0	0.0356	0.0148	2.7900	2.8048	0.0053
8	0.0	0.0376	0.0150	2.7900	2.8050	0.0053
9	0.0	0.0424	0.0153	2.7900	2.8053	0.0055
10	0.0	0.0530	0.0161	2.7900	2.8061	0.0057
11	0.0	0.0505	0.0161	2.7900	2.8061	0.0057
12	0.0	0.0527	0.0170	2.7900	2.8070	0.0061
13	0.0	0.0415	0.0155	2.7900	2.8055	0.0055
			T_E			
1	0.0024	0.1156	0.0371	3.5000	3.5371	0.0105
2	0.0047	1.6550	0.1286	3.5000	3.6286	0.0355
3	0.0046	11.5101	0.6448	3.5000	4.1448	0.1556
4	0.0000	0.0399	0.0166	3.5000	3.5166	0.0047
5	0.0000	0.0617	0.0228	3.5000	3.5228	0.0065
6	0.0	0.0621	0.0204	3.5000	3.5204	0.0058
7	0.0003	0.0475	0.0107	3.5000	3.5107	0.0030
8	0.0013	0.0432	0.0120	3.5000	3.5120	0.0034
9	0.0013	0.0445	0.0135	3.5000	3.5135	0.0038
10	0.0003	0.0519	0.0123	3.5000	3.5123	0.0035
11	0.0000	0.0459	0.0137	3.5000	3.5137	0.0039
12	0.0000	0.0682	0.0165	3.5000	3.5165	0.0047
13	0.0002	0.1762	0.0363	3.5000	3.5363	0.0103
			$T_H - T_E$			
1	0.0000	0.1409	0.0369	3.9600	3.9969	0.0092
2	0.0012	3.2256	0.1537	3.9600	4.1137	0.0374
3	0.0005	17.5944	0.8665	3.9600	4.8265	0.1795
4	0.0000	0.0491	0.0124	3.9600	3.9724	0.0031
5	0.0000	0.1203	0.0295	3.9600	3.9895	0.0074
6	0.0	0.1052	0.0262	3.9600	3.9862	0.0066
7	0.0104	0.6492	0.0255	3.9600	3.9855	0.0064
8	0.0027	0.0402	0.0137	3.9600	3.9737	0.0034
9	0.0012	0.0433	0.0091	3.9600	3.9691	0.0023
10	0.0002	0.0170	0.0037	3.9600	3.9637	0.0009
11	0.0000	0.0105	0.0016	3.9600	3.9616	0.0004
12	0.0000	0.0286	0.0025	3.9600	3.9625	0.0006
13	0.0000	0.2194	0.0352	3.9600	3.9952	0.0088
			C_E			
1	0.0097	0.0568	0.0219	39.2040	39.2259	0.0006
2	0.0065	0.5707	0.0693	39.2040	39.2733	0.0018
3	0.0037	0.2645	0.0810	39.2040	39.2850	0.0021
4	0.0000	0.0003	0.0001	39.2040	39.2041	0.0000
5	0.0000	0.0012	0.0002	39.2040	39.2042	0.0000
6	0.0000	0.0017	0.0001	39.2040	39.2041	0.0000
7	0.0034	0.0691	0.0227	39.2040	39.2267	0.0006
8	0.0015	0.0542	0.0137	39.2040	39.2177	0.0003
9	0.0016	0.0639	0.0117	39.2040	39.2157	0.0003
10	0.0004	0.0242	0.0058	39.2040	39.2098	0.0001
11	0.0005	0.0201	0.0023	39.2040	39.2063	0.0001
12	0.0001	0.1794	0.0437	39.2040	39.2477	0.0011
13	0.0010	0.0173	0.0050	39.2040	39.2090	0.0001
			C_H			
1	0.0279	0.6243	0.2480	34.2250	34.4730	0.0072
2	0.0833	8.3643	1.1849	34.2250	35.4099	0.0335
3	0.1613	22.4327	2.1599	34.2250	36.3849	0.0594
4	0.0000	0.0319	0.0002	34.2250	34.2252	0.0000
5	0.0000	0.0348	0.0004	34.2250	34.2254	0.0000
6	0.0000	0.0017	0.0003	34.2250	34.2253	0.0000
7	2.0129	6.1693	3.7219	34.2250	37.9469	0.0981
8	0.2417	1.4145	0.6162	34.2250	34.8412	0.0177
9	0.0636	0.5113	0.2169	34.2250	34.4419	0.0063
10	0.0127	0.1635	0.0630	34.2250	34.2880	0.0019
11	0.0024	0.0048	0.0212	34.2250	34.2462	0.0006
12	0.0018	1.2646	0.0760	34.2250	34.3010	0.0022
13	0.0022	0.1550	0.0324	34.2250	34.2574	0.0009
			$C_H - C_E$			
1	0.0001	0.8576	0.3258	39.6010	39.9268	0.0082
2	0.0170	4.0350	1.2384	39.6010	40.8394	0.0303
3	0.0345	25.0813	2.3230	39.6010	41.9240	0.0554
4	0.0000	0.0331	0.0002	39.6010	39.6012	0.0000
5	0.0000	0.0113	0.0007	39.6010	39.6017	0.0000
6	0.0000	0.0070	0.0004	39.6010	39.6014	0.0000
7	1.5530	6.7234	3.6270	39.6010	43.2280	0.0839
8	0.1075	1.4739	0.6207	39.6010	40.2217	0.0154
9	0.0054	0.8330	0.2351	39.6010	39.8361	0.0059
10	0.0009	0.5034	0.0778	39.6010	39.6788	0.0020
11	0.0006	0.1858	0.0269	39.6010	39.6279	0.0007
12	0.0000	0.0972	0.0151	39.6010	39.6161	0.0004
13	0.0000	0.2277	0.0446	39.6010	39.6456	0.0011
			D_H			
1	0.0149	2.5698	0.6151	185.0000	185.6151	0.0033
2	0.0444	3.2148	0.8953	185.0000	185.8953	0.0048
3	0.0278	2.3541	0.6870	185.0000	185.6870	0.0037
4	0.0000	2.0611	0.4675	185.0000	185.4675	0.0025
5	0.0001	2.5230	0.5236	185.0000	185.5236	0.0028
6	0.0001	1.6391	0.3240	185.0000	185.3240	0.0018
7	0.0145	2.7803	0.6605	185.0000	185.6605	0.0036
8	0.0213	2.3901	0.5878	185.0000	185.5878	0.0032
9	0.0113	2.1731	0.5675	185.0000	185.5675	0.0031
10	0.0074	2.1746	0.5365	185.0000	185.5365	0.0029
11	0.0379	2.2602	0.5288	185.0000	185.5288	0.0029
12	0.0022	3.0344	0.6744	185.0000	185.6744	0.0036
13	0.0035	1.9620	0.4912	185.0000	185.4911	0.0026
			D_E			
1	0.1076	3.1291	1.2546	130.0000	131.2546	0.0096
2	0.2057	29.7515	3.7555	130.0000	133.7555	0.0281
3	0.5648	56.5606	7.0416	130.0000	137.0416	0.0514
4	0.0003	0.4544	0.2672	130.0000	130.2672	0.0021
5	0.0001	1.0829	0.2710	130.0000	130.2710	0.0021
6	0.0001	0.5800	0.2231	130.0000	130.2231	0.0017
7	4.7603	13.2247	8.1159	130.0000	138.1159	0.0588
8	0.8924	3.0896	1.6132	130.0000	131.6132	0.0123
9	0.2548	1.3072	0.7344	130.0000	130.7344	0.0056
10	0.0776	1.1927	0.4637	130.0000	130.4637	0.0036
11	0.0210	1.8019	0.4209	130.0000	130.4209	0.0032
12	0.0007	3.0527	0.6326	130.0000	130.6326	0.0048
13	0.0395	2.9270	0.7037	130.0000	130.7037	0.0054
			$D_H - D_E$			
1	0.0004	4.6080	1.7694	230.0000	231.7694	0.0076
2	0.0236	26.8335	4.2726	230.0000	234.2725	0.0182
3	0.1232	114.7940	7.7212	230.0000	237.7212	0.0325
4	0.0000	1.8452	0.3756	230.0000	230.3756	0.0016
5	0.0000	3.6190	0.7347	230.0000	230.7347	0.0032
6	0.0000	2.6489	0.5530	230.0000	230.5530	0.0024
7	4.1769	14.5512	8.1753	230.0000	238.1753	0.0343
8	0.6105	3.4889	1.7137	230.0000	231.7137	0.0074
9	0.1076	1.6689	0.7331	230.0000	230.7331	0.0032
10	0.0133	3.6920	0.2429	230.0000	230.2429	0.0011
11	0.0303	0.3225	0.1019	230.0000	230.1019	0.0004
12	0.0002	1.3898	0.1234	230.0000	230.1234	0.0005
13	0.0032	3.7066	0.7626	230.0000	230.7626	0.0033

* - To get variances on original scales, multiply values in table by 1.0, 1000., and .001 for T, C, and D variables, respectively.

4.6.3 Error Analysis (continued)

been estimated by supplying the model with a set of independent
variables which were subject to error and subsequently comparing
dependent variable observations with model predictions. Generally,
the same set of independent variables has been used in simulating
each case, so that the same covariance matrix of independent
variable errors [Cov (x_i, x_j) in equation (8)] would apply.
However, the sensitivities of the state variables with respect to
the independent variables [$\partial y(t)/\partial x_i$ in equation (8)] may change
as a function of case, suggesting that total independent variable
error may also change with case, just as total parameter error has
been shown to change.

Probably the most important concern with regard to the effects
of this type of error is the possibility that consistent errors or
biases in one or more of the independent variables may have intro-
duced biases in the parameter estimates. One variable which would
be of potential concern in this regard is U_6 , the chloride flux
from the Allied Chemical waste beds. As demonstrated in Section 4.6,
the hypolimnion and gradient variables are highly sensitive to the
values of this variable. This indicates that the optimal parameter
estimates, particularly a_{14} , would also be sensitive to bias in
U_6 . Fortunately, as discussed in Section 4.3, the value assumed
for this variable has been verified by two independent methods:
(1) Ninemile Creek flow and chloride concentration data, and (2) a
mass balance on the Allied Chemical solvay plant developed from soda

4.6.3 Error Analysis (continued)

ash production, reaction stoichiometry, and reaction efficiency data.
Because of the possibility for introduction of bias, the quality of
the independent variable data can be critical to proper implementation
of the empirical parameter estimation methods employed above.

4.7 Conclusions and Recommendations

4.7.1 Model Adequacy

One distinctive feature of the model which has been developed,
estimated, and applied above is its high level of spatial and temporal
aggregation. Monthly-average boundary conditions have been employed
to drive a system consisting of two mixed layers. This is to be
compared with the M.I.T.[1,2] and W.R.E.[9] reservoir models discussed
in Section 4.2. These models employ a minimum of 20 mixed layers
of 1- to 2-meter thicknesses and require daily-average boundary
conditions. Aggregation in space and time go together, because the
former tends to enhance the stability of the system and to render
it less sensitive to high-frequency variations in boundary conditions.

This aggregation has resulted in a number of distinctive
properties. First, the model obviously has less potential resolution
in time and space. The resolution was sufficient for generally
characterizing vertical exchange in the lake , as evidenced by the
ability of the model to simulate observed temperature and chloride
variations in the hypolimnion and epilimnion with time. Second, the model
requires less data to implement. Obviously, the specification of
daily-average meteorologic and hydrologic boundary conditions would

4.7.1 __Model Adequacy__ (continued)

be a much more demanding requirement. Third, the model is less expensive to implement, because it involves fewer state variables and longer time steps. As noted in Section 4.2, the M.I.T. model, with spatial and temporal resolutions of one meter and one day, respectively, consumes 2.4 minutes of CPU time per year of simulation, while the model developed here uses 0.028 CPU minutes per year*. Use of the parameter estimation routine would probably have been economically infeasible if a model with the spatial and temporal resolution of the M.I.T. model were employed.

The standard errors of estimate of the model developed here are compared with those typical of the M.I.T. model, as derived from the study of Parker et al.[11] in Table 4.7-1. For predictions of surface temperature, the standard errors for the latter averaged 2.15°C and ranged from 1.38 to 2.98°C in a sample of seven different reservoirs and years. These can be compared with the standard error of 1.67°C obtained for predictions of epilimnion temperature in Onondaga Lake. Corresponding values for predictions of reservoir outlet temperatures averaged 1.90°C and ranged from 1.01 to 2.88°C using the M.I.T. model, compared with a value of 1.87°C obtained above for prediction of hypolimnion temperature. Thus, it appears that the two models have about the same residual standard errors, despite the differences in their levels of spatial and temporal resolution.

---

* Some of the differences in CPU time may. be attributed to differences in computers. The computer model employed in the 1975 study of the M.I.T. model was not specified by the authors.[25]

Table 4.7-1

Comparisons of Standard Errors of Estimate for Temperature Predictions in
Onondaga Lake with Those Typical of the M.I.T Deep Reservoir Model

| Model | Lake or Reservoir | Years | STANDARD ERRORS OF ESTIMATE ($^\circ$C) | |
			Surface Temperature	Bottom Temperature
ESP	Onondaga	1968-74	1.67[a]	1.87[b]
MIT[d]	Fontana	1966	1.74	1.19
MIT	Douglas	1969	2.03	2.11
MIT	Cherokee	1967	2.07	2.70
MIT	Norris	1972	2.31	-
MIT	So. Holston	1953	2.44	2.88
MIT	Hiwassee	1947	1.38	1.01
MIT	Fort Loudon	1971	2.98	1.53

a — epilimnion-average.

b — hypolimnion-average.

c — Model of Harleman et al. [11,13], as evaluated by Parker et al.[25];
bottom temperature refers to outlet level.

4.7.1 Model Adequacy (continued)

One feature which renders the model developed above somewhat unrealistic is the assumption of constant epilimnic and hypolimnic volumes. Because of thermocline migration, these volumes actually vary with season and could possibly be influenced by some of the conditions specified by the various cases in Section 4.7. Realism could be enhanced by incorporating another state variable which would permit simulation of thermocline migration. This would involve a modification of the exchange function (F_{11}), but mechanical energy balance considerations would still apply. Because the relative changes in volume could have substantial effects on the dynamics of non-conservative components (in particular, algal growth rates), this element of realism should probably be added to the model before expanding it to permit direct simulation of water quality components.

4.7.2 Data Adequacy

Generally, two types of data have been employed above to serve two purposes: definition of system conditions and definition of boundary conditions. The adequacy of the former as a basis for parameter estimation is indicated by the relatively small contributions of parameter variance to total model prediction variance, as noted in Section 4.6.3. The adequacy of the boundary condition data is more difficult to assess. Much of the auto-correlation and periodicity in the residuals has been attributed to concomittant errors in the specification of meteorologic or hydrologic boundary conditions (e.g.,

4.7.2 Data Adequacy (continued)

solar radiation). These residual properties have been shown to introduce uncertainty into the estimates of the parameter covariance matrix and confidance region. Residual standard errors can actually be attributed to three components: measurement error, model error, and independent variable error. An assessment of these components is not possible without independent evidence. The primary concern over the potential effects of independent variable error is that consistent errors or biases in these data may have given rise to biases in the parameter estimates. Such biases, in turn, may have influenced the policy recommendations derived from model applications. It would be difficult if not impossible to make the analysis robust to such errors.

4.7.3 Methods Adequacy

This work has provided a partial basis for assessment of the values and limitations of Bard's parameter estimation algorithms in modelling studies of this type. A number of modifications and additions to these algorithms have been made, specifically to permit : (1) mapping of objective function values and residual sums of squares for specified parameter values; (2) use of Runge-Kutta integration; (3) computation of derivatives of the objective function with respect to the parameter values using finite-difference methods; (4) estimation of the limits of the parameter confidence region. These additions have enhanced the flexibility of the parameter estimation program, which has demonstrated reasonable potential.

The methods have been employed to estimate optimal sets of parameters

4.7.3 Methods Adequacy (continued)

for each year individually and for all years combined. In the yearly exercises, in which two parameters were estimated, convergence was achieved using the Gauss-Newton maximization algorithm in from 3 to 7 interations (derivative evaluations), 8 to 19 objective function evaluations, or from 15 to 33 equivalent function evaluations. When all data were combined and four parameters were estimated, convergence was reached in 7 iterations, 16 objective function evaluations, or 44 equivalent function evaluations. The numbers of iterations are typical of those required in nonlinear estimation problems[33]. The use of finite-difference methods, as opposed to sensitivity equations, for computation of derivatives does not appear to have impaired convergence rates. In these problems, the objective functions and parameter values changed fairly rapidly over the first one or two iterations and rapidly approached assymptotic values as the solutions were approached.

Most of the cost of estimating the parameters is attributed to numerical integration of one set of state variable equations throughout the time period for each equivalent function evaluation required. As noted above, such costs could become prohibitive as the level of temporal and spatial resolution in the model increases. The algorithms appear to be most practical for use with aggregated models. If high resolution is not required, the use of aggregated models offers an additional advantage in that it would entail the aggregation of observations, which would also serve to remove some of the measurement error. This, in turn, could improve convergence rates by rendering the solution a more well-defined region in parameter space.

4.7.3 Methods Adequacy (continued)

One of the most important benefits derived from use of the algorithms is that both the first and second moments of the parameters can be estimated. These can provide a basis for estimation of the confidence region and for use of parameter stability criteria for model verification. The parameter and residual covariance matrices can also be used to estimate the covariance matrix of model predictions. Comparisons of parametric and residual variance components can lead to an assessment of the adequacy of the amount of data employed for parameter estimation purposes. Generally, because of nonlinearity in the model and serial dependence and non-normality in the residuals, the estimates of the residual and parametric covariance matrices are only approximate. For simple models, better estimates could be obtained using Monte-Carlo techniques.

Applications of the parameter estimation algorithms should be restricted to cases in which the independent or forcing variables are well-known. Biases in these variables could lead to biases in parameter estimates, a distorted picture of system dynamics, and biases in model predictions and policy recommendations. In such situation, prior parameter estimates, based upon independent experimental evidence might be employed, if available. Alternately, the unknown boundary conditions could be treated as unknown parameters and estimated along with the unknown system parameters, provided that the dimensions of the problem do not become excessive.[*] Although not exercised above,

---

[*] "Excessive" would probably be defined by economic constraints. As distributed, the Bard program can handle up to 20 unknown parameters. By increasing the sizes of the appropriate dimension statements, this upper limit could be increased. The feasibility of estimating more than four parameters simultaneously has not been demonstrated here. Generally, convergence problems might be encountered as the number of estimated parameters increases if some or all of the parameters are highly correlated with each other.[2]

4.7.3 Methods Adequacy (continued)

the option to include the effects of prior distributions of parameter esti-
mates on the objective function would seem potentially useful. In such a
case, the optimal parameter estimates would be based upon information obtained
both from the system under study and from independent evidence obtained
from experiments or from other natural systems. The information content
of the prior parameter distributions could improve the convergence properties
of the algorithms, which could be particularly effective when the size of
the unknown parameter vector is large.

In approaching parameter estimation problems of this type, prelimi-
nary studies should involve mapping of objective function and residual
sums of squares for observed variables over feasible regions of parameter
space. This can provide a feeling for the general location of the solution
and for the shape of the response surface, including the locations of any
local optima. Comparing the residual sums of squares for various dependent
variables can also be useful as a means of detecting independent variable
errors or model inadequacies. If such errors or inadequacies are not
significant, and if there are no biases in the measurements, minimal sums
of squares for all predicted variable residuals should all be located in
approximately the same region of parameter space. This type of preliminary
analysis can also provide reasonable initial guesses for implementation
of the formal parameter estimation routine.

4.7.4 Implications for Outfall Design

Model simulations indicate that the mode of discharge of the combined minicipal/industrial effluent could have relatively dramatic effects on the degrees of vertical mixing in Onondaga Lake. Discharge into the hypolimnion would apparently result in permanent stratification under the hydrologic and meteorologic conditions in the 1968 to 1974 pariod. At the other extreme, simulations indicate that discharge into the epilimnion with an infinite initial dilution ratio would minimize the impact of the industrial salt loading and induce a mixing regime substantially more vigorous than currently exists and not much different from what would exist if the unnatural salt discharge were eliminated. As discussed in Section 4.7, the path between these two limiting cases as a function of initial dilution ratio is uncertain because it depends upon the value of a parameter (a_{21}) whioh cannot be estimated a priori without a more detailed study of the hydrodynamics of the outfall site. Simulations with a dilution ratio of 16 and a conservative value of 1.0 for a_{21} (Case 12), are not greatly different from simulations with an infinite initial dilution ratio (Case 13). If allowance for this conservatism is made, it would appear that discharge into the epilimnion at a dilution ratio of 16 or greater would likely be equivalent to infinite intial dilution,in that it would essentially destroy the driving force for sinking of the effluent into the hypolimnion. It is unclear whether the current plan to discharge through a shoreline, surface outfall would achieve this level of initial dilution. A ratio of 16 would not be particularly difficult to achieve if an off-shore diffuser were employed.

4.7.4 Implications for Outfall Design (continued)

Decreases in density gradients over the course of the 1968-1974 period have been attributed to : (1) climatologic variations (possibly including (a) warmer winters in 1973-74, (b) higher wind speeds in spring of 1974, and (c) residual effects of the high flows experienced during Hurricane Agnes in June of 1972) and (2) the increased frequency of cooling water withdrawal from the hypolimnion by Allied Chemical in later years. Some of the water quality improvements observed over this period may have been partial responses to enhanced mixing rates. These include increased hypolimnic dissolved oxygen levels and reduced dominance of blue-green algae.

These observations suggest that future water quality could be sensitive to the design of the Metro STP outfall. If the outfall were designed to achieve a high rate of initial dilution of the effluent in ambient epilimnic waters, resultant enhanced mixing could improve hypolimnic oxygen levels. Based upon the analysis of Chapter 3, it could also result in somewhat higher transparencies and a less favorable environment for the growth of blue-green algae. Specfic water quality impacts could be more thoroughly investigated with an expanded model. Increasing the intitial dilution ratio would serve other purposes by reducing the impact of the effluent in the immediate discharge zone and possibly abating aesthetic problems by reducing the driving force for rapid precipitation of calcium carbonate, as discussed in Section 4.1. If economics are considered, it would appear that the flexibility and relative inexpense of the current plan for a surface, shoreline outfall are attractive. Should this alternative prove unacceptable from a lake mixing or aesthetic viewpoint, the analysis here indicates that

4.7.4 Implications for Outfall Design (continued)

an extended outfall should provide for dilution of the effluent in the epilimnion, if permanent density stratification is to be avoided.

Advective mixing provided by cooling water withdrawal from the hypolimnion and discharge to the epilimnion is significant compared with wind-induced mixing rates during midsummer seasons. Continued use of hypolimnic waters by Allied Chemical would result in a variety of environmental and economic impacts. Allied chlorinates its cooling water to minimize the impact of reducing compounds on its equipment[7]. Enhanced dissolved oxygen levels in the influent water would lower chlorine demand costs, while simultaneously permitting constant use of hypolimnic waters, the lower temperatures of which would reduce volume requirements and associated pumping costs. If advective flows into the hypolimnion due to density currents are eliminated (e.g., Case 13), withdrawal from the hypolimnion may result in a lower average thermocline level, a result not derivable directly from the model because of its constant volume constraints. Allied's withdrawal rate is about sufficient to lower the thermocline from 6 meters (approximate spring-time level) to about 10 meters in a 120-day period. Coupled with normal wind-induced thermocline migration, this may cause the hypolimnic volume to shrink more rapidly during the summer season than under current condi-tions, in which the hypolimnion is partially replenished by flow from Ninemile Creek. The hypolimnion would likely not be destroyed completely due to cooling water withdrawal, because Allied's deep intake ports would be exposed to the epilimnion, should the thermocline reach the vicinity of the 15-meter level.

REFERENCES - CHAPTER 4

1. Bard, Y., "A Function Maximization Method with Application to Parameter Estimation", New York Scientific Center Report 322.0902, I.B.M., May 1967.

2. Bard, Y., "Nonlinear Parameter Estimation and Programming", Program No. 360D-13,6,003, Share Program Library Agency, Triangle Universities Computation Center, Research Triangle Park, N.C., December 1967.

3. Bard, Y., Nonlinear Parameter Estimation, Academic Press, New York and London, 1974.

4. Banks, R.B., "Some Features of Wind Action on Shallow Lakes", Journal of the Environmental Enginsering Division, American Society of Civil Engineers, Vol. 101, No. EE5, pp. 813-827, October 1975.

5. Bella, D., "Dissolved Oxygen Variations in Stratified Lakes", Journal of the Sanitary Engineering Division, American Society of Civil Engineers, Vol. 96, No. SA5, Proc. Paper 7628, pp. 1129-1146, October 1970.

6. Blanton, J.O., "Vertical Entrainment into the Epilimnia of Stratified Lakes", Limnology and Oceanography, Vol. XVIII, No. 5, pp. 697-704, September 1973.

7. Clough, R., Allied Chemical Corporation, Syracuse, New York, personal communication, 1976.

8. Dake, J.M.K. and D.R.F Harleman, "Thermal Stratification in Lakes: Analytical and Laboratory Studies", Water Resources Research, Vol. 5, No. 2, pp. 404-495, April 1969.

9. Fletcher, R. and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", The Computer Journal, Vol. 6, No. 2, pp. 163-168, 1963.

10. Handbook of Physics and Chemistry, 36th Edition, 1955.

11. Harleman, D.R.F. and M. Markofsky, "A Predictive Model for Thermal Stratification and Water Quality in Reservoirs", MIT Department of Civil Engineering, Laboratory for Water Resources and Hydrodynamics, Report No. 134, January 1971.

12. Hoel, P.G., Introduction to Mathematical Statistics, Fourth Edition, John Wiley and Sons, 1971.

REFERENCES - CHAPTER 4 (continued)

13. Huber, W.C. and D.R.F. Harleman, "Laboratory and Analytical Studies of Thermal Stratification in Reservoirs", MIT Hydrodynamics Laboratory Technical Report No. 112, October 1968.

14. Huber W.C., D.R.F. Harleman, and P.J. Ryan, "Temperature Prediction in Stratified Reservoirs", Journal of the Hydraulics Division, American Society of Civil Engineers, Vol. 98, No. HY4, Proc. Paper 8839, pp. 645-646, April 1972.

15. Hutchinson, G.E., A Treatise on Limnology: Vol. 1 - Geography, Physics, and Chemistry, Wiley, New York, 1957.

16. International Business Machine Corp., System/360 Scientific Subroutine Package, Program No. 360A-CM-03X, Fifth Edition, 1970.

17. Johnston, J., Econometric Methods, Second Edition, McGraw-Hill Book Co., New York, 1972.

18. Kato, H. and O.M. Phillips, "On the Penetration of a Turbulent Layer into Stratified Fluid", Journal of Fluid Mechanics, Vol. 37, Part 4, pp. 643-655, 1969.

19. Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus, Hydrology for Engineers, McGraw-Hill Book Co., Second Edition, 1975.

20. Lorenzen, M. and R. Mitchell, "Theoretical Effects of Artificial Destratification on Algal Production in Impoundments ",Environmental Science and Technology, Vol. 7, No. 10, pp. 939-944, October 1973.

21. O'Brien and Gere, Engineers, "Onondaga Lake Monitoring Program", Annual Reports to Onondaga County, Department of Public Works, 1970-74.

22. O'Brien and Gere, Engineers, "Supplement to Wastewater Facilities Report - Metropolitan Sewage Treatment Plant Expansion", prepared for Onondaga County, April 1971.

23. Onondaga County, New York, "Onondaga Lake Study", US EPA, Water Quality Office, Publication No. 11000 FAE 4/71, April 1971.

24. Orlob, G.T. and L.G. Selna, "Temperature Variations in Deep Reservoirs", Journal of the Hydraulics Division, American Society of Civil Engineers, Vol. 96, No. HY2, Proc. Paper 7063, pp. 391-410, February 1970.

REFERENCES - CHAPTER 4 (continued)

25. Parker, F.L., B.A. Benedict, and C. Tsai, "Evaluation of Mathematical Models for Temperature Prediction in Deep Reservoirs", National Environmental Research Center, Office of Research and Development, US EPA, Document No. EPA-660/3-75-038, June 1975.

26. Rooney, J., "Evaluation of Interim Basin Plan IBP-NY-07-07, for Lake Onondaga, New York and Project WPC-NY-659 (Syracuse Metro STP)", EPA Region II Ofice, New York 1973.

27. Snedecor, G.W. and W.G. Cochran, <u>Statistical Methods</u>, Sixth Edition, Iowa State University Press, 1967.

28. Snodgrass, W.J., "A Predictive Phosphorus Model for Lakes - Development and Testing", Ph.D. Thesis, University of North Carolina at Chapel Hill, Environmental Sciences and Engineering, 1974.

29. Stefan, H. and D.E. Ford, "Temperature Dynamics in Dimictic Lakes", <u>Journal of the Hydraulics Division, American Society of Civil Engineers</u>, Vol. 101, No. HY1, pp. 97-114, January 1975.

30. Sundaram, T.R., C.C. Easterbrook, K.R. Prech, and G. Rudinger, "An Investigation of the Physical Effects of Thermal Discharges into Cayuga Lake", Cornell Aeronautical Laboratory, Buffalo, New York, Report VT-2616-0-2, November 1969.

31. Sundaram, T.R., R.G. Rehm, and G.E. Merritt, "A Study of Some Problems on the Physical Aspects of Thermal Pollution", Cornell Aeronautical Laboratory, Buffalo, New York, Report VT-2790-A1, June 1970.

32. Sundaram, T.R. and R.G. Rehm, "Formation and Maintenance of Thermocline in Temperate Lakes", <u>American Institute of Aeronautics and Astronautics</u>, Vol. 6, No. 2, pp. 1322-1329, 1971.

33. Sundaram, T.R. and R.G. Rehm, "Effects of Thermal Discharges on the Stratification Cycle of Lakes", <u>American Institute of Aeronautics and Astronautics</u>, Vol. 10, No. 2, pp. 204-210, 1972.

34. Turner, J.S. and E.B. Kraus, "A One Dimensional Model of the Seasonal Thermocline I.A Laboratory Experiment and its Interpretation", <u>Tellus</u>, Vol. 19, pp. 88-97, 1967.

REFERENCES - CHAPTER 4 (continued)

35. Turner, J.S., "The Influence of Molecular Diffusivity on Turbulent Entrainment Across a Density Interface", _Journal of Fluid Mechanics_, Vol. 33, pp. 639-656, 1968.

36. Thompson, E.S., "Computation of Solar Radiation from Sky Cover", _Water Resources Research_, Vol. 12, No. 5, pp. 859-865, October 1976.

37. US EPA, Region II, "Environmental Impact Statement on the Wastewater Treatment Facilities Construction Grants for the Onondaga Lake Drainage Basin", May 1974.

38. U.S. Weather Service, "Local Climatological Data", U.S. Department of Commerce, 1965-74.

39. Velz, C.J., _Applied Stream Sanitation_, Wiley-Interscience, New York, 1970.

40. Vollenweider, R.A., "Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, with Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication", Organization for Economic Cooperation and Development, Paris, 1970.

37. Walker, W.W., "Techniques for Parameter Estimation in Nonlinear Dynamic Systems Applied to Two Ecosystem Models", Technical Paper No. 750512, Onondaga Lake Modelling Project, Environmental Systems Program, Harvard University, 1975.

42. Water Resources Engineers, Inc., "Mathematical Models for the Prediction of Thermal Energy Changes in Impoundments", Federal Water Quality Administration, Water Pollution Control Research Series 16130 EXT 12169, December 1969.

43. Williams, J., _Introduction to Marine Science_, Little and Brown Co., Boston and Toronto, 1964.

44. Wu, J., "Wind Stress and Surface Roughness at Air-Sea Interface", _Journal of Geophysical Research_, Vol. 74, pp. 444-455, January 1969.

5.0 OVERVIEW

The previous chapters have demonstrated the use of a variety of quantitative techniques for assessments of lake water quality problems. These techniques have been generally characterized in Chapter 1. Rather than reiterate here the specific conclusions already expressed at the end of each chapter, it would be of value to summarize what has been learned about the potential roles and limitations of these methods and approaches.

Preliminary data analyses have been useful in summarizing important relationships, both in the cross-sectional studies of Chapter 2 and in the spatial and temporal studies of Chapter 3. In Chapter 2, marked stratification of all independent and dependent variables with lake trophic state was observed. In Chapter 3, a variety of temporal associations were evident. For example, the disappearance of blue-green algae occurred simultaneously with the reductions in ambient phosphorus levels, chromium levels, and density gradients. Lake phosphorus concentrations decreased simultaneously or immediately following the implementation of detergent restrictions, combined sewer maintenance programs, and diversion of a raw sewage discharge to primary treatment facilities. Because of the general problem of multicollinearity, either in variables characterizing different lakes or in events occurring in or around Onondaga Lake, data analyses alone have not provided sufficient bases for inferences of causation. At a practical level, we have had to rely upon independent evidence as a basis for functional understanding

5.0 OVERVIEW (continued)

of system behavior. The empirical modelling approach of Section 2.5 would not permit clear separation of the effects of depth and hydraulic residence time upon lake phosphorus dynamics. Some more mechanistic modelling efforts might be helpful in this regard. It is unclear whether current understanding of multi-species algal population dynamics would permit us to attribute the apparent disappearance of blue-green algae from Onondaga Lake to a specific factor. Thus, preliminary data analyses have provided important descriptive information, but not the functional understanding required to predict lake behavior.

Regression techniques for estimating parameters have been applied to linear, nonlinear, and nonlinear/dynamic models. In the last case, the use of nonlinear programming algorithms has demonstrated reasonable potential, subject to economic constraints on model complexity. The variety of assumptions inherent in the use of maximum likelihood criteria for parameter estimation have been discussed in Chapter 1. A difficulty which has been encountered in Chapter 2 and Chapter 4 modelling efforts concerns the possible effects of errors (as bias or variance) in the independent variables on the estimates of the parameter vector and of the parameter and residual covariance matrices. In Chapter 4, observed serial correlation in the residuals was attributed partially to such errors. A Monte Carlo approach to this problem would be feasible only in relatively simple systems. Alternative methods for obtaining prior estimates of such errors are not well-developed. This tends to emphasize

5.0 OVERVIEW (continued)

the importance of independent variable data as bases for modelling efforts. Specification of models, particularly with regard to complexity, should not be done without considering the quality and quantity of available independent variable data. The same comments apply to the importance of parameter estimates which are derived exclusively from prior information.

As a corollary to this, assessments of lake water quality problems can be particularly dependent upon estimates of the quantities of materials entering the lake from various sources. Thus, mass flux boundary conditions represent an important type of independent variable in these analyses. If a key component is identified (e.g., phosphorus), in many instances, management policy recommendations can be based upon accurate estimation and comparison of the various problem sources, without use of a particular lake model, other than the assumption that the particular component is controlling lake water quality. In other cases, lake models may be required to estimate the degree of source control necessary to achieve specific water quality objectives. The estimation and verification of such models depend upon the accuracy of source estimates. Thus, the state-of-the-art of source models, as discussed in Section 2.3, suggests the importance of source monitoring, as well as lake monitoring, in providing an analytical basis.

Error analyses have been useful in assessing model and data

5.0 OVERVIEW (continued)

adequacies and in providing bases for model predictions on probabilistic

terms. In Section 1.5, prediction errors were described as consisting

of three basic components: parameter error, independent variable error,

and residual error. The last was further dissected into model and

measurement error components. In applications, it has been difficult

to obtain complete separation of these components. For example, if it

is estimated by comparing model predictions with system observations,

residual error actually contains an independent variable error component.

Prior estimates of errors in independent variables and in specified

parameters are usually difficult to obtain with much accuracy. Posterior

estimates of the parameter covariance matrix are likewise approximate in

the case of a nonlinear model, particularly if the residuals are auto-

correlated. It must also be considered that the error equation (1.5-2)

is based upon a first-order approximation. Thus, three factors suggest

the approximate nature of error analyses, particularly in cases of com-

plex models: (1) smearing of error components; (2) the approximate

natures of the parameter, independent variable, and residual covariance

matrix estimates; (3) first-order truncation of the fundamental equation.

Despite these factors, comparisons of error sources can still yield

useful information for assessing model and data adequacies, because,

as demonstrated in Sections 2.5.3 and 4.6.3, the terms of the total

error equation often differ by orders of magnitude. The approximate

nature of the total error estimate for a given model prediction suggests

that this second moment should not be relied upon too heavily as a basis

5.0 OVERVIEW (continued)

for probabilistic projections and rational designs, particularly if a relatively complex model is employed.

Generality is a key model attribute. In order to be applied properly in addressing specific management problems in specific natural systems, models should be general enough to be valid both under present and projected states of the systems. Parameter stability criteria have been employed in Chapters 2 and 4 as partial bases for assessments of model generality. The parameters of the phosphorus retention model developed in Chapter 2 were shown to be stable across lake trophic states. While this model was apparently general enough to give unbiased predictions in different types of lakes, its validity for simulating behavior of single lake in time was not substantiated. In Chapter 4, two principal parameters of the vertical stratification model were found to be reasonably stable when estimated independently, based upon data from separate years. However, the model was applied to predict lake behavior under conditions of salt disposal which were substantially different from those under which estimation took place. The theoretical basis of the model was relied upon to permit its use in extrapolating or projecting system behavior beyond previously observed states. Demonstration that the same model and parameter estimates could successfully simulate behavior of a different lake would have provided stronger evidence for model generality and further justification for its use in a projective mode.

5.0   OVERVIEW (continued)

The qualified successes of empirical approaches have shown that
average water quality conditions in a cross-section of lakes in the
same geographical region can be associated with such factors as
phosphorus loading, mean depth, and hydraulic residence time. However,
the accuracy and validity of these models for use in predioting response
of a given lake to changes in the associated factors have yet to be
adequately determined. As discussed in Sections 2.4.2 and 2.6, theore-
tical ecosystem models have not as yet demonstrated wide generality for
use in eutrophication assessments, as evidenced by apparent variations
in parameter values appropriate for simulating different lake ecosystems.
This suggests inadequacies in the data and/or models and indicates that
our ability to project the detailed behavior of a given lake ecosystem
much beyond previously observed states is still rather limited.

Based upon the weaknesses inherent in these analyses, it would
seem that modelling efforts should strive to demonstrate generality
both among lakes and along temporal dimensions within lakes. This
suggests that time series data from more than one lake should serve
as a basis for model estimation and verification. The success of such
an approach would depend upon the availability of adequate data and
upon the feasibility of expressing essential functional relationships
in concise terms, based upon current understanding of lake ecosystems.
Applied to key parameters, the estimation techniques demonstrated above
would partially eliminate the necessity of having to use parameter

5.0 OVERVIEW (continued)

values derived from laboratory experiments, the results of which are often of limited validity under field conditions. The stability of optimal parameter estimates along temporal dimensions and among lakes could be used as a partial basis for assessment of model generality. Demonstration of parameter stability would help to insure that the estimation step has been more than a curve-fitting exercise and that important functional aspects of system behavior have been captured. Comparisons of optimal parameter estimates with ranges of values found in the literature or measured experimentally could provide further evidence. However, it is re-remphasized that demonstration of parameter stability or satisfaction of any criterion for model "verification" in these non-experimental situations would not establish model validity, but only test whether that validity could be rejected. Thus, application of a rigorous battery of tests would reduce the dependence of model evaluation upon subjective judgment, but could not eliminate it entirely.

APPENDIX A — Results of t-Tests for Horizontal Mixing in Onondaga Lake

APPENDIX A (continued)

APPENDIX A (continued)

COMPONENT: CALCIUM
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	39	397.1423	36.3142	408.6970	64.0251	-6.3413	16.9832	-0.5410
2	40	850.4013	27.2243	827.9993	24.5519	-4.3583	16.9852	-0.3773
3	40	745.0331	36.2671	715.1990	25.4132	6.7033	11.4170	0.0442
4	40	812.0348	21.1070	813.1040	30.0046	-1.0006	16.5405	-0.1102
5	40	890.3413	26.3470	880.3401	31.3010	-6.1013	20.3044	-0.4343
6	40	843.0147	33.7453	912.3700	32.6010	-6.1666	21.9104	-0.3732
7	40	835.0117	34.2500	833.5402	35.3340	21.3314	19.6901	1.4126
8	40	460.3110	36.2112	675.1321	36.4403	-4.0531	11.3175	-0.3361
9	40	905.0363	43.0070	903.1343	31.4100	3.7033	18.3043	0.0973

COMPONENT: ALKALINITY
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	22	171.6142	2.6057	162.4164	2.2790	0.0722	2.0067	1.0041
2	23	149.0321	2.5516	143.6430	2.7612	3.1609	1.7750	1.7010
3	23	161.0114	2.6011	144.0274	2.3714	1.0453	1.4004	0.7072
4	23	172.3042	2.5147	170.0041	2.1163	1.4521	1.7411	0.6711
5	23	165.0153	2.6432	151.7000	2.7342	-0.6042	2.5471	-0.2370
6	23	157.0561	3.4505	162.1907	3.4102	0.7307	2.6519	0.2601
7	22	201.1750	3.4371	204.0730	3.4204	3.2500	3.6450	0.6996
8	23	169.2204	3.4470	164.0075	2.5621	4.4201	1.1724	3.2043
9	23	194.0056	2.5516	192.6756	2.6570	3.7351	1.3750	2.1156

COMPONENT: PH
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	42	7.2010	0.6410	7.2253	0.0446	-0.1009	0.0329	-3.0704
2	43	7.3444	0.0466	7.7710	0.0503	-0.1370	0.0574	-2.7042
3	42	7.3452	0.0166	7.0022	0.0331	-0.0712	0.0235	-2.0122
4	42	7.4517	0.0231	7.5043	0.0301	-0.0153	0.0222	-3.0001
5	42	7.6112	0.0232	7.6013	0.0270	-0.0092	0.5144	-2.1902
6	42	7.3452	0.0302	7.4122	0.0234	-0.0704	0.0100	-1.0560
7	32	7.6243	0.0330	7.6250	0.0334	-0.0047	0.0194	-0.2400
8	42	7.6532	0.0446	7.7366	0.0461	-0.0971	0.0243	-3.3331
9	42	7.1516	0.0204	7.6300	0.025A	-0.0120	0.0140	-2.7040

COMPONENT: IRON
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	33	0.2267	0.0203	0.2009	0.0171	0.0119	0.0231	0.1660
2	37	0.1501	0.0113	0.1972	0.0140	-0.0011	0.0137	-0.0014
3	37	0.1503	0.0143	0.1917	0.0122	0.0044	0.0132	0.3030
4	37	0.1732	0.0107	0.1904	0.0205	-0.0144	0.3122	-1.1050
5	37	0.2232	0.0214	0.2614	0.0273	-0.0102	0.0092	-2.0764
6	37	0.0310	0.0144	0.2410	0.0170	-0.0200	0.0130	-1.4071
7	36	0.2243	0.0134	0.3242	0.0352	-0.0103	0.0344	-2.0017
8	37	0.2243	0.0121	0.2902	0.0137	0.0091	0.0130	0.1004
9	37	0.2271	0.0194	0.2444	0.0197	-0.0143	0.0104	-1.1732

COMPONENT: CRACKING
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	41	0.0203	0.0044	0.0322	0.0030	-0.0019	0.0027	-0.0654
2	42	0.0322	0.0047	0.0332	0.0034	0.0080	0.0027	0.0174
3	42	0.0320	0.0025	0.0320	0.0031	0.0010	0.0010	0.3740
4	42	0.0210	0.0059	0.0319	0.0091	-0.0012	0.0023	0.3204
5	42	0.0304	0.0010	0.0323	0.0004	-0.0019	0.0014	-1.1500
6	42	0.0214	0.0044	0.0311	0.0047	-0.0024	0.0016	-1.1744
7	16	0.0229	0.0009	0.0304	0.0031	-0.0007	0.0020	-2.1041
8	42	0.0221	0.0054	0.0323	0.0034	-0.0002	0.0019	-0.1374
9	42	0.0270	0.0097	0.0244	0.0034	-0.0017	0.0012	-1.4100

COMPONENT: COPPER
PAIRED T-TEST FOR SIGNIFICANT DIFFERENCES BETWEEN STATIONS AT EACH DEPTH

LEVEL	N	AV1	SDN	AV2	SDP	AV1-AV2	SDN	T
1	37	0.0442	0.0009	0.0501	0.0043	-0.0018	0.0046	-0.3763
2	36	0.0444	0.0047	0.0453	0.0046	0.0033	0.0041	0.7750
3	36	0.0452	0.0017	0.0440	0.0034	0.0034	0.0027	2.0009
4	30	0.0412	0.0040	0.0443	0.0038	0.0149	0.0044	3.3713
5	36	0.0450	0.0037	0.0443	0.0032	0.0027	0.0039	0.7027
6	36	0.0431	0.0079	0.0435	0.0070	0.0022	0.0043	0.7050
7	27	0.0411	0.0045	0.0501	0.0034	0.0110	0.0014	1.4106
8	36	0.0340	0.0054	0.0310	0.0034	0.0023	0.0023	0.7050
9	36	0.0419	0.0094	0.0445	0.0055	0.0034	0.0024	1.4174

APPENDIX B - GRID Displays of Onondaga Lake Water Quality Data

TEMPERATURE

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : TEMPERATURE (DEGREES C)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.00 AND 29.00 MEAN = 10.10 ST. DEV. = 6.57

TEMPERATURE GRADIENT
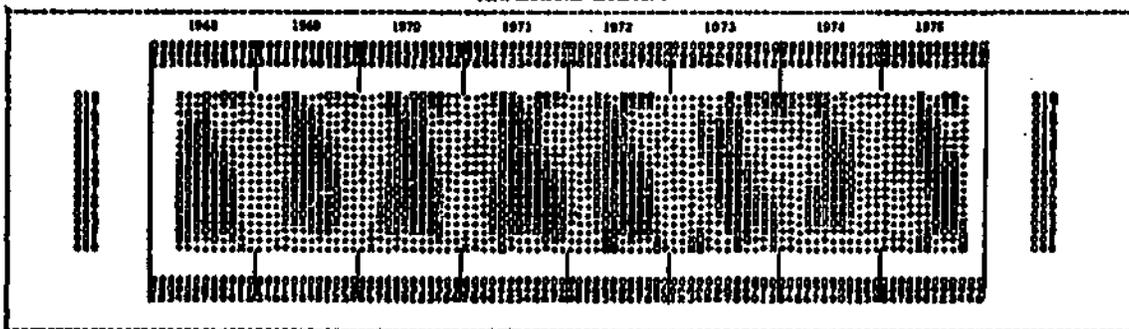
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : TEMPERATURE GRADIENT (DEG-C/M)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.0 AND 1.00 MEAN = 0.33 ST. DEV. = 0.29

APPENDIX B (continued)

CHLORIDE
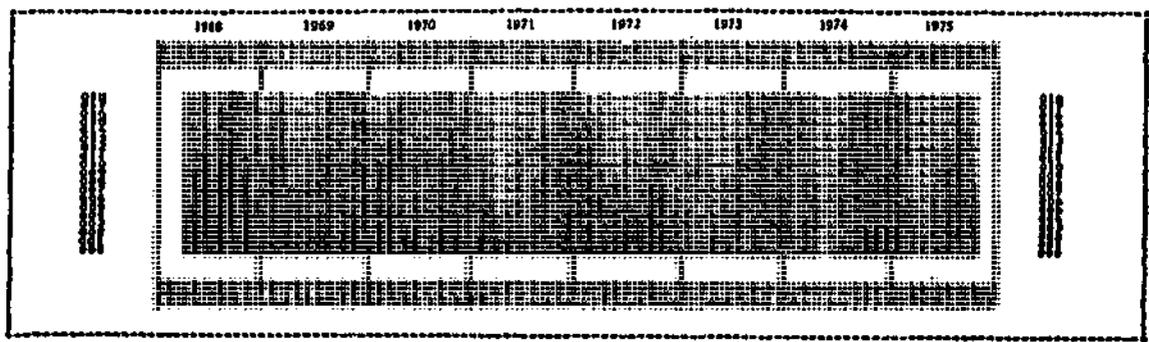
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : CHLORIDE (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 1000.00 AND 2200.00 MEAN = 1610.26 ST. DEV. = 344.35
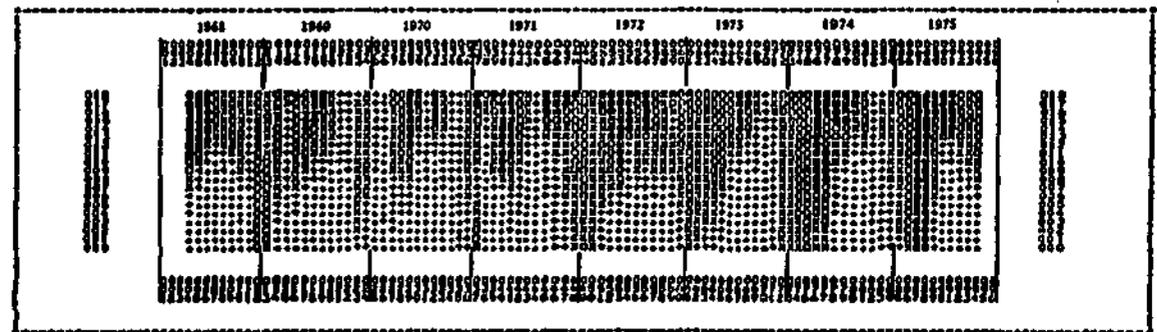
DISSOLVED OXYGEN

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : DISSOLVED OXYGEN (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 1.00 AND 9.00 MEAN = 3.49 ST. DEV. = 3.09

ORTHO-PHOSPHORUS

| 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |
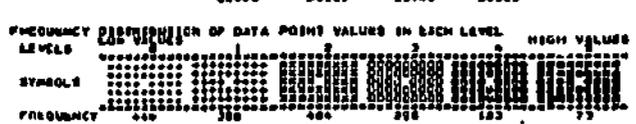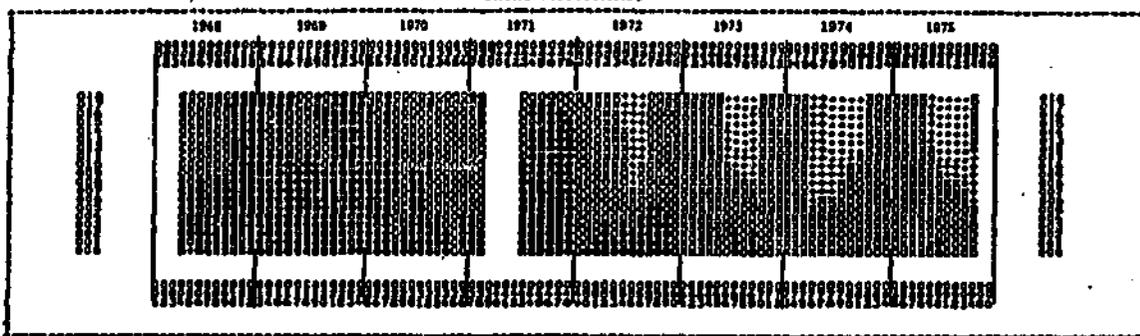
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : ORTHO-PHOSPHORUS (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.10 AND 1.50 MEAN = 0.73 ST. DEV. = 0.07

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL

LEVELS · · · · · · · · · · · · · · · · · · HIGH VALUES

SYMBOLS

FREQUENCY

TOTAL INORGANIC PHOSPHORUS

| 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |
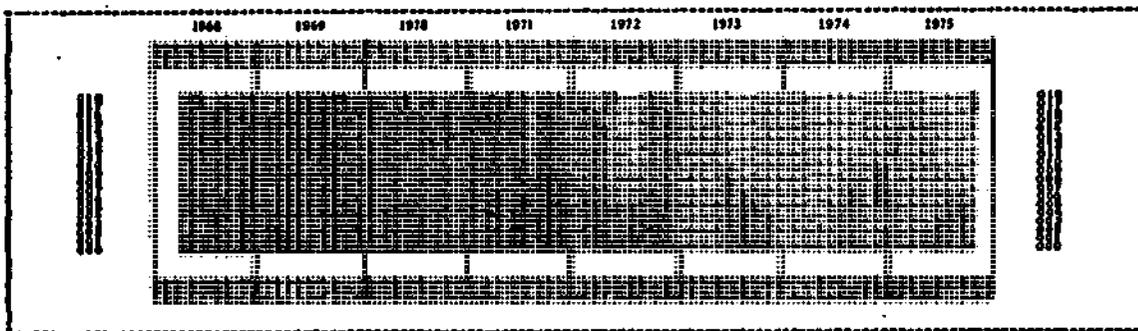
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : TOTAL INORGANIC PHOSPHORUS (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.50 AND 4.00 MEAN = 1.40 ST. DEV. = 1.22

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL

LEVELS · · · · · · · · · · · · · · · · · · HIGH VALUES

SYMBOLS

FREQUENCY

NITRATE NITROGEN

| 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |
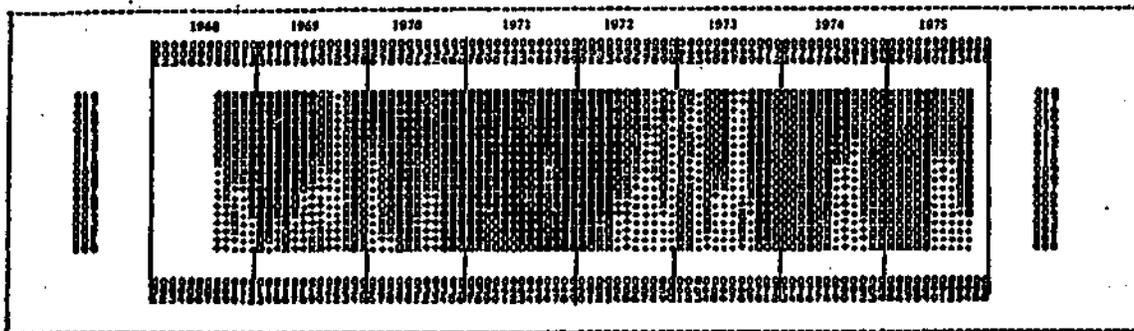
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : NITRATE NITROGEN (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.02 AND 0.38 MEAN = 0.24 ST. DEV. = 0.06

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL HIGH VALUES

LEVELS

SYMBOLS

FREQUENCY

NITRITE NITROGEN

| 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |
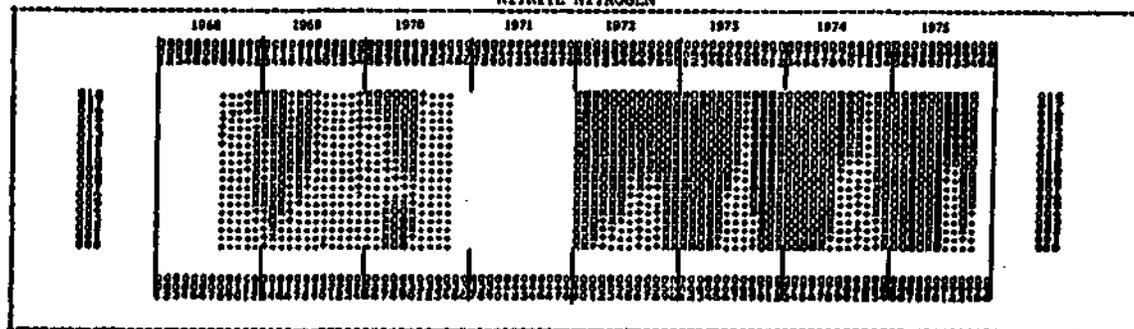
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : NITRITE NITROGEN (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.02 AND 0.35 MEAN = 0.06 ST. DEV. = 0.11

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL HIGH VALUES

LEVELS

SYMBOLS

FREQUENCY

AMMONIA NITROGEN

UNONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (16 = SURFACE)

COMPONENT : AMMONIA NITROGEN (MG/L)

ORGANIC NITROGEN

UNONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (16 = SURFACE)

COMPONENT : ORGANIC NITROGEN (MG/L)

SILICATE

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : SILICATE (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.00 AND 8.00 MEAN = 4.73 ST. DEV. = 1.55
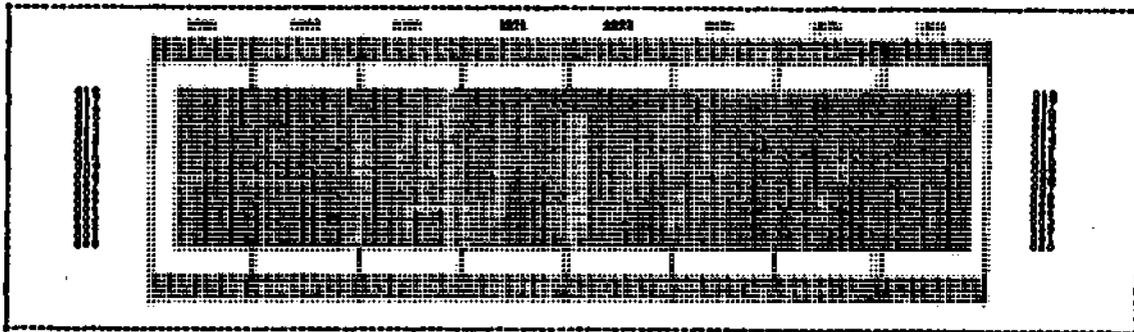
S-DAY BOD

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (10 = SURFACE)

COMPONENT : 5-DAY BOD (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 1.00 AND 10.00 MEAN = 6.26 ST. DEV. = 6.68

APPENDIX B (continued)

CALCIUM
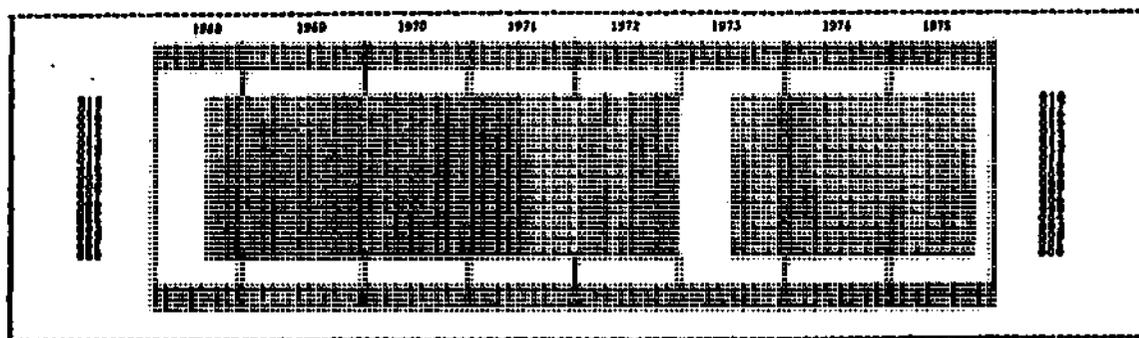
| | 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |

ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (16 = SURFACE)

COMPONENT : CALCIUM (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 256.00 AND 1200.00 MEAN = 673.00 ST. DEV. = 316.00

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL   HIGH VALUES

LEVELS

SYMBOLS

FREQUENCY   48   836   482   374   148   166

COPPER

| | 1968 | 1969 | 1970 | 1971 | 1972 | 1973 | 1974 | 1975 |
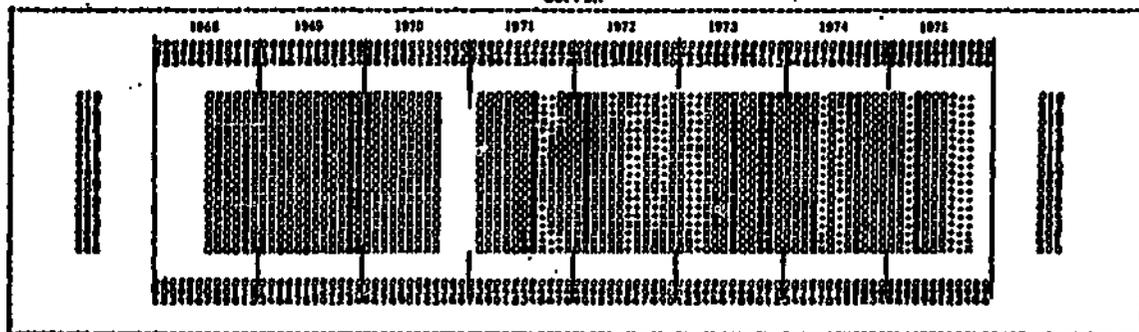
ONONDAGA LAKE DATA PLOT - ENVIRONMENTAL SYSTEMS PROGRAM - HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (16 = SURFACE)

COMPONENT : COPPER (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.01 AND 0.16 MEAN = 0.08 ST. DEV. = 0.04

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL   HIGH VALUES

LEVELS

SYMBOLS

FREQUENCY   116   387   636   616   131   41

APPENDIX B (continued)

CHROMIUM
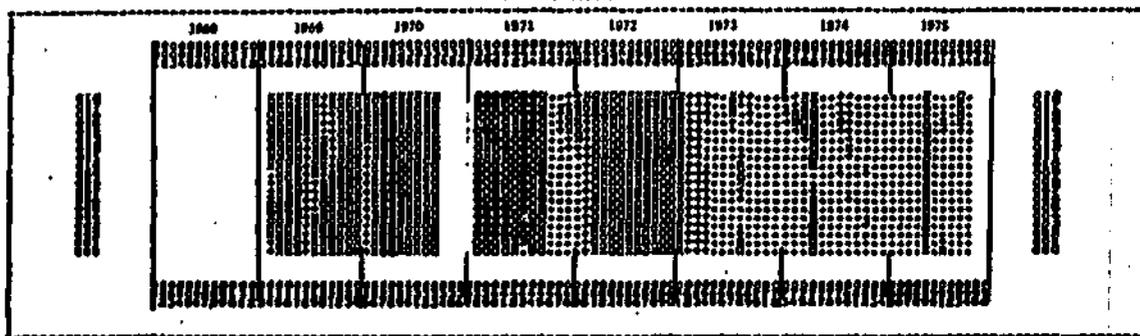
ONONDAGA LAKE DATA PLOT — ENVIRONMENTAL SYSTEMS PROGRAM — HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (18 = SURFACE)

COMPONENT : CHROMIUM (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.01 AND 0.10 MEAN = 0.03 ST. DEV. = 0.04
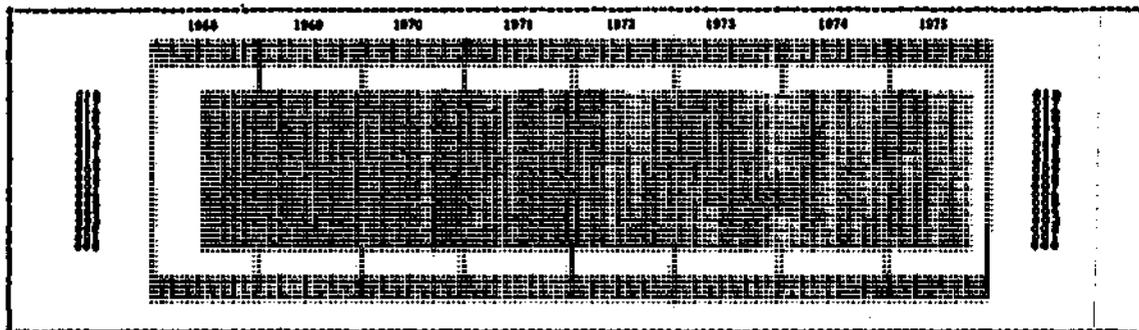
IRON

ONONDAGA LAKE DATA PLOT — ENVIRONMENTAL SYSTEMS PROGRAM — HARVARD UNIVERSITY

HORIZONTAL SCALE : TIME IN MONTHS FROM JANUARY 1968

VERTICAL SCALE : DEPTH IN METERS ABOVE LOWEST SAMPLING POINT (18 = SURFACE)

COMPONENT : IRON (MG/L)

DATA MAPPED IN 4 LEVELS BETWEEN EXTREME VALUES OF 0.05 AND 0.60 MEAN = 0.50 ST. DEV. = 0.10

APPENDIX C - Line Plots of Volume-Averaged Onondaga Lake Water Quality Data
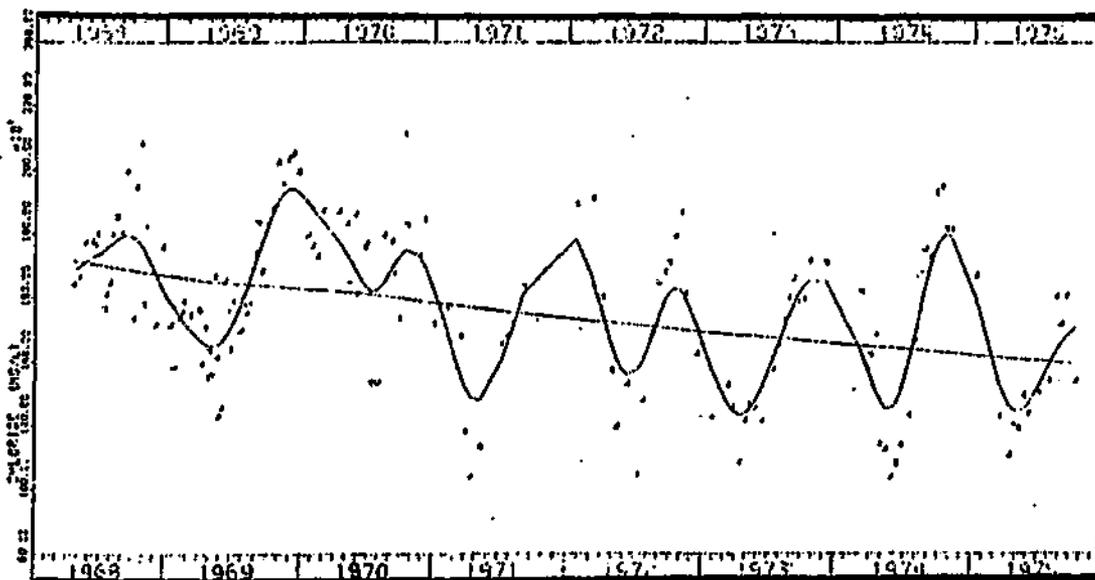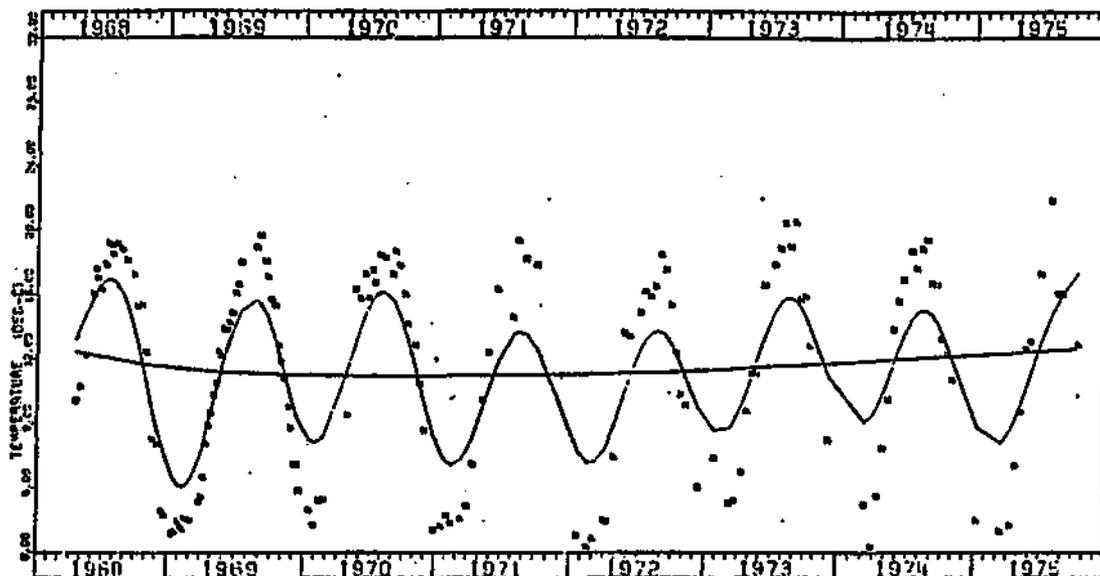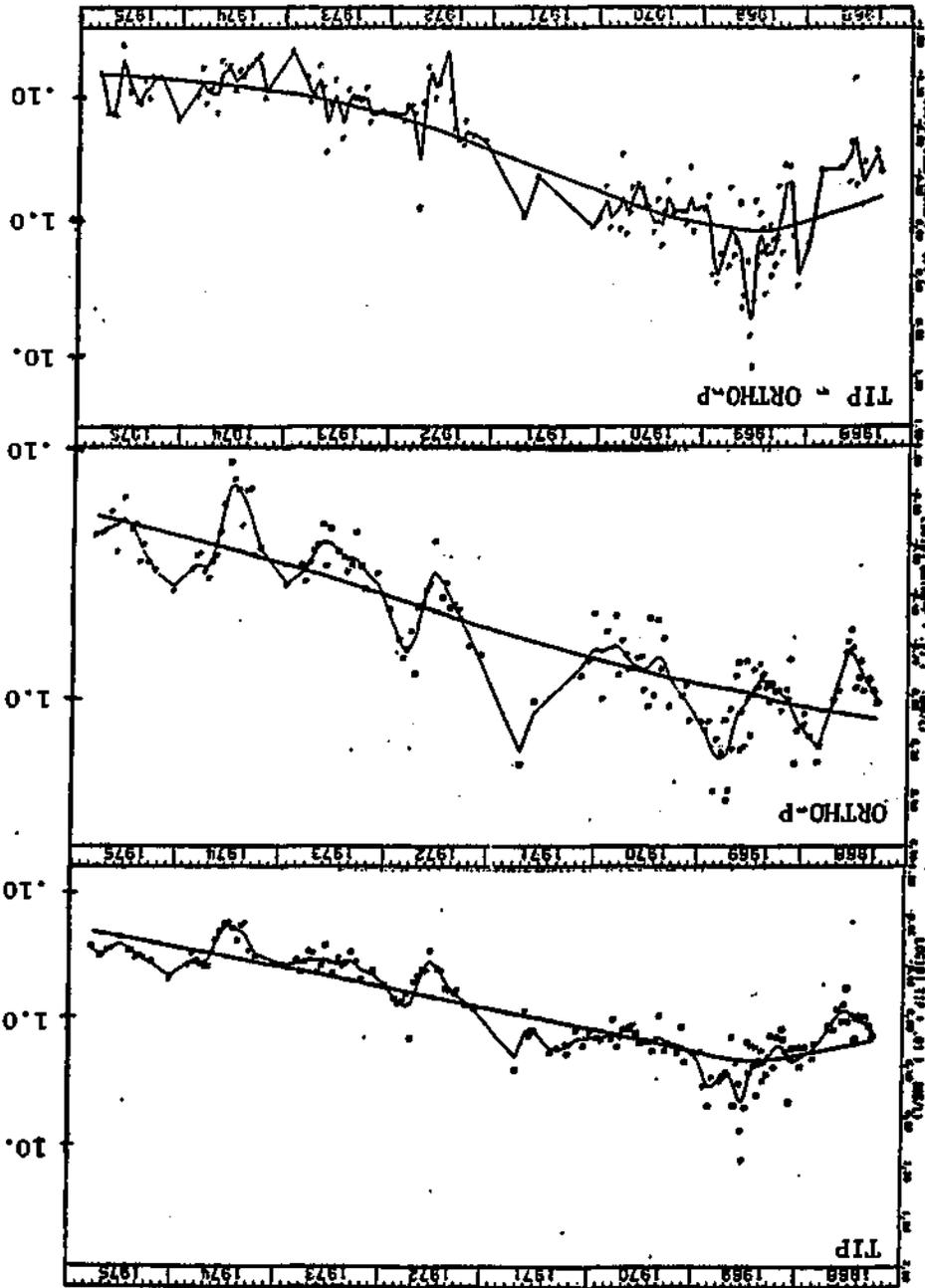
APPENDIX C (continued)
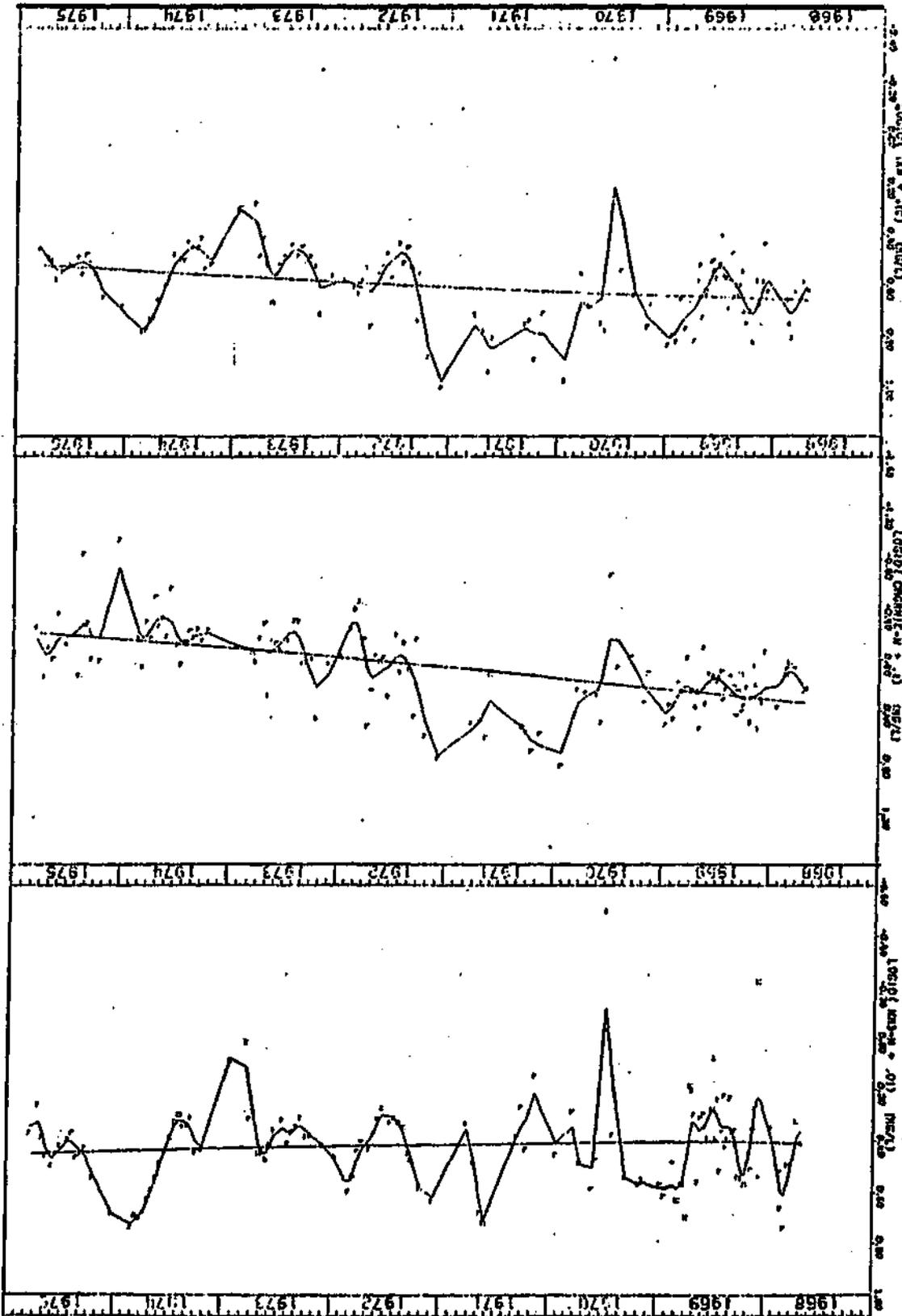
A-13

APPENDIX C (continued)

APPENDIX C (continued)
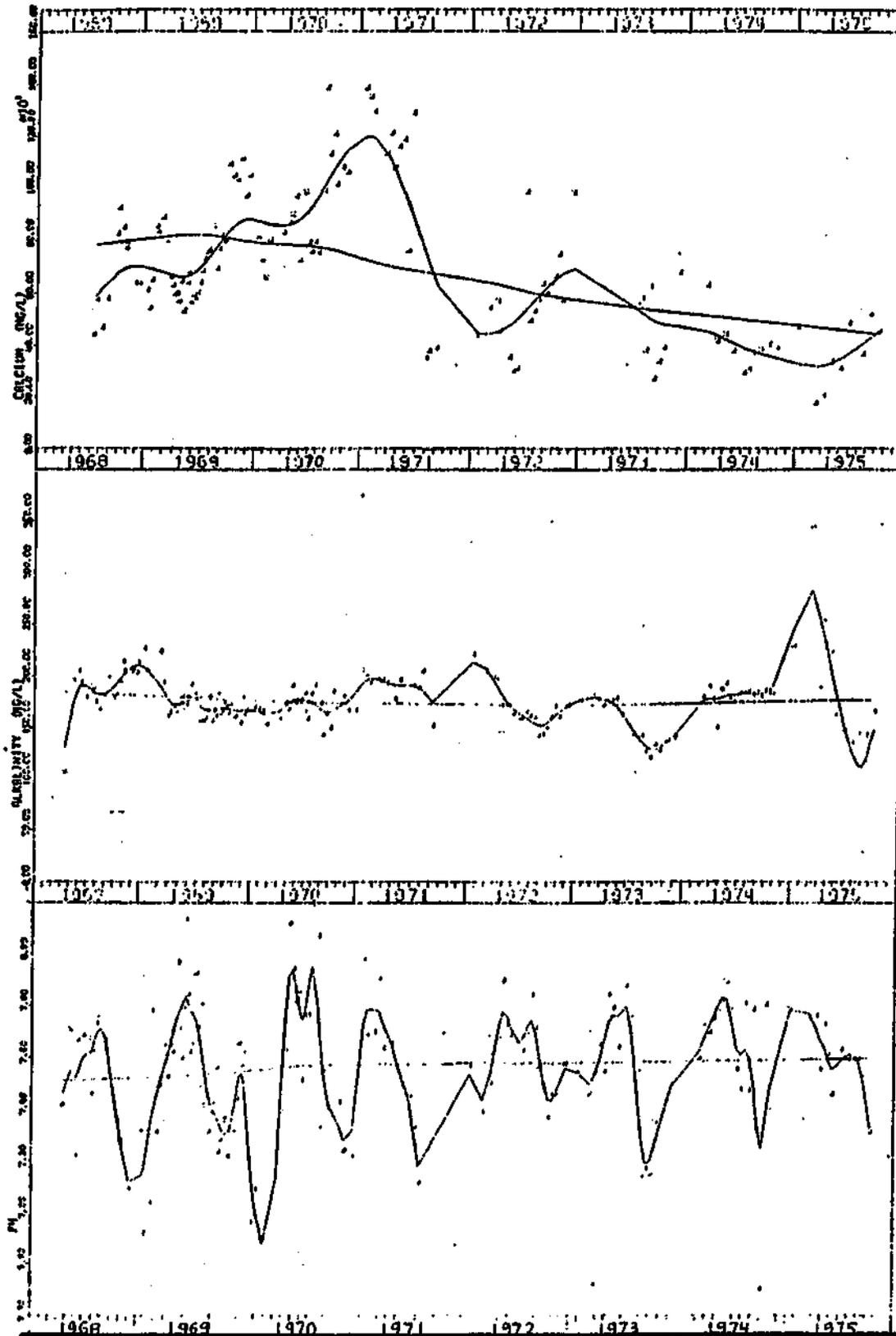
APPENDIX C (continued)
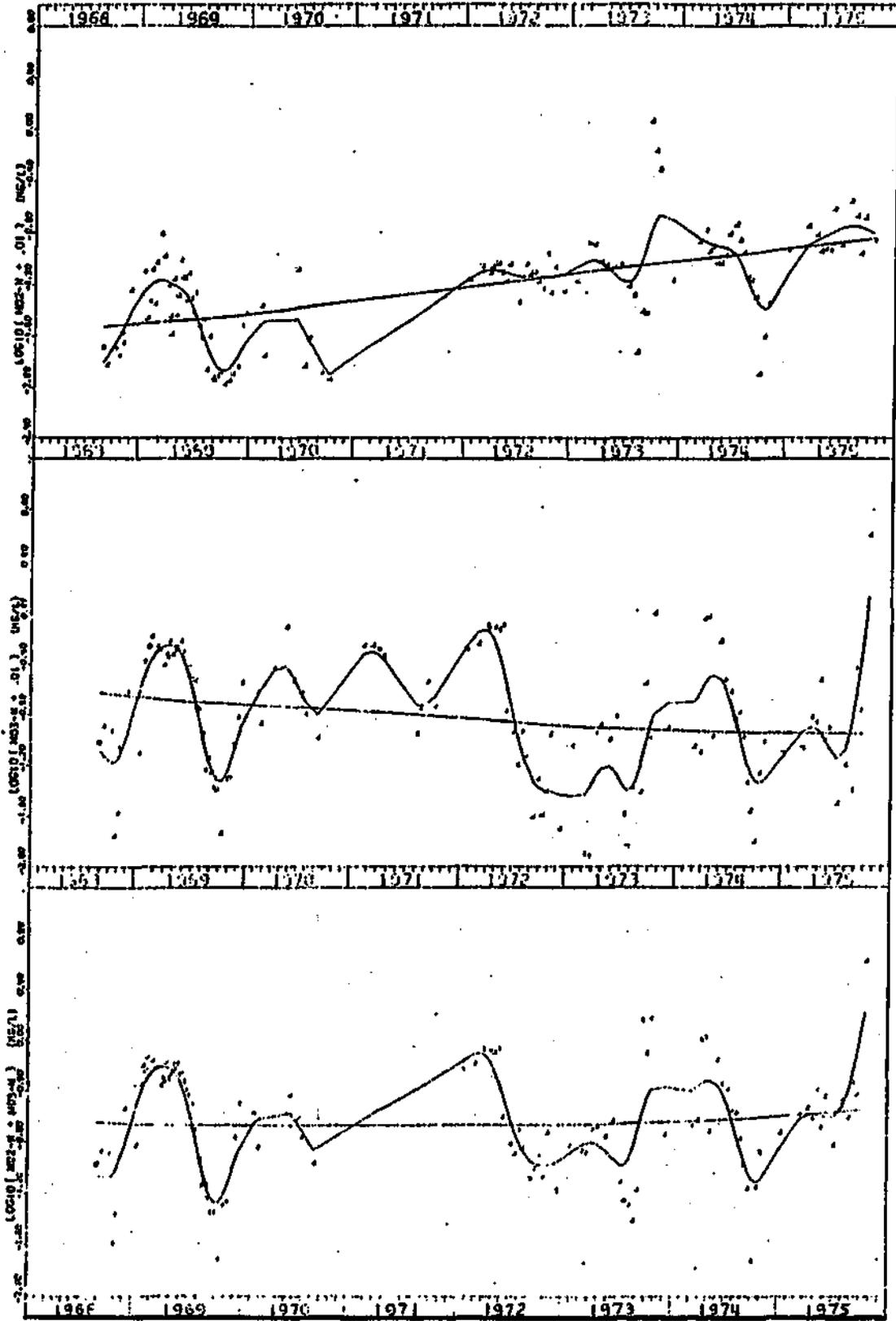
APPENDIX C (continued)
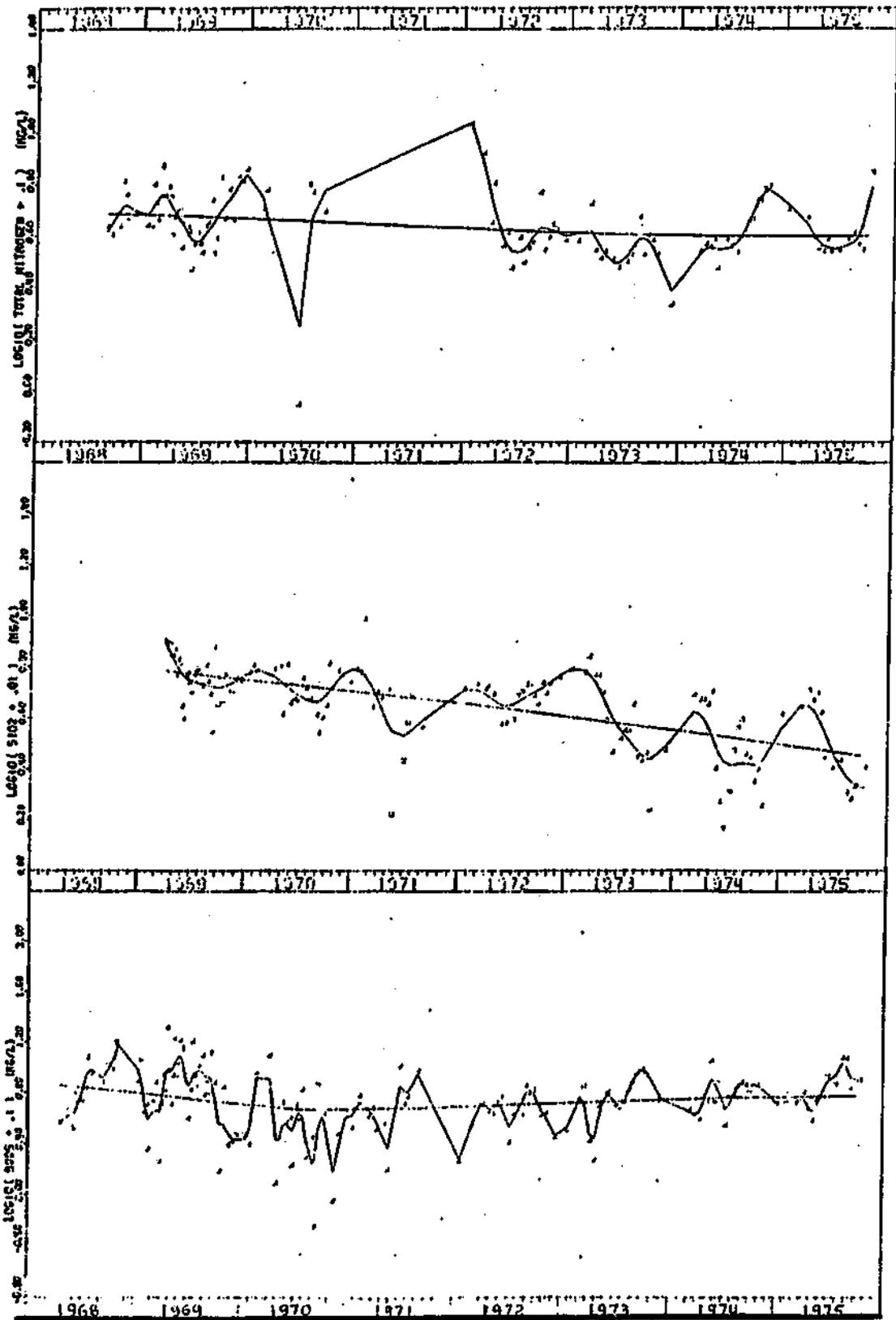
APPENDIX C (continued)
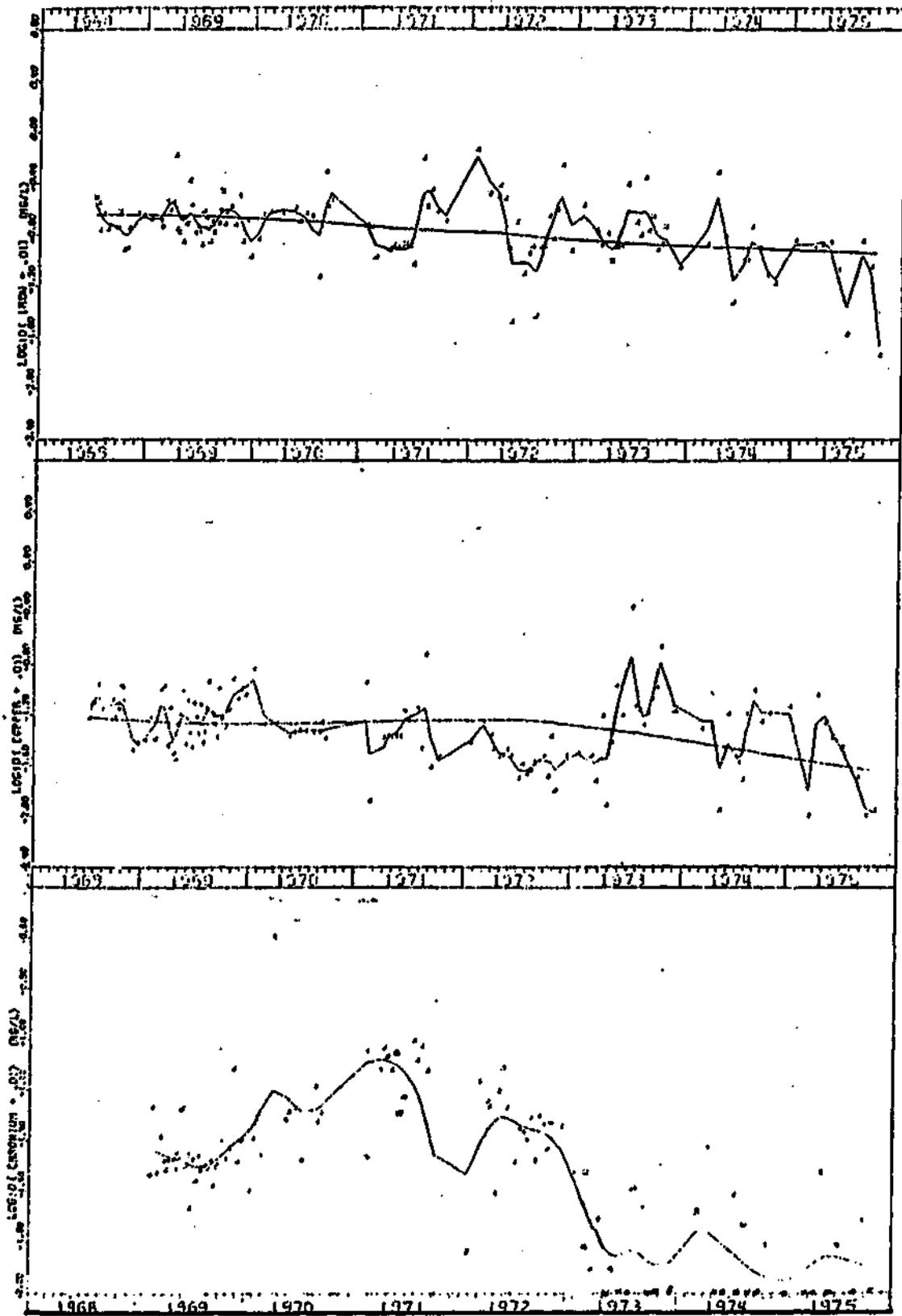
APPENDIX D – Mass Balances on Onondaga Lake

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Hill
8 - Metro S.T.P.
9 - Total Input

FLOW (10^9 M^3)

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

FLOW (10^9 M^3)

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
• - Total Input

ORTHO-PHOSPHORUS (10³ Metric Tons)

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

ORTHO-PHOSPHORUS (10³ Metric Tons)

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CONDENSED INORGANIC PHOSPHORUS (10^3 Metric Tons)

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

CONDENSED INORGANIC PHOSPHORUS (10^3 Metric Tons)

APPENDIX D (continued)

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Loy Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX D (continued)

CUMULATIVE INPUTS

1 - Ley Creek
2 - Onondaga Creek
3 - Harbor Brook
4 - Ninemile Creek
5 - Ungauged Inflows
6 - Allied Chemical
7 - Steel Mill
8 - Metro S.T.P.
* - Total Input

CUMULATIVE BALANCE

I - Total Input
O - Total Output
S - Change in Storage
A - Accumulation

APPENDIX E (continued)