This is to certify that the

thesis entitled

AN EMPIRICAL STUDY OF FACTORS AFFECTING
BLUE-GREEN VERSUS NONBLUE-GREEN ALGAL
DOMINANCE IN LAKES

presented by

JONATHAN TAYLOR SIMPSON

has been accepted towards fulfillment
of the requirements for

Master of Science degree in Resource Development

Kenneth H. Reckhow

Major professor

Date February 21, 1980

O-7639

AN EMPIRICAL STUDY OF FACTORS AFFECTING BLUE-GREEN VERSUS NONBLUE-GREEN ALGAL DOMINANCE IN LAKES

By

Jonathan Taylor Simpson

A THESIS

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

MASTER OF SCIENCE

Department of Resource Development

1979

ABSTRACT

AN EMPIRICAL STUDY OF FACTORS AFFECTING BLUE-GREEN VERSUS NONBLUE-GREEN ALGAL DOMINANCE IN LAKES

By

Jonathan Taylor Simpson

In many lakes, the use and enjoyment of the water is limited due to the dominance of undesirable blue-green algae. Exploratory data analysis techniques were applied to 90 north temperate lakes included in the EPA National Eutrophication Survey to examine empirical relationships between: 1) the chemical and physical variables that affect algal dominance in lakes; and 2) the dominant algal type.

Single variable box plots and bivariate-discriminant plots document the importanct of the inorganic nitrogen concentration and hydraulic detention time in determining blue-green versus nonblue-green algal dominance in eutrophic lakes. The multivariate statistical technique of discriminant analysis was applied to 68 high alkalinity lakes in the data set to: 1) further identify variable relationships; and 2) construct a simple predictive model for algal dominance. Application of results are discussed in an ecological and management context.

ACKNOWLEDGMENTS

This research was supported by funds provided by the United States Department of the Interior, Office of Water Research and Technology as authorized under the Water Resources Research Act of 1964, as amended.

I wish to thank committee members Dr. Kenneth H. Reckhow, Dr. Darrell L. King, Dr. Eckhart Dersch and Dr. Clarence D. McNabb for their comments and criticisms regarding this work. In addition I would like to acknowledge the contributions made by Mrs. Janine Niemer (typing) and Mr. J. Paul Schnieder (graphics consultant).

A special note of thanks must go to my friends and fellow graduate students Craig N. Spencer and Michael N. Beaulac for their professional and personal concern throughout these past two years. I look forward to a lifetime of interaction with them.

Particular gratitude is extended to my major professor, Dr. Kenneth H. Reckhow, and to Darrell L. King. Dr. King, with his unique spirit, was unselfish with his time and patience and provided much of the limnological insight that made up the "glue" for this paper. Academic advisor, instructor, co-researcher, editor and personal friend describe the roles Dr. Reckhow played during my graduate career. His guidence was an essential element in this research from beginning to end. The contributions he made are proudly acknowledged.

Finally, I'd like to thank my parents, Mr. & Mrs. John T. Simpson for their love and support throughout my college years. It is to them, that I dedicate this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE | Page

LIST OF ABBREVIATIONS

EPA-NES -- Environmental Protection Agency-National Eutrophication Survey

$CaCO_3$ -- Calcium carbonate

CO_2 -- Carbon dioxide

SPSS -- Statistical Package for the Social Sciences

U.S. -- United States

MEASUREMENTS

cms -- cubic meter per second

cm -- centimeter

^oC -- degrees centigrade

g -- gram

km -- kilometer

l -- liter

m -- meter

μg -- microgram

mg -- milligram

yr -- year

CHAPTER I

INTRODUCTION

Problem Statement

In recent times there have been increased demands placed on our nation's lakes, streams, and reservoirs as recreation centers, and as sources for domestic, industrial and agricultural water supplies. These demands are due to: 1) the increased wealth, mobility and leisure time of our growing population; and 2) dwindling supplies of groundwater available convenient to the population centers. However, the subsequent introduction of cultural discharges to a lake in the form of untreated or inadequately treated industrial and municipal wastes, agricultural and urban runoff, and septic tank leachate has often resulted in serious water quality deterioration.

Barlowe (1976), states that "effective resource planning and policy-making calls for broad comprehension of the relevant information concerning resource situations, the problems that exist or are expected to arise and the possible solutions of these problems." In addition to purely physical and biological knowledge, an holistic approach to resource problems and policy necessarily includes tests for economic and institutional feasibility as well. Thus, a lake restoration program must be economically acceptable to the affected parties as well as meeting the rational goal of anticipated benefits equaling or exceeding expected costs. Institutional acceptability requires the examination of the legal, political and social constraints within the bounds of administrative workability. From this perspective, water quality problems can become very complex and demanding to those working within the field.

It is within this context that the need for adequate information to make sound decisions is essential if we are to realize the main objective of proper resource development, that is, "policy which enables a nation to provide citizens with opportunities for obtaining high levels of life both in the near and more distant future" (Barlowe, 1976). In addition, this information must be furnished at a reasonable cost to satisfy those who must pay for it.

The Federal Water Pollution Control Act of 1972 (Public Law 92-500) established as a national goal, restoration and maintenance of the chemical, physical, and biological integrity of the nation's waters. In response to the Federal commitment for comprehensive national, regional, and state water management practices, the United States Environmental Protection Agency originated the National Eutrophication Survey (EPA-NES). The purpose of the survey was to develop information on select freshwater lakes and reservoirs so that the problems of water quality can be more adequately addressed.

It is important to be aware, however, that water quality itself may be viewed from many perspectives, according to the desires, uses and goals of the interested population. For example, a bass fisherman's designation of a "quality" lake may be altogether different from the designation assigned to the same lake by a public health specialist.

In other words, levels of water quality may be defined by many quantitative and qualitative terms. Nutrient concentrations, algal biomass, bacteria pollution, blue-green algal dominance, concentration of suspended sediment, loading of organic matter, oxygen depletion, and toxic pollution all may be considered a measure of water quality. Each definition can represent a real problem to those affected by it.

Vollenweider (1968) points out the difficult, yet important, task of separating the problem of eutrophication from other problems of water pollution. He defines eutrophication (in the true sense) as a term that may be applied to anything, including both external and internal sources, which plays a part in: 1) accelerating nutrient loading; and 2) increasing (ambient) nutrient levels and water productivity to the point that nuisance conditions exist.

Vollenweider's definition places emphasis on the causes of eutrophication, namely increased nutrient enrichment by cultural sources. However, nutrient enrichment per se carries no inherently clear message to lakeside property owners, policy-makers, zoning boards, etc. What is much clearer to the non-professional are the obnoxious symptoms (e.g., algal blooms) and effects (e.g., lower property values around the lake) exhibited by increased eutrophy in a lake.

An "abundance of primary production" is perhaps the best elucidation of the term "eutrophication". However, in-lake restoration techniques differ according to the type of eutrophic problem a lake faces. King (1979) identified three basic nuisance problems that may be found within an eutrophic lake:

1) an abundance of algae and/or the dominance of an undesirable algal-type;

2) an abundance of macrophyte growth;

3) an undesirably low concentration of dissolved oxygen in all or parts of the lake.

Although each of the above problems are either directly or indirectly related to primary production (stimulated by nutrient loading), in order to optimize in-lake restoration strategies, it is useful to treat each of the listed problems as a separate management problem worthy of investigation.

It is the purpose of this paper to concentrate on the dominance of an undesirable algal type as a basic nuisance problem.

Blue-Green Algae: An Undesirable Algal-Type

Palmer (1962) states that of the approximately 18,000 algal species identified, only a small number are notable nuisance species. In particular, the dominance of blue-green algae in a lake is a very visible and objectionable sign of eutrophic conditions. Blue-greens are prokaryotic in cell structure and thus resemble bacteria in many respects. This characteristic and their relatively large size and/or organizational type (i.e., coenobial or filamentous) make them undesirable as a food source to zooplankton and other higher organisms. Certain common blue-green genera such as Anabaena, Microcystis, Aphanizomenon and Oscillatoria are buoyant due to pseudovacuoles and may collect in large, unsightly mats or "blooms". Their death and subsequent decomposition on the lake surface produces a distinct "septic" ordor which detracts from a variety of lake uses. Most species of bluegreens are also notorious slime producers and some genera such as Anabaena and Oscillatoria are troublesome filter cloggers (Palmer, 1962).

Of all the species of freshwater algae, only the blue-greens exhibit toxic properties (Palmer, 1962). Animal ingestion of some blue-greens (or their toxic by-products) reportedly resulted in prostration and convulsions followed by death. Incidents of animal poisoning have been documented in many states including Michigan (Stewart et al., 1950). It is very possible many similar cases go unreported because of unfamiliarity with blue-green toxicity.

Among the selection advantages possessed by the blue-greens, as compared to other freshwater algae, included an ability to function at high light intensity and temperatures (Jackson, 1965; Fogg, 1965) and at low free carbon dioxide concentrations (King, 1970). There is also evidence that some blue-greens can chemically inhibit growth of other algae (Boyd, 1973). An additional advantage afforded some species of blue-greens is the ability to fix elemental nitrogen (i.e., transforming it into a biologically useable form) (Dugdale and Neess, 1961).

The above mentioned physiological advantages coupled with buoyancy and their incompatability with the traditional food chain enhances the likelihood that blue-green algae may become the dominant algal type in culturally impacted lakes. Because of the obnoxious and toxic properties exhibited by most species of blue-green algae, dominance of this particular algal type in a lake may be defined as a symptom or index of poor water quality. The research contained in this paper hinges on this definition and thus the desirability to develop management strategies that affect blue-green algal dominance.

Shapiro et. al. (1977) state that environmental manipulation which causes a dominance shift from blue-green algae to other types of freshwater algae has great potential as a lake restoration technique. He predicts that someday "it will be possible to manipulate small or moderate size lakes to improve them. Part of the manipulation may involve converting the population of algae from forms inedible by zooplankton to forms that can be eaten by them. In other words, we consider it possible to bring about a blue-green to green shift in whole lakes."

Obviously, the feasibility of a management strategy that manipulates algal dominance requires careful examination and assessment of factors such as essential and limiting growth requirements, and relative competitive abilities among the algal-types. In general, phosphorus is thought to be the most important factor in stimulating and maintaining eutrophic symptoms. While this statement may be true in terms of total algal biomass, the qualitative makeup of the population is a more complex issue. For example, King (1970, 1972) presents evidence that (especially in lakes undergoing cultural eutrophication) factors such as nitrogen, carbon, and light may be very important in producing at least one eutrophic symptom, the dominance of blue-green algae.

Study Objectives

The objective of this research was to statistically examine data from a cross section of north temperate lakes to identify which factors are most important in determining whether or not a lake is dominated by blue-green algae. Among the variables examined were chemical, and physical parameters found, in theoretical and experimental work, to have an influence on algal dominance. Empirical relationships were studied using exploratory data analysis techniques such as bivariate plotting and correlation analysis. The multivariate statistical technique of discriminant analysis was also used as a tool to describe the relationships among variables, and to develop a predictive model.

Chapter II briefly reviews limnological relationships and conditions that tend to favor the dominance of certain algal types. Chapter III describes the Environmental Protection Agency's National Eutrophication Survey (EPA-NES) which provided the data base used in this study. Included in this chapter are the materials, methods and lake selection criteria used by the EPA and this investigator. Deficiencies in sampling design and technique are noted along with the assumptions used for this investigation.

Chapter IV describes the exploratory data analysis techniques undertaken to familiarize the investigator with the data and variable inter-relationships. This chapter also introduces discriminant analysis as an exploratory aid and as a classification tool. Chapter V applies the data analysis techniques described in Chapter IV to an EPA-NES data set. This chapter provides documentation of the factors that appear to most affect algal-type dominance. Discriminant analysis is then applied to the EPA-NES data to: 1) further define the differences between blue-green dominated lakes and lakes dominated by more desirable algae; and 2) develop a simple predictive model for algal dominance. Chapter VI summarizes the results of this research, within a management context.

Concluding Comments

This research is intended as a heuristic empirical examination of the factors involved in algal dominance using a large and uniformly collected data set (EPA-NES). From this analysis, it is possible to draw at least tentative conclusions regarding a sub-section of the entire cultural eutrophication problem facing lakes in the U.S. In a larger sense, this research represents a different approach to the investigation of algal dominance in lakes. Previous studies of this topic have generally been experimental in nature, and thus based upon small-scale, human-constructed systems. This research, in contrast, is based on unperturbed, natural (in the experimental sense), full-scale lake systems. Each approach has its merits and drawbacks, however, by pursuing both methods, knowledge and mutual corroboration can be gained that ultimately may lead to an understanding of this complex issue.

CHAPTER II

A REVIEW OF LIMNOLOGICAL RELATIONSHIPS AFFECTING ALGAL GROWTH AND DOMINANCE

The Paradox of the Plankton

Even today the "paradox of the plankton"[1] and the apparent deviation from the general theory of competitive exclusion remains an example of the complexity of lake systems. This phenomenon is perhaps not so remarkable, however, if one considers the spatial and temporal heterogeneities that occur in a lake and our inability to sample throughout time and space.

For example, Hutchinson (1961) suggests that the competitive exclusion theory is unsuitable within a dynamic environmental context such as a lake. He proposes that the multitude of external and internal interactions present within a lake system are constantly creating "new" environments that each species must contend with, and these new environments may or may not be optimal habitats for growth or even survival for any particular species.

It has also been established that different plankton species may be found at intermediate depths within the photic zone (Lund, 1965; Moss, 1969; Baker and Brook, 1971). Richerson et al. (1970) suggest

---

[1]The "paradox of the plankton" is a phrase coined by G. E. Hutchinson (1961) to describe the phenomenon of species co-existing in an apparently isotrophic environment where the species must compete for the same limited supply of nutrients.

that the water column is, in fact, a three-dimensional array of habitats. For example, Moss (1972), in the summer stratified waters of Gull Lake, observed stratified populations of Synechococcus sp., Cryptomonas sp., and Rhodomonas sp., within the water column. This phenomenon of species stratification perhaps indicates that competitive exclusion does indeed occur in selected habitats within a lake system.

In the water column of most temperate lakes many species of algae may be found co-existing at any one time, however, there is usually an overall dominance of one or two species (Fogg, 1965). Qualitative changes in dominance occur on a seasonal basis and generally in a fairly predictable pattern from year to year in natural lakes (Wetzel, 1975).

Figure 1 is a simple schematic representation of lake and watershed characteristics that affect algal dominance. The natural base character of a lake (e.g., alkalinity, ionic makeup, base sediment, etc.) is generally a result of the surrounding watershed (King, 1979). The geomorphometry, geology, climate, terrestrial plant development, etc. also determine the natural nutrient loading to the lake. Cultural development within the watershed, and the importation, disposal, and subsequent migration to the lake of nutrients (e.g., contained in sewage, fertilizers, detergents, etc.) represent the artificial load. Seasonal conditions control and modify watershed inputs and internal physical, chemical and biological factors operate to produce dynamic habitats within the water column. The dominant algal type is theoretically the species that can best compete in the niches defined by the habitat(s).

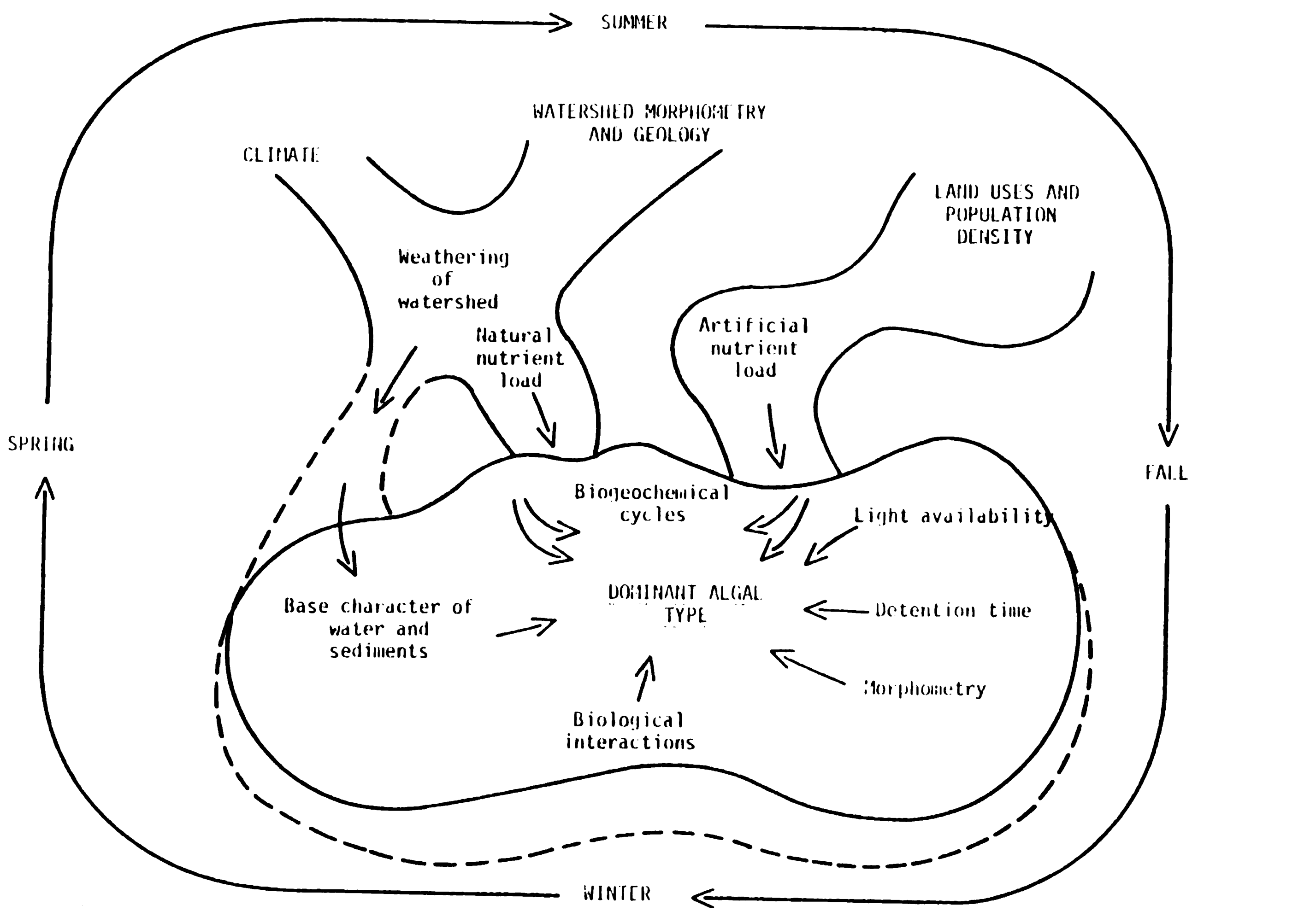

Figure 1.--Factors affecting algal dominance

No two lakes are completely alike by virtue of the uniqueness of natural and cultural characteristics within individual watersheds. Therefore, the generalities implied by summaries like Figure 1 may, on occasion, constrain or otherwise hamper the assessment of lake problems. Suffice to say, the ecosystem approach to lake management should allow for a reasonable amount of flexibility and imagination on the part of limnologists and policy-makers because the solutions to water quality problems are unique to each lake.

An intense review of ecological and physiological conditions necessary and/or optimal for specific algal-types is beyond the scope of this research. However, it would be beneficial to briefly describe the environmental factors thought to be of importance in determining blue-green algal dominance versus nonblue-green algal dominance in lakes.

A classical explanation of the seasonal periodicity of phytoplankton dominance is based on two important physiological factors, light and temperature (Hutchinson, 1967). Solar radiation, of course, is the driving force for all primary production. In the spring, the increased availability of light, coupled with cold waters and renewed nutrient accessability (via overturn and spring runoff) create favorable conditions for diatom growth. Hutchinson (1967) suggests that when the diatom pulse has lowered the phosphate concentration and depleted most of the available silica, the yellow-green algae, Dinobryon sp. are likely to dominate. During the summer and on into the fall, more light and higher water temperatures favor the dominance of green algae, which is followed by, in some cases, blue-green dominance. In the fall when the water cools and overturn occurs, diatoms are again likely to dominate.

This general succession trend has been shown to hold true in many temperate lakes (Birge and Juday, 1922; Rodhe et al., 1958; Lund, 1964; and Moss, 1972). However, causal conclusions based only on light and temperature variations as the determinants of dominance are incomplete. There are at least three important additional factors that determine whether or not a lake is dominated by blue-greens or some other algal-type. These factors are:

1) turbulence and hydraulic detention time;
2) nutrient limitation; and
3) biological mechanisms.

Turbulence and Hydraulic Detention Time

Of primary importance to an algal population is the ability to stay in the photic zone at least long enough to reproduce and maintain the population. When sinking rates increase beyond this equilibrium point, population shrinkage begins. Wind-induced Langmuir circulation and other currents and up-wellings are effective in keeping an individual algal cell within the photic zone. In quiescent waters, however, sinking is usually inevitable due to the fact that cells are heavier than water. Nevertheless, sinking is not without its advantages. Motion provides a constant renewal of dissolved nutrients surrounding the cell, which yields a steeper concentration gradient than if the cell was to remain stable (Munk and Riley, 1952).

The most effective adaption to the problem of sinking is displayed by many species of blue-greens. Gas vacuoles within the cell decrease the cell's density below that of water which allows it to float and thus be no longer dependent on turbulence to remain in the photic zone. Buoyance may be regulated by changing the gas vacuole/cell volume ratio which thereby allows "searches" within the water column for environmental conditions most optimal for growth (Ward and King, 1976; Reynolds, 1972).

Wetzel (1975) notes that the release of extracellular algal products is of much greater importance than realized, in terms of sinking rates and successional dynamics of algal populations. Specifically, excreated cellular organic compounds can function as organic polymers that will effectively flocculate the algal cells and cause them to sink rapidly (Pavonii et al., 1971). Generally, high rates of extracellular leakage, flocculation and subsequent higher rates of sinking occurs during periods of stress. Stress of algae may be theoretically induced by any number of conditions but reasons most often cited include high population density, high or low light intensities and carbon dioxide deficiencies (Ward and King, 1976; Pritchard et al., 1962). Thus, under stressful conditions, such as those listed above, the blue-green algae possess a great advantage since their response to stress does not usually include sinking (due to their buoyant properties).

Flushing via the lake outflow is another potentially important reason for the loss of algal cells. This loss is generally a function of hydraulic detention time and may be a significant factor in determining dominance in some lakes. For example, a lake with a very short detention time (e.g., two or three days) would likely be turbulent and thus dominated by small motile green or attached algae. Blue-green algae, on the other hand, would be more likely to dominate in lakes with long detention times since the water would tend to be less turbulent (which would allow their buoyant properties to become a selective advantage). Note that in some lakes, under heavy rain/flood conditions, flushing may be a very effective population control mechanism for all unattached algal types.

Nutrient Limitation

The nutritional needs of algae vary from species to species, but it is generally agreed that at least 25 or so elements and substances are required for growth. Essential macronutrients include carbon dioxide, oxygen, nitrogen, sulfur, phosphorus, potassium, magnesium and calcium.[1] Essential micronutrients include iron, boron, zinc, copper, molybdenum, manganese, cobalt and sometimes sodium, chlorine and vanadium. Many algal species also require forms of organic nutrients such as thiamine, vitamin B_{12} and biotin.

Each nutrient, of course, has at least the potential to be a limiting factor of algal growth. Cellular metabolic demands for the various nutrients are not equal, however. Table 1 presents the plant demand for various nutrients and the average supply found in world river waters. In most natural waters, especially waters exposed to cultural

---

[1]Silicon is essential for growth of diatoms, some flagellates and perhaps some higher plants. It plays no essential role in the growth of other plants as far as is known (Vallentyne, 1974).

Table 1: Concentrations of essential elements for plant growth in living tissues of freshwater plants (demand), in mean world river water (supply) and the plant:water ratio of concentrations (demand:supply) (from Vallentyne, 1974).

Element	Demand Plants %	Supply Water %	Demand:supply Plant:water (approx.)
Oxygen	80.5	89	1
Hydrogen	9.7	11	1
Carbon*	6.5	0.0012	5,000
Silicon**	1.3	0.00065	2,000
Nitrogen*	0.7	0.000023	30,000
Calcium	0.4	0.0015	1,000
Potassium	0.3	0.00023	1,300
Phosphorus*	0.08	0.000001	80,000
Magnesium	0.07	0.0004	<1,000
Sulfur	0.06	0.0004	<1,000
Chlorine	0.06	0.0008	<1,000
Sodium	0.04	0.0006	<1,000
Iron	0.02	0.00007	<1,000
Boron	0.001	0.00001	<1,000
Manganese	0.0007	0.0000015	<1,000
Zinc	0.0003	0.000001	<1,000
Copper	0.0001	0.000001	<1,000
Molybdenum	0.00005	0.0000003	<1,000
Cobalt	0.000002	0.000000005	<1,000

*Concentrations of carbon, nitrogen, and phosphorus in water are given for inorganic forms only.

**Silicon is essential for growth of diatoms, some flagellates and perhaps some higher plants (Equisetum). It plays no essential role in the growth of other plants so far as is known.

perturbations (in the form of articially added nutrients), it is very unlikely that the elements in low demand relative to supply will limit growth. Instead, it is the nutrients with the greatest demand/supply ratio that are most likely to serve as growth limiters. As indicated in Table 1, the three most important are carbon, nitrogen and phosphorus.

The inducement of nutrient deficiencies is recognized as a sound ecological management strategy for controlling eutrophic conditions. Liebig's Law of the Minimum[1] states the premise under which the focusing on a single nutrient is made an attractive control mechanism. Unfortunately, complications frequently arise when attempting to identify the actual or potential limiting nutrient. This is because the uptake of nutrients and other growth factors occur simultaneously, and therefore limitations are interactive and dynamic. In other words, the concentration at which a particular nutrient is limiting varies with (and according to) the levels of other factors. In addition, some algae exhibit the ability of luxurious uptake and storage of certain nutrients. Blue-green algae in particular, have been known to exist at very low concentrations of inorganic nutrients (Hutchinson, 1967). This phenomenon is presumably due to an ability to accumulate intercellular reserves of nutrients such as phosphorus and nitrogen (Fogg et al., 1973).

---

[1]Liebig's Law of the Minimum: Under steady state conditions, the essential material available in amounts most closely approaching the critical minimum needed will tend to be the limiting material.

In spite of the above mentioned complications phosphorus is usually identified as the most important limiting nutrient in natural water bodies (Sawyer, 1947; Thomas, 1969; Vallentyne, 1974). The close link between phosphorus and eutrophic conditions is evidenced by empirical research relating phosphorus loading and concentration with primary production as measured by chlorophyll _a_ concentrations (Vollenweider, 1968, 1975; Dillon and Rigler, 1975; Schindler, 1978). Although the relationship between phosphorus and chlorophyll _a_ yields high correlation, predictive regression lines derived from empirical phosphorus models still reveal considerable individual lake scatter (Shapiro, 1979). This scatter would indicate that phosphorus concentration should not be the only lake parameter examined when assessing algal problems in a lake. Shapiro (1979) maintains that algal abundance and type is a function of biological factors within the limits imposed by phosphorus, and thus the prediction scatter is due to the variability set forth by other growth factors.

Whenever phosphorus is the most important limiting factor of growth, increasing the availability of phosphorus will generally stimulate algal growth. Correspondingly, there will be an increase in demand for all other growth factors and a rate increase for many biogeochemical cycles which will, in turn, impact all components of the lake ecosystem (Vollenweider, 1968). King (1972, 1979) states that the natural phosphorus limit can be nullified by excessive cultural inputs of phosphorus into a lake and this phenomenon can lead to the establishment of nitrogen and/or carbon limits in the lake system.

The existence of either of these limits will tend to favor the dominance of blue-green algae. That is, blue-greens will usually dominate in lakes exhibiting low inorganic nitrogen or low carbon dioxide concentrations. In order to understand this phenomenon, it is necessary to examine portions of the carbon and nitrogen cycles in more depth.

Under increased growth conditions (stimulated by increased phosphorus availability) the demand for carbon dioxide (CO_2) for photosynthesis can: 1) deplete the supply of CO_2 from biological respiration; and 2) exceed the rate of supply of CO_2 from the atmosphere. These conditions can cause the plants to extract CO_2 from the bicarbonate-carbonate alkalinity of the lake water. This extraction results in the decrease of the equilibrium CO_2 concentration and an increase in the pH (King, 1970).

Pritchard et al., (1962) report that low carbon dioxide concentrations will induce extracellular excretion of organic material by algae. Thus, continued reduction of the CO_2 concentration (from the alkalinity) will eventually cause the more desirable green algae and diatoms to flocculate and sink out of the photic zone. Therefore, the blue-green algae would likely dominate under these circumstances because they remain unaffected by this stress response (i.e., gas vaculoes keep them buoyant). King (1972) proposes that blue-green algae begin to dominate when the CO_2 is reduced to 7.5 μmoles CO_2/l. Other investigations indicate that a shift from green to blue-green dominance occurs at no fixed CO_2 concentration but rather at various points where the CO_2 concentration and light intensity interact (King, 1978; King and Hill, 1978).

It is important to note that the bicarbonate-carbonate alkalinity in a lake serves as a reserve carbon source for photosynthesis. Thus, low alkalinity lakes are more likely to exhibit carbon limiting conditions (e.g., depletion of the carbon reserve from the alkalinity) than high alkalinity lakes. Therefore, a shift to blue-green algal dominance, induced by carbon limiting conditions, is most likely to occur in low alkalinity lakes.

King (1978, 1979) states that various mechanisms of nitrogen stripping can force a lake to a nitrogen limit. For example, in a productive lake, bacterial degradation of plant material will accelerate the nitrogen cycle and increase the frequency that an atom of nitrogen will appear as the ammonium ion (King, 1978, 1979). At the higher pH levels (e.g., generated by the extraction of carbon dioxide from the alkalinity) the ammonium ion is converted to ammonia gas and lost to the atmosphere. Wind induced mixing of the water column will tend to accelerate this ammonia nitrogen loss. In addition, nitrogen loss can also be induced under anaerobic conditions (e.g., caused by the bacterial degradation of plant material) via denitrification and the production and subsequent release of nitrogen gas (N_2).

Blue-green algae that possess heterocysts are associated with the phenomenon of nitrogen fixation (Fogg et al., 1973).[1] Obviously, the ability to fix elemental nitrogen would present a tremendous selective advantage in lakes with a nitrogen limit. Thus, a lake faced with an

---

[1]Heterocysts are specialized cells found in most filamentous blue-green species, with the exception of Oscillatoriaceae (Fogg et al., 1973).

increasing rate of nitrogen loss would also be increasing the likelihood of the dominance of nitrogen-fixing blue-green algae.

Further, one result of nitrogen stripping is a decrease in the lake nitrogen concentration. Thus, the presence of an inadequate supply of nitrogen relative to algal demand may (as in the case of carbon) cause leakage of organic material from the cells, followed by the flocculation and sinking of the non-buoyant algal populations, leaving the buoyant blue-greens (both nitrogen fixing and otherwise). Horne and Fogg (1970) found a correlation between blue-green nitrogen fixing activity and increased concentration of dissolved organic compounds in the water. It is possible that this observed phenomenon was as artifact of the dominance transfer from "leaky" greens to positively buoyant blue-greens.

Biological Mechanisms

The effects of fish predation, zooplankton grazing and parasitism are potentially significant in determining algal species composition (Edmonson, 1972; Hutchinson, 1973; Lund, 1965). For example, Canter and Lund (1948) found that the chytridiaceous fungus *Rizophidium planktonieum* is parasitic on the diatom *Asterionella formosa* (Hass.) in lakes in the English Lake District. They report that during certain times of the year, the fungus causes a parasitic epidemic on the diatom population which greatly reduces the number of living cells in the colony. Another example of parasitism, and of special importance to blue-green populations, is the fact that infections of a number of viruses, cyanophages and myxobacteria can cause lysis of blue-green cells (Shilo, 1971). It is proposed that this method of biological control may become a prevalent strategy in the future in ridding a lake of obnoxious blue-green algal bloom.

The consumption (grazing) of algal cells by zooplankton and herbivorous fish populations can also greatly influence seasonal succession and dominance in a lake due to size and/or species specificity of grazing (Wetzel, 1975). Therefore, competitive advantage is afforded to those species or algal-types that are less effectively grazed. Due to their bacteria-like nature, size organizational type (i.e., coenobial or filamentous) and perhaps some toxic properties, the blue-green algae are very rarely eaten by organisms in the food chain. That is, they have essentially no predators. The advantages in this regard are obvious.

In addition to direct predation, other indirect biological mechanism may also play roles in determining the quality of algae in a lake. For example, Helfrich (1976) tested the impact of zooplanktivorous fish on the algal community. He found that the stocking of fathead minnows in otherwise fish-free ponds led to a significant reduction in the total number of zooplankton. This decrease of zooplankton: 1) reduced the grazing pressure on the algal population; 2) caused an increase in the total density of algae; and 3) ultimately caused a shift in algal composition from one dominated by green algae, diatoms and cryptomonads to one dominated by blue-green algae. As a final note, Helfrich (1976) suggests that "the success of future water quality management strategies oriented toward the regulation of objectional growths of planktonic and filamentous algae requires not only the ability to accurately assess the physiochemical processes involved, but a more complete understanding of and appreciation for biological forces which drive aquatic ecosystems."

Concluding Comments

Much debate centers around the "true" causitive factors of algal dominance, some authors provide suggestions beyond the factors covered in this chapter. For example, Keenan (1973) attributes blue-green dominance to hydrogen ion concentration but suggests the relationship with phosphorus as well. Morton and Lee (1974) demonstrated that the iron concentration may be an important parameter in population shifts. It has been also documented that a shift from blue-greens to greens occurs along with changes in water chemistry, when small lakes and reservoirs are mixed mechanically or with compressed air (Symon, 1969; Bernhardt, 1967; Wirth et al., 1970).

Dr. Joseph Shapiro has been actively researching the subject of green algal versus blue-green algal dominance with the idea that dominance manipulation may become a feasible lake restoration strategy (Shapiro et al., 1977). He has demonstrated in isolated plastic enclosures (in the field) that algal populations may be shifted from blue-greens to greens by adding nitrogen and phosphorus and lowering the pH with CO_2 of hydrochloric acid.

Despite the existance of other evidence, this investigation has concentration on general factors that are most often thought to influence blue-green algal dominance in lakes. Based on the literature review presented herein, it may be concluded that if phosphorus is continuously added to a lake, the phosphorus limit on aquatic plant production will be eventually removed. This condition will most probably lead to carbon and/or nitrogen limits and the dominance of blue-green algae. The rate at which this process occurs is dependent to a large extent on:

1) the alkalinity of the lake water (King, 1979);

2) the cell-loss mechanisms present in the lake (e.g., hydraulic flushing and turbulent water versus quiescent water; and

3) parasitism, grazing and other biological effects.

CHAPTER III

THE NATIONAL EUTROPHICATION SURVEY

Objectives

The Environmental Protection Agency's National Eutrophication Survey (EPA-NES) provided the empirical lake data that were used in this investigation. Since lake survey methods vary greatly in general practice, it is useful to describe the EPA-NES in some detail.

The objective of the EPA-NES was to develop a water quality data file of select freshwater lakes and reservoirs in order to aid national, regional, and state agencies in the formulation of plans and practices for the maintenance and restoration of water quality. For this reason, survey emphasis was placed on the identification of nutrient sources, lake nutrient concentrations and trophic conditions. Between the years 1972 and 1975, EPA-NES personnel collected data from over 800 lakes throughout the continental United States. Lakes included in the survey were jointly selected by state water pollution agency personnel and the EPA Regional Office.

Lake Selection Criteria

The lakes examined in this study are from the northeastern and northcentral regions of the United States and were surveyed by the EPA in the years 1972 and 1973. These lakes were, in most cases, selected for the EPA-NES because of actual or potential eutrophic problems.

Listed below are the general guidelines for lake inclusion in the EPA-NES during the years 1972 to 1973 (USEPA, 1975).

1) Lakes with one or more municipal sewage treatment plants discharging either directly into the lake or into an inlet tributary within approximately 40 kilometers of the lake;

2) Lakes 40 hectares or larger in size;

3) Lakes with a mean hydraulic detention time of at least 30 days.

These guidelines were flexible, however, and if a lake falling outside these criteria presented a special concern to state personnel, it was also included.

Since the above described EPA-NES selection criteria were biased toward lakes exhibiting eutrophic conditions, the survey data must be used with caution. This is especially true when using information derived from the EPA-NES to generalize, assess or describe the entire population of U.S. lakes. At this time, there is no randomly selected large-scale inventory of lakes that may be used to compare the EPA-NES lakes with the "true" population of U.S. lakes.

Three hundred forty-nine lakes from the northeastern and north-central regions were surveyed by the EPA-NES. Examination of the individual lake reports published by the EPA, however, revealed that many of the lakes had incomplete or inadequate data or utilized a poor sampling design. Therefore, for the purposes of this study, it became necessary to develop additional lake selection criteria. EPA-NES lakes that possessed any of the following were excluded from this study:

1) Incomplete assessment of nutrient sources and/or loading estimation;

2) Unnatural conditions that potentially affect lake quality such as dredging, drawdowns or chemical control of algae;

3) Hydraulic detention time of less than three days;

4) No inlets or outlets;

5) Nonrepresentative sampling site selection (e.g., sites within bays rather than open waters);

6) Horizontally non-homogeneous water quality data;

7) Chains of lakes with aggregated data;

8) Shorelines that were swampy and supported excessive aquatic weed growth;

9) Missing physical, biological and/or chemical data.

Of the 349 lakes that were sampled by the EPA-NES in the northeastern and northcentral regions, 90 passed this study's selection criteria described above.

Field Sampling Methods and Analysis

Each EPA-NES lake was sampled three times during the year; in the spring (between March 7 and July 1), in the summer (between July 5 and September 18) and in the fall (between September 19 and November 14). Sampling was conducted by three-man teams using a pontoon-equipped Bell UH-1H helicopter. The selection of individual lake sites were based on a lake's characteristics and known problem areas. The number of sampling sites for a given lake varied "in accordance with lake size, morphological and hydrological complexity, and practical considerations of time, flight range and weather" (USEPA, 1975).

Table 2 summarizes the lake parameters and the methods of sampling utilized by the EPA-NES. Depth, turbidity and temperature were

Table 2: EPA-NES sample analysis summary (modified from USEPA, 1975)

Parameters	Sample Volume	Field Treatment	Where Preformed	Depth
Temperature			In situ	Continuous
Depth			In situ	Continuous
pH	4-ounce	Refrigerated	Field lab	Select levels
Dissolved oxygen	300-ml	Hach chemicals	Field lab	Select levels
Algae identification	4-ounce	Lugol's solution	NERC-LV	Photic zone integration
Total phosphorus	4-ounce	Unfiltered, $HgCl_2$	NERC-LV	Select levels
Nitrite-nitrate-N, Ammonia-N, Total alkalinity	4-ounce	Filtered, $HgCl_2$	NERC-LV	Select levels

measured in-situ using an "Interocean Systems" sensor package, submersible pump and multiconductor cable. Interval sample depths for nutrient, pH, alkalinity and dissolved oxygen were selected after an initial pass of the sensor package through the water column. Nutrient and alkalinity samples were preserved with aqueous mercuric chloride and forwarded to the National Environmental Research Center-Las Vegas (NERC-LV) for analysis. Table 3 summarizes the analytical methods and precision undertaken at NERC-LV.

The mobile field laboratory determined at the end of each sampling day:

1) pH (Beckman Electromate portable pH meter and combination electrode);

2) dissolved oxygen concentration (titration with phenylarsine oxide); and

3) chlorophyll a concentration (flourometric procedure described by Yentsch and Menzel, 1963).

At each sample station, a 130 ml water sample was obtained for phytoplankton analysis. The sample was collected between the surface and a depth of 4.6 meters or at the lower limit of the photic zone (1% of surface incident light), whichever was greater. In shallow waters (less than 4.6 meters), the sample was collected between the surface and the lake bottom. Four ml of Acid-Lugol's preservative solution was added to the samples at the time of collection and then shipped to NERC-LV for examination. There, the samples were concentrated by settling and examined by EPA personnel. A species list was compiled and phytoplankton were enumerated using a Neubauer Counting Chamber at 400x magnification. All forms were counted within each

Table 3: Analytical methods and precision of laboratory analysis (modified from USEPA, 1975)

Parameter	Method	Precision
Total phosphorus	Persulfate oxidation followed by single reagent methods involving colorimetric determination of antimony-phosphomolybdate complex (for dissolved orthophosphate).	+0.005 mg/l P or + 5%
Nitrite-N	Diazotization of sulfanilamide by nitrite coupled with N-(1-naphthyl)-ethylene diamine.	+0.001 mg/l N or + 2%
Nitrite-Nitrate-N	Cadmium reduction followed by above NO_2 method.	+0.01mg/l N or +5%
Nitrate-N	Determined by difference of two preceeding reactions	+0.01mg/l N or +5%
Ammonia-N	Alkaline phenol hypochlrite reaction producing indophenol blue.	+0.005mg/l N or +5%
Total alkalinity	Methyl orange colorimetric.	+0.5mg/l or 5% as $CaCO_3$

field until a minimum of 100 fields had been viewed, or until the dominant form had been observed a minimum of 100 times. Additional explanation of EPA-NES materials and methods can be found in EPA-NES Working Paper No. 175 (USEPA, 1975).

Definition of the Independent and Dependent Variables

One of the stated purposes of the EPA-NES was to document the conditions found in the survey lakes. The research contained in this paper is based on the assumption that the parameter values obtained by the EPA-NES are truly representative of the summer conditions of the 90 lakes selected under the criteria previously described.

The following are parameters compiled for this research from the individual lake "Working Papers" of the EPA-NES:

1) Dominant summer algal genera, which was determined by EPA-NES personnel based on relative cell size and cell concentration;

2) Median summer total phosphorus concentration (mg/l);[1]

3) Median summer inorganic nitrogen concentration (mg/l);

---

[1]For this parameter, and likewise for other applicable parameters, the median value was chosen to represent and characterize the lake of interest. This was done to eliminate the disproportional influence of extreme values on the "central tendency" statistic. That is, when the data are skewed, the mean can be greatly influenced by outlier values, while the median is essentially insensitive to extreme values (Reckhow, 1979).

For example, it was often observed that the very bottom waters contained a phosphorus concentration measurement an order of magnitude or more greater than the values measured in the epilimnion. Although this fact is interesting from a limnological point of view, it would have biased upward the general, "average" value desired to describe lake conditions.

4) Median summer pH;

5) Median summer water temperature (centigrade);

6) Median summer dissolved oxygen concentration (mg/l);

7) Median summer alkalinity (mg $CaCO_3$/l);

8) Total annual areal phosphorus loading (g/m^2-yr);

9) Total annual areal nitrogen loading (g/m^2-yr);

10) Total annual water inflow (m^3/yr), which was provided to the EPA-NES by the U.S. Geological Survey;

11) Hydraulic detention time (yr);

12) Mean lake depth (m);

13) Lake volume (m^3);

14) Lake area (m^2);

15) Watershed area (m^2);

Additional variables were also calculated based on the above listed variables. These included:

16) Areal water loading (q_s in m/yr)

where:

$$q_s = \frac{\text{annual total water inflow}}{\text{lake surface area}}$$

17) Average influent phosphorus concentration (INP in mg/l)

where:

$$INP = \frac{\text{total annual areal phosphorus load x hydraulic detention time}}{\text{mean lake depth}}$$

18) Average influent nitrogen concentration (INN in mg/l)

where:

$$INN = \frac{\text{total annual areal nitrogen load x hydraulic detention time}}{\text{mean lake depth}}$$

19) Free carbon dioxide concentration, which was calculaked from pH and alkalinity data according to Park (1969) (μmoles/l)

The parameters 2-19 were considered to be the independent variables for the purposes of the empirical study. Parameter 1, the dominant algal genera, were grouped into blue-green genera and nonblue-green genera and these groups were considered to be the dependent variables.[1]

---

[1]It is evident that the identification and description of summer algal dominance (blue-green versus nonblue-green) taken from a small aliquot from a specific sample parcel from a larger water body is subject to possible sampling error. On any given day, complications in algal sampling can arise due to non-random distribution of populations (e.g., into clumps or patches, both vertical and horizontal). For the purposes of this study, however, the sampling techniques utilized by the EPA-NES has been assumed to be adequate enough to truly characterize the resident algal population.

CHAPTER IV

AN INTRODUCTION TO EXPLORATORY DATA ANALYSIS

Introduction

This research is focused on blue-green algal dominance as an eutrophic lake condition that may be desirable to manage. In order to assess the feasability of this management objective, however, it is important to identify which factors are most involved in determining dominance. The literature survey presented in Chapter II provides some insight into this phenomenon, but most of the studies are either theoretical or experimental in nature. An objective of this research is to introduce empirical data analysis as a new approach in understanding the problem of undesirable algal dominance in lakes.

This chapter describes the empirical data analysis techniques that were used in the next chapter to examine the relationships among: 1) the chemical and physical lake variables (i.e., the independent variables); and 2) the dominant algal type (i.e., blue-green algae versus nonblue-green algae as the dependent variables).

Analysis of Data Variability Using Box Plots

An important first step in the analysis of an empirical data set is an assessment of data distributions. The "mean" and "standard deviation" are common statistics used to describe central tendency and data spread respectively. These statistics are useful in this regard, however, they can be misleading because they can be influenced by extreme values (Reckhow, 1979). Therefore, for data sets with skewed distributions (as is the case for most water quality related data sets: Montgomery, personal communication), it is generally more informative to use robust statistics such as the median (central tendency) and the interquartile range (data spread).[1]

One graphical technique for displaying batches of data is the box plot (McGill et al., 1978). This technique is based on order statistics[2] and the plot itself is constructed from five values from the (ordered) data set. These values are: 1) the median; 2) the minimum value; 3) the maximum value; 4) the 25% value; and 5) the 75% value. Figure 2 presents the basic configuration of the box plot.

One question that can be answered via empirical data analysis is; "For which of the independent variables do the dependent variable groupings differ the most?" In this regard, box plots can be very useful since an independent variable can be subdivided by dependent variable groupings, and then the groups may be plotted side-by-side. Visual comparisons of such box plots may be enhanced by the incorporation of the statistical significance of the median into the plot. This is achieved by notching the box at a desired confidence level. For example, if the 95% confidence level notches around two medians do not

---

[1]The interquartile range is the difference between the value at the 75% level and the value at the 25% level.

[2]Order statistics are based on the ordering of the data points from low value to high value.

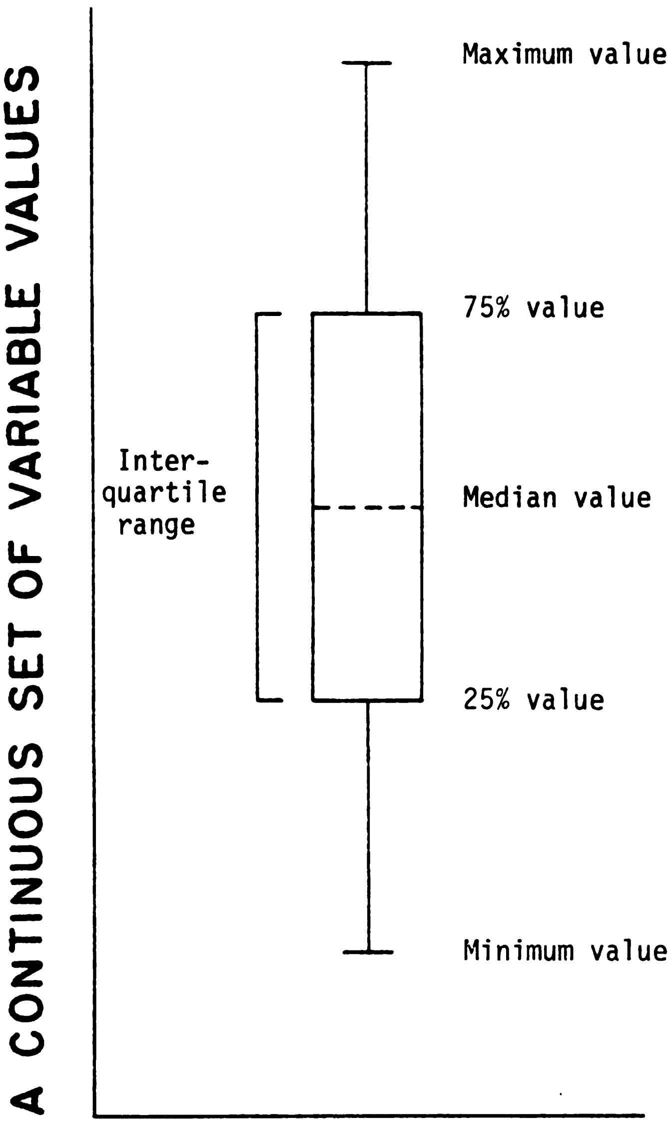

Figure 2: The basic configuration of a box plot

overlap in the display, the medians are roughly significantly different at the 95% confidence level. The height of the notch above and below the median is ± Cs, where C is a constant between 1.96 and 1.39 and s is the standard deviation of the median (see McGill et al., 1978 and Reckhow, 1979 for details). Figure 3 displays, for variable y, a pair of box plots; Group A and Group B. Note that the notches do not overlap and therefore group medians may be considered significantly different (at the predefined confidence level).

Box plots do not represent the only method of estimating the discriminating power of independent variables, however. Snedecor and Cochran (1967) suggest that the following function be used to measure the classification effectiveness of an independent variable, given two predefined dependent variable groups.

$$\frac{u_2 - u_1}{2\sigma} \tag{4.1}$$

where:

u_1 = variable mean value for group 1

u_2 = variable mean value for group 2

σ = variable standard deviation (both groups combined)

"The discriminating power of a variable increases as relationship (4.1) increases (greater distance between the means), and a value for relation (4.1) of 1.5 indicates a high degree of classification success (94% if the variable is normally distributed)" (Reckhow, 1977).

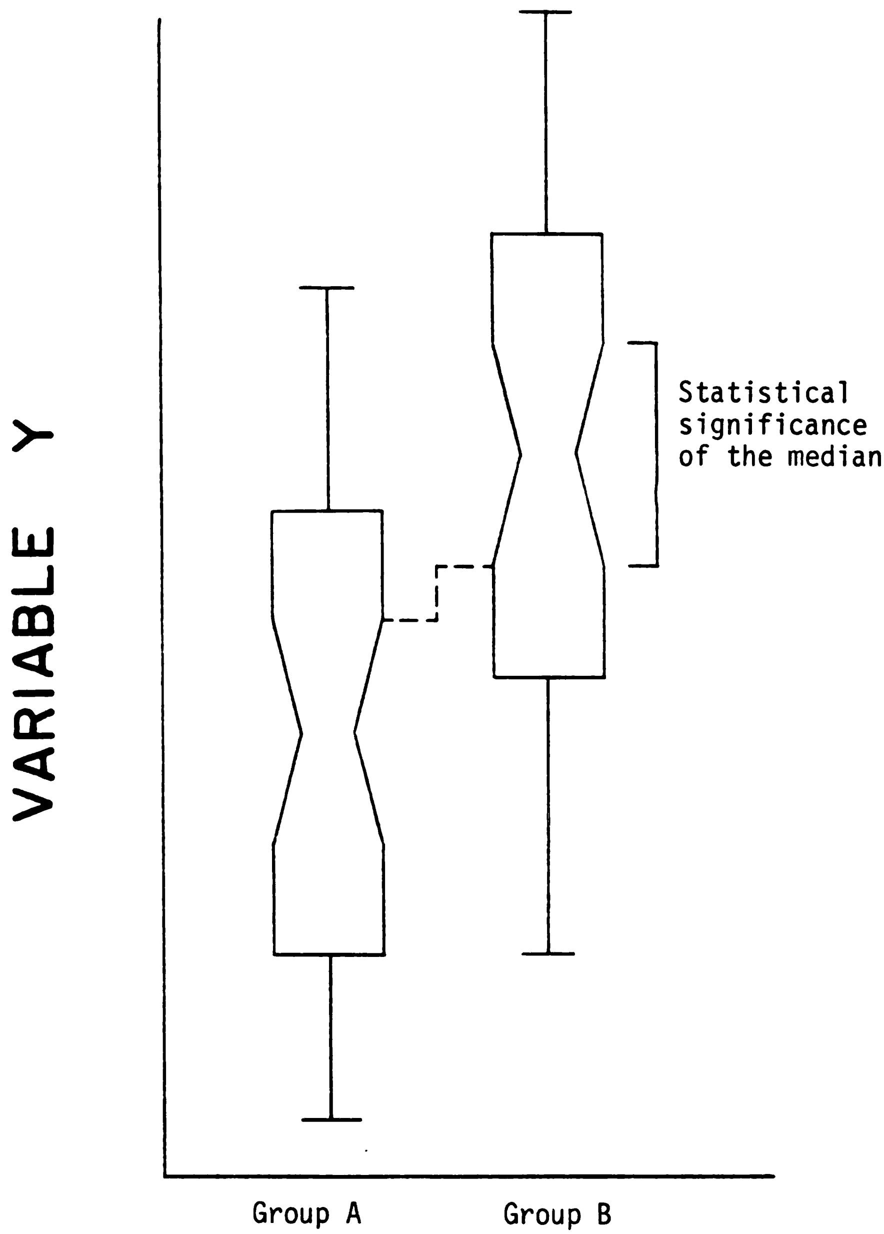

Figure 3: Box plots possessing significantly different medians

Analysis of Bivariate Relationships

Another phase of exploratory data analysis is the examination of relationships between variables of interest. Careful study of variable relationships facilitates the construction of a conceptual model (such as Figure 1) and may suggest improvement of variable expression.

The association between two variables is often expressed in terms of a correlation coefficient. A correlation coefficient is a summary of the linear relationship and indicates the degree to which variation of one variable relates to the variation of the other. The correlation coefficient (r) may be calculated by the following equation:

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (4.2)$$

where:

$\sum_{i=1}^{n}$ = the sum of the elements; from i=1 to i=n

$\bar{x},\bar{y}$ = sample mean for variables x and y respectively

x_i, y_i = elements from variables x and y respectively

The value r may range from -1 to +1, the stronger the linear association, the higher the absolute value of r.

Reckhow (1979) notes two important limitations of the correlation coefficient: 1) since r evaluates a linear relationship, it will not reflect an intrinsically nonlinear association; and 2) r may be biased towards a higher absolute data value if the data set is not normally distributed.[1] For these reasons it is often desirable to evaluate a correlation coefficient with the aid of a bivariate plot. Plots may visably provide evidence of nonlinearity and the need for re-expression. One useful modification of bivariate plotting is to identify different groups of interest within the plot. By constructing these bivariate-discriminant plots, variable relationships can be examined within the groups and between the groups or well as within the total data set.

Discriminant Analysis

Many research situations require an investigation of the functional relationship between independent and dependent variables. Regression analysis is a common method of optimizing a linear fit (y versus x_1, x_2 ... x_n) to the data points. That is, multiple regression "best" combines the independent variables in a linear equation to describe or predict the dependent variable. This analysis is not especially appropriate if the dependent variable has discrete states, however. For example, regression analysis can be aptly used to develop a function to predict phosphorus concentration, which can assume a continuous set of values (Reckhow, 1979). On the other hand, if the dependent variable has distinct or discrete states (e.g., blue-green and nonblue-green algal dominance), the multivariate statistical procedure of discriminant analysis is better suited to describe the functional relationship between independent and dependent variables.

---

[1] In some instances non-normal distributions may be normalized via a data transformation. For example, a commonly observed distribution among water quality data is the log-normal. That is, a logarithmic transformation of the data results in a normal distribution.

The concept of discriminant analysis is fairly old (introduced by R. A. Fisher in 1936) but infrequently used in comparison to multiple regression analysis. Since the procedure may be unfamiliar to the reader, discriminant analysis and its usefulness in exploratory data analysis and as a modeling technique is described in more detail below.

The principle objective of discriminant analysis is to discriminate, classify or otherwise distinguish between two or more groups of cases using a set of independent variables (Morrison, 1969). This multivariate statistical procedure estimates a linear combination of the independent variables that "best" classify the cases (e.g., the EPA-NES lakes) into one of the predefined dependent variables classes (e.g., algal-type dominance). In other words, the discriminant function attempts to maximize statistical distinction along a single dimension.

The discriminant function is of the form:

$$D_i = d_{i1}z_1 + d_{i2}z_2 + \dots + d_{ip}z_p + d_o \qquad (4.3)$$

where:

D_i = the discriminant score

d_{ij} = the discriminant coefficients

z_j = the raw score for the discriminating variable (e.g., lake parameter values such as nitrogen concentration, mean depth, etc.)

d_o = a constant

Thus, for a two group example:

if $D_i < D_{i(crit)}$, a case is classified as a member of group 1 (e.g., blue-green dominated lake)

if $D_i > D_{i(crit)}$, a case is classified as a member of group 2 (e.g., nonblue-green dominated lake) where:

$D_{i(crit)}$ = the critical value for the discriminant score that separates the groups

$D_{i(crit)}$ is defined by the classification boundary, if there are two independent variables the boundary is a straight line.[1] The boundary defined by three independent variables is a two-dimensional plane in a three-dimensional space, etc. In other words, the classification boundary is a n-1 dimensional hyperplane in n space (Morrison, 1969).

The discriminant analysis procedure assumes that the independent variables are normally distributed and that the independent variable variances are the same for all groups. In succinct terms, the discriminant procedure assumes that the group covariance matrices are equal (Nie et al., 1974).

Two research objectives may be met using discriminant analysis:

1) the development of a predictive model that can be used to estimate group memberships of unknown cases; and

2) the identification of the relative importance of the independent variables in the discriminant function.

Obviously, the value of assessing the relative importance of discriminating variables hinges on how well they actually discriminate. Likewise, a predictive model is only as valuable as the discriminating power defined by the independent variables. Therefore, both objectives are necessarily developed concurrently in the discriminant analysis procedure, regardless of the initial priorities.

---

[1]The line is straight only if the assumption of equal covariance matrices between groups is not violated (Reckhow, personal communication).

The discriminatory power contained in the discriminant function can be assessed by classifying known cases and then determining the number correctly classified. This classification step is necessary to demonstrate that classification results are better than one would expect by chance. If the percentage of correct predictions are judged to be significant, one may then begin to draw meaningful information from the investigations into variable relationships and the relative importance of the independent variables as well as the function's potential success as a predictive model (Frank et al., 1965). On the other hand, if the predictive capabilities of the discriminant function are no greater than can be expected by chance, investigations into relative variable importance and the function's use as a model is of no relevant consequence.

Assuming that a discriminant function's predictive capabilities are judged to be significant, one can logically proceed to the second objective; identification of the relative importance of the independent variables in the function. This can be accomplished by examining the function's standardized discriminant coefficients (Nie et al., 1974). As in regression or factor analysis, the weight of the coefficient represents the relative contribution of the independent variable in the function. The sign of the coefficient determines the direction of the independent variable's effect on the dependent variable. However, if the variables are correlated, the standardized discriminant coefficients are much less meaningful (Reckhow, personel communication).

When this situation occurs, the relative importance of the independent variables should be judged primarily on the examination of box plots and the Snedecor-Cochran statistic (Equation 3.1).

In many research situations, more independent variables are available to the researcher than practically or statistically needed to achieve satisfactory discrimination results. Therefore, step-wise discriminant analysis techniques that enter variables into the function one at a time are useful in determining the "best" discriminating variables. In step-wise discriminant analysis, the independent variable that enters the function first is the best discriminating variable judged by group (dependent variable) mean value separation. Subsequent variables are included in the function based on their discriminatory power in combination with the previously selected variable(s). One common statistic used for step-wise variable selection is the partial F ratio. This "F statistic" is a test for statistical significance of additional dependent variable separation (discrimination) created by a new variable beyond that already achieved by previously entered variables. Thus, the variable with the highest partial F ratio (conditioned on the variables already present in the function) is selected for inclusion at each step.[1] The selection procedure may

---

[1]It should be noted that the methods of variable inclusion in the discriminant function do not "order" the relative discriminant importance of the variables. It can only approximate the "best" discrimination variables since the procedure does not assess every possible subset of variables but rather sequentially selects the "next best" discriminator conditioned on the variables previously selected (Nie et al., 1974).

be halted when the partial F ratio is deemed too small to be of significance. Wilk's Lambda is also a common measure of group discrimination and is a direct inverse function of the "F statistic". That is, the variables which maximize the F ratio, minimize Wilk's Lambda (Nie et al., 1974).

Morrison (1974) notes that the "F statistic" is the multidimensional analog of the traditional "t-test" for statistical difference between group means. He points out that the concept of statistical significance must be approached with caution since levels vary with sample size. Therefore, statistics like the partial F ratio and Wilk's Lambda are poor indicators in themselves of the ability to which independent variables can discriminate. Thus, the question; "How well do the independent variables discriminate?" must include tests of correct classification.

One issue of concern is the application of discriminant analysis to a situation where the majority of a population (or sample) belongs to one dependent variable group. Thus, population (or sample) distribution alone favors the classification of a case into that one group over any other groups. In this instance, one may either eliminate cases from the majority group(s) until the groups are approximately even or incorporate prior probabilities into the discriminant function. These probabilities, however, affect only the constant value and have no effect on the discriminant coefficients (Frank et al., 1965). Thus, if the research purpose is only to assess variable importance in the discriminant function (via the examination of standardized discrimination coefficients), prior probabilities may be disregarded and all available cases used to calculate the discriminant function.

The discriminant information contained in the discriminant function can be re-expressed as classification functions, one for each group. These functions are derived from the pooled within-group covariance matrix and the centroids for the discriminating variables. The classification functions are of the form:

$$C_i = c_{i1}v_1 + c_{i2}v_2 + \dots c_{ip}v_p + c_o \tag{4.4}$$

where:

C_i = the classification score for group i

c_{ij} = the classification coefficients

v_j = the raw score for the discriminating variable

c_o = a constant

An individual case is classified into the group that yields the highest classification score of all the group classification functions (Nie et al., 1974).

In the usual procedure of classification function development, all individual cases are assumed to have equal probability of group membership. However, if sample or actual population distributions are known, or if the risks associated with misclassification represent a special concern, an adjustment of classification probabilities should be performed (Morrison, 1974). Such is the case, as previously mentioned, when group memberships are of grossly different size.

An important consideration when using the discriminant or classification functions in a predictive capacity, is the uncertainty inherent in the prediction. Reckhow (1978) presents probability equations that take into account variable uncertainties. He also provides a method which expresses the classification equations in a probabilistic form.

This method is presented below:

$$P_i = \frac{1}{e^{-(\Sigma c.f._i)} + 1} \qquad (4.5)$$

where:

P_i = the probability associated with group i classification

$\Sigma c.f._i$ = the sum of group i classification functions

The Statistical Package for the Social Science (SPSS) (Nie et al., 1974) also present a method by which classification probabilities can be estimated:

$$P(G_j/X) = \frac{P_j|D_j|^{-1/2}e^{-\sigma_j^2/2}}{\sum_{i=1}^{g} P_j|D_j|^{-1/2}e^{-\sigma_j^2/2}} \qquad (4.6)$$

where:

$P(G_j/X)$ = the probability associated with group j classification

P_j = the prior probability for group j

D_j = the group covariance matrix for group j

g = the number of groups

σ_j^2 = the chi-square distance from each group centroid

Although uncertainty is inherent in any classification prediction, when constructing the discriminant or classification functions, the independent variable uncertainty (in the model-development data set) is automatically incorporated in the model (Reckhow, 1977; Walker, 1977). Therefore, if the data for a specific application case are gathered and assessed in a manner similar to that for the development data set, the probabilities (with that level of uncertainty) require no further error analysis.

Discriminant analysis computer programs usually present classification information in a table or matrix (Morrison, 1969). This table described how many cases from the original data are assigned, via the classification function, to the dependent variable groups. Thus, the percentage correctly classified is the percentage of cases assigned to their true original groups. This method, however, produces an upward bias of classification "success" since the data set is optimally fitted to the functions when it is constructed (Morrison, 1969). That is, the discriminant function automatically maximizes the percentage of cases correctly classified.

One method to reduce classification bias is to use a percentage of the data cases to construct the discriminating function(s) and then use the remaining percentage to test for significant classification results. Note that a large data set is generally required to apply this method. The "jackknife" technique is similar to this in theory, except that, in the jackknife approach, an individual case is omitted and a discriminant function constructed using the remaining cases

(Mosteller and Tukey, 1977). The omitted case is then classified based on the computed discriminant function. The process is subsequently repeated until every case in the data set has been omitted and classified according to its own "unbiased" discriminant function. The overall result of the jackknife method is the construction of an "unbiased" correct classification table that can be used to assess the discriminatory power of the function's discrimination variables.

Reckhow (1978) states that although the discriminant function(s) and the classification functions convey nearly the same discriminant-type information, they may be best used for different purposes. He suggests that the discriminant function, when displayed in a graphical form, is more useful for multi-lake analysis or cross-sectional comparisons while the classification functions, when expressed in a probabilistic form, are more useful for single-lake, longitudinal studies.

Evaluation techniques and exploratory data analysis must all be considered in conjunction with the goals and objectives of the research when assessing the usefulness of discriminant function(s) or classification functions as a predictive model. Clearly, a predictive model that requires independent variable values that cannot be obtained or estimated within normal budgetary constraints is useless as a management tool. Reckhow (1979) also points out that:

> "It is critical that an appropriate model be selected (assuming that an appropriate model can be found) for the intended purpose, and that the individual applying the model has a clear understanding of the model's limitations."

He further states that:

> "It should be the modeler's responsibility to clearly document the proper use and limitation on the use of his/her model. For an empirical model, this documentation should include a statement of all limitations ... associated with the data set used to develop the model. In addition, any biases noted within the range of application should also be indicated. Another important but generally neglected statement of information about the model concerns the type of issues and decisions appropriate for model application and the value of the information provided by the model toward this application."

CHAPTER V

AN EMPIRICAL ANALYSIS OF EPA-NES DATA

Introduction

In the analysis of a large multivariate data set it is important that the variables themselves (e.g., data distributions) and the relationships among the independent variables and between the dependent and independent variables (e.g., correlations) are well understood. This chapter documents the exploratory data analysis that was used in this research to examine relationships between the chemistry and physics of select EPA-NES lakes and the dominant algal-type in those lakes. Discriminant analysis was also used in the exploration of multivariate relationships and in the construction of a predictive model.

The Statistical Package for the Social Sciences (SPSS), an integrated system of computer programs, was used for the calculation of the statistical material described in this chapter. SPSS was accessed via the Control Data Corporation Computer (CDC 6500) on the Michigan State University Campus.

Preliminary Statistics

Tables 5 and 6 present summary statistics that can be used to analyze data distributions and variable relationships. Table 5 presents the mean, median, standard deviation, minima, and maxima of the independent variables derived from the ninety EPA-NES lakes preselected for use in this research. Table 4 is the key to the data tables and Table 6 presents a matrix of independent variable correlation coefficients.

Not surprisingly, there are relationships among some of the physical lake variables; notably, mean depth, lake volume and hydraulic detention time. There is also high correlation between average influent phosphorus concentration and lake total phosphorus concentration. Other empirical studies have also verified this close relationship (Vollenweider, 1968; Reckhow, 1977). It is interesting to note, however, that average influent nitrogen concentration and lake inorganic nitrogen concentration do not correlate nearly as well. Table 6 also indicates relatively strong relationships between pH and: 1) carbon dioxide concentration; and 2) alkalinity. This is to be expected because of the intimate relationship among the CO_2, bicarbonate and carbonate equilibrium concentrations and pH (Wetzel, 1975).

Analysis of EPA-NES Data Using Box Plots and the Snedecor-Cochran Statistic

As discussed in Chapter IV, summary statistics, like those presented in Table 5, can be misleading if the data set has a skewed distribution. Therefore, graphical presentations of data, such as box plots, are desirable since they often convey more information than simple summary statistics alone.

The box plot is useful in two aspects of data analysis: 1) it can lead to a thorough examination of data distributions; and 2) it

Table 4: Key to EPA-NES data tables

A_L = Lake area (km^2)

Z = Mean depth (m)

V = Lake volume ($10^6 m^2$)

T = Hydraulic detention time (yr)

A_B = Basin (watershed) area (km^2)

Q = Mean annual total water inflow (cms)

q_s = Areal water loading (m/yr)

Prec = Mean annual total precipitation (cm/yr)

Temp = Median summer water temperature (°C)

DO = Median summer dissolved oxygen concentration (mg/l)

pH = Median summer pH (unitless)

AlK = Median summer alkalinity (mg $CaCO_3$/l)

P = Median summer total phosphorus concentration (mg/l)

N = Median summer inorganic nitrogen concentration (mg/l)

CO_2 = Median summer free carbon dioxide concentration (μmoles/l)

L_P = Total annual phosphorus load (g/m^2-yr)

L_N = Total annual nitrogen load (g/m^2-yr)

INP = Average influent phosphorus concentration (mg/l)

INN = Average influent nitrogen concentration (mg/l)

Table 5: Statistics for the EPA-NES data set

	Mean	Median	Standard Deviation	Minimum	Maximum
A_L (km^2)	11.02	3.94	16.80	0.10	81.12
Z (m)	5.44	4.15	5.01	0.90	31.60
V($10^6 m^2$)	83.37	15.45	203.17	0.20	1170.60
T (yr)	1.569	0.302	3.639	0.003	21.000
A_B (km^2)	2231.32	176.50	10521.71	4.00	96324.00
Q (cms)	13.59	1.40	40.84	0	287.10
q_s (m/yr)	1.8	0.5	3.5	0.0	20.0
Prec. (m)	91.1	90.9	16.8	58.0	156.0
Temp (^oC)	21.9	21.8	4.2	10.4	30.1
DO (mg/l)	6.0	6.8	2.5	0.0	10.4
pH (unitless)	7.89	7.96	6.73	5.2	9.2
AlK (mg $CaCO_3$/l)	120	120	76	10	334
P (mg/l)	0.139	0.051	0.266	0.005	1.420
N (mg/l)	0.75	0.29	1.04	0.05	4.62
CO_2 (μmoles/l)	103.7	53.8	137.4	1.2	600.0
L_p (g/m^2-yr)	8.28	1.60	19.14	0.03	129.43
L_N (g/m^2-yr)	140.3	35.4	305.1	1.5	2382.8
INP (mg/l)	0.32	0.11	0.79	0.01	5.70
INN (mg/l)	3.68	2.45	3.83	0.51	25.25

Table 6: Matrix of correlation coefficients for the EPA-NES variables

	A_L	Z	V	T	A_B	Q	q_s	Prec	Temp	DO	pH	Alk	P	N	CO_2	L_P	L_N	INP	INN
A_L	1	.26	.91	.34	.53	.52	-.31	-.18	.20	.27	.11	-.04	-.29	-.17	-.18	-.37	-.36	-.14	-.09
Z		1	.63	.56	-.02	-.04	-.28	-.13	-.39	-.25	.04	.07	-.55	-.21	.12	-.41	-.31	-.30	-.16
V			1	.51	.42	.40	-.37	-.21	-.01	.12	.10	-.01	-.47	-.22	-.09	-.47	-.42	-.24	-.14
T				1	-.53	-.56	-.92	-.25	-.12	-.00	.22	.09	-.28	-.32	-.23	-.80	-.87	.12	.26
A_B					1	.93	.56	-.09	.11	.11	.01	.06	-.05	.19	.08	.44	.54	-.22	-.26
Q						1	.65	.06	.16	.11	-.10	-.09	-.09	.11	.12	.41	.52	-.29	-.35
q_s							1	.22	-.00	-.12	-.21	-.06	.16	.26	.29	.78	.88	-.20	-.31
Prec								1	.40	-.04	-.40	-.41	-.11	.14	.19	.06	.14	-.25	-.25
Temp									1	.38	.04	-.15	.26	.18	-.23	.14	.04	.25	.16
DO										1	.13	-.23	.03	-.15	-.41	-.06	-.14	.07	-.05
pH											1	.69	.30	.14	-.78	.10	-.01	.48	.50
Alk												1	.25	.29	-.13	.22	.20	.43	.57
P													1	.39	-.24	.61	.34	.78	.51
N														1	.05	.49	.55	.33	.50
CO_2															1	.08	.21	-.34	-.27
L_P																1	.91	.41	.19
L_N																	1	.10	.08
INP																		1	.81
INN																			1

can be conveniently used to compare variable groupings. Recall (from Chapter IV) that the graphical technique of the box plot is based on order statistics and can include the following information (Reckhow, 1979):

1) the median, which is a robust measure of central tendency;

2) the interquartile range, which is a robust measure of variability (the difference between the value at the 75% level and the value at the 25% level);

3) the range, which is the maximum value minus the minimum value;

4) an indication of skew, which is obtained by comparing the symmetry above and below the median;

5) a measure of significant difference between medians, which is obtained by notching the boxes to some confidence interval.[1]

Figures 4 through 9 present box plots grouped by algal dominance (blue-green dominated lakes versus nonblue-green dominated lakes) for the independent variables deemed potentially influential by this author, based on an initial examination of group statistics and the literature review in Chapter II.

Figures 4 through 9 reveal four independent variables that display significant difference between algal group medians (at the 95% confidence level). Empirical conclusions based on these box plots are listed below:

1) Lakes dominated by blue-green algae are generally lower in CO_2 concentration than are lakes dominated by nonblue-green algal-types.

---

[1]The height of the notch above and below the median is ± Cs, where C is a constant between 1.96 and 1.39 and s is the standard deviation of the median. A value of 1.7 was chosen to represent C in this paper (see McGill et al., 1978 for details).

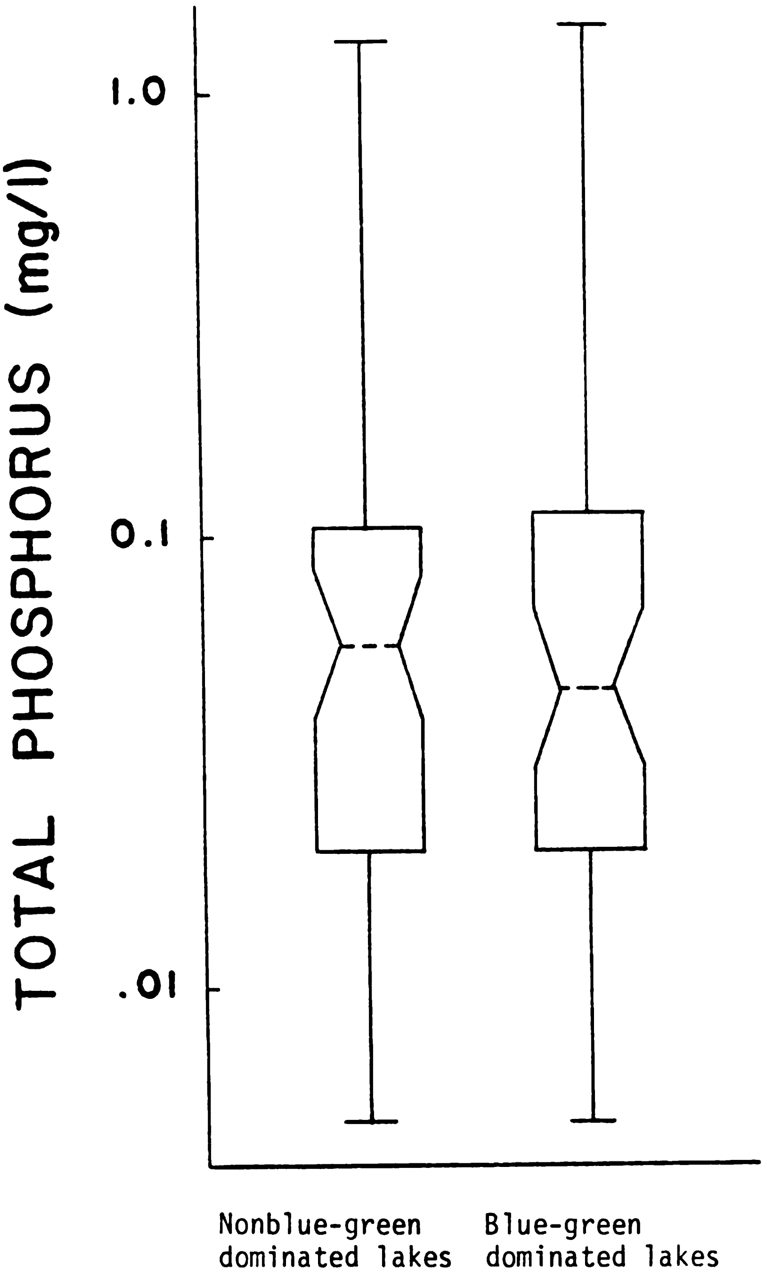

Figure 4.--Box plots of total phosphorus by algal-type dominance

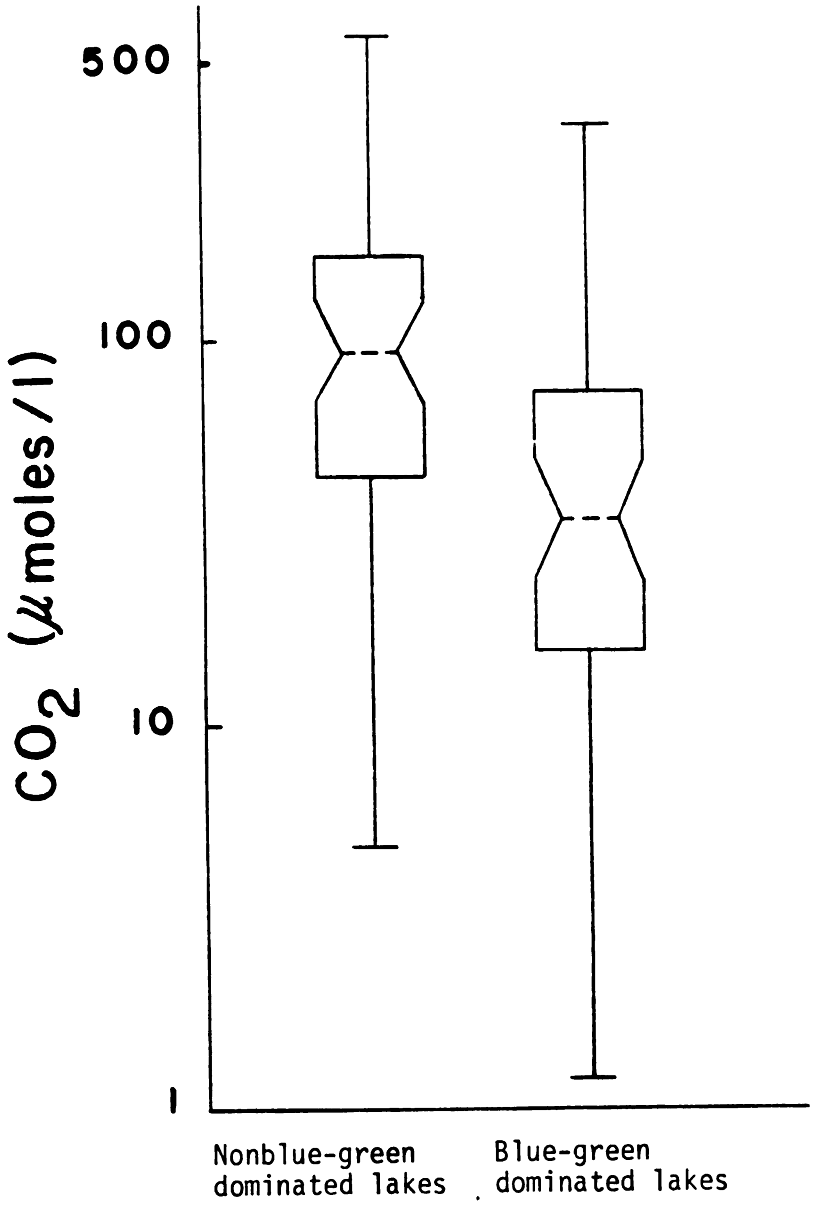

Figure 5.--Box plots of carbon dioxide by algal-type dominance

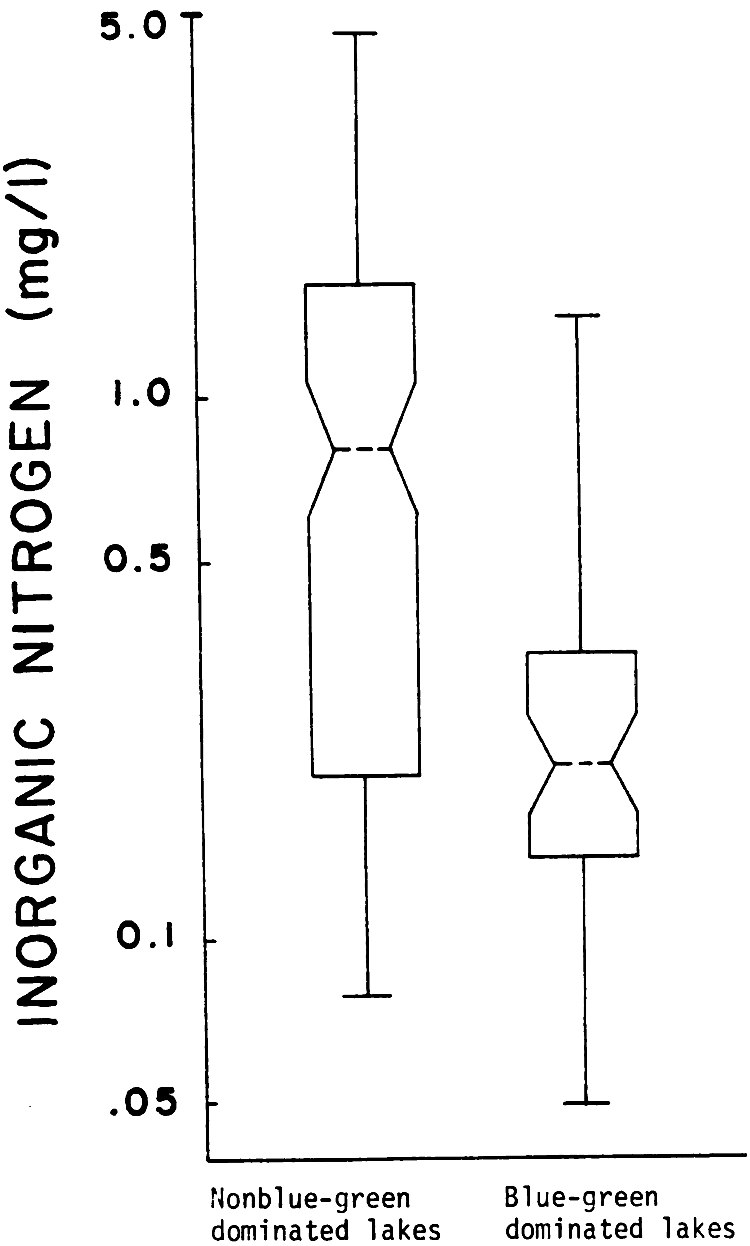

Figure 6.--Box plots of inorganic nitrogen by algal-type dominance

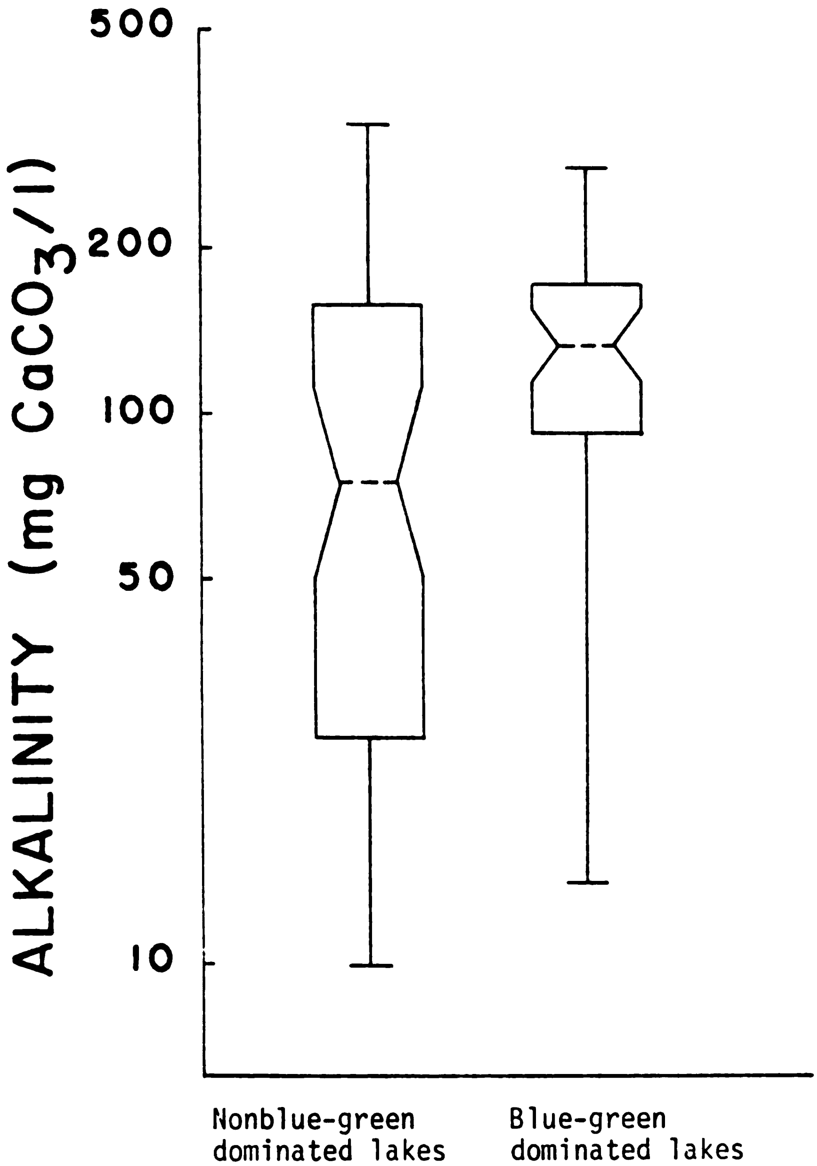

Figure 7.--Box plots of alkalinity by algal-type dominance

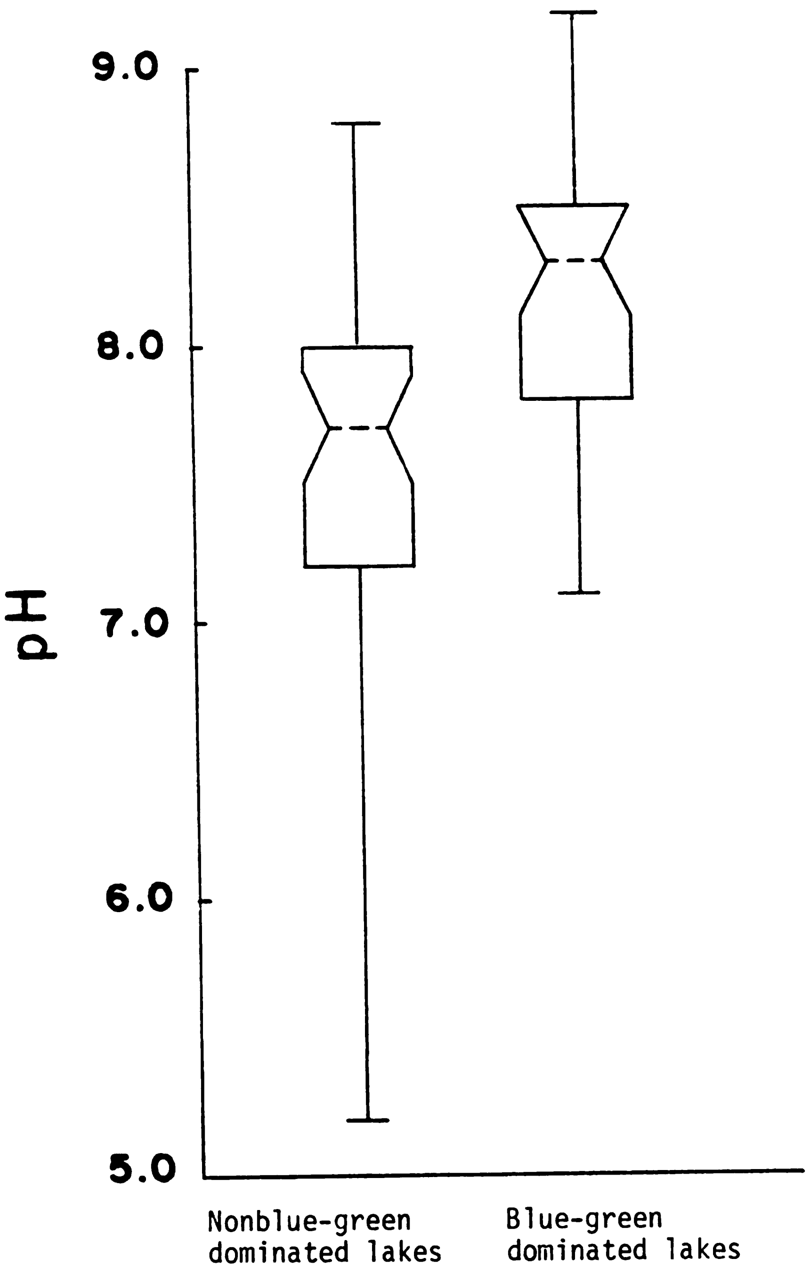

Figure 8.--Box plots of pH by algal-type dominance

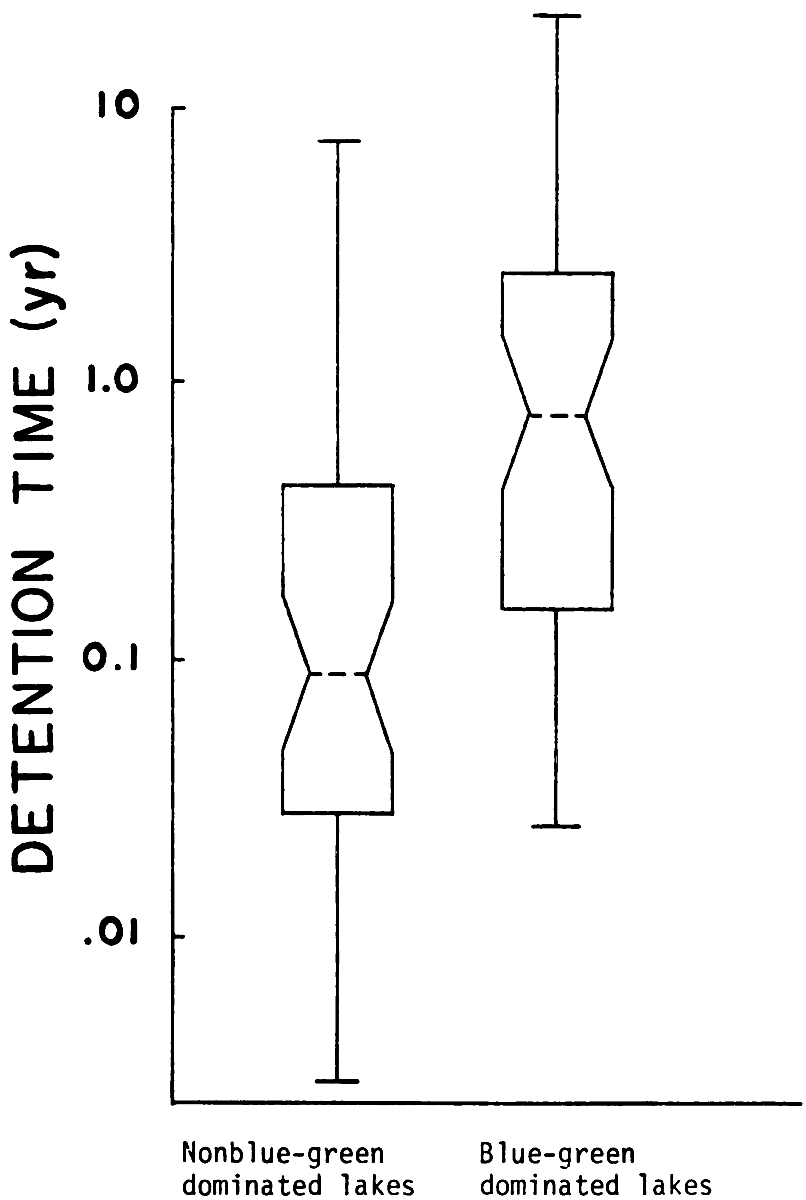

Figure 9.--Box plots of detention time by algal-type dominance

2) Lakes dominated by blue-green algae are generally lower in inorganic nitrogen concentration than are lakes dominated by nonblue-green algal-types.

3) Lakes dominated by blue-green algae are generally higher in pH than are lakes dominated by nonblue-green algal-types.

4) Lakes dominated by blue-green algae generally have a longer hydraulic detention time than lakes dominated by nonblue-green algal-types.

5) The other independent variables in the data set do not display a significant difference of group medians (at the 95% confidence level). However, visual inspection reveals that blue-green algae also tend to dominate in lakes with a higher alkalinity than nonblue-green dominated lakes.

It is interesting to note that lake phosphorus concentration, which is frequently used as an explicit measure of eutrophic conditions, does not discriminate the algal groups well (see Figure 4). Recall, however, that the EPA-NES selection criteria favored the inclusion of lakes impacted by cultural sources, notably sewage treatment plants (USEPA, 1975). These sources undoubtedly increase the amount of phosphorus available for aquatic plant production. It is generally agreed that this phenomenon will: 1) increase algal growth (Vollenweider, 1968); and 2) increase the biogeochemical cycles of many other essential growth nutrients (King, 1979). These events can ultimately serve to reduce the importance of phosphorus in determining the total amount of algal growth. Thus, it may be proposed that the natural phosphorus limit has been nullified in many of the EPA-NES lakes (due to excessive phosphorus loading) and that other factors such as nitrogen and carbon may be limiting algal growth. This shift of limiting factors also plays a role in determining algal dominance

(King, 1970, 1972). Therefore, perhaps the qualitative importance of phosphorus has also been nullified in many of the EPA-NES lakes, thus explaining the lack of discriminating power. On the other hand, it may also be that phosphorus is simply not a good discriminating variable, regardless of limiting conditions.

When the algal demand for carbon dioxide exceeds the rate of supply from the atmosphere and biological sources, the plants will extract it from the bicarbonate-carbonate alkalinity of the water (King, 1970, 1972). This extraction results in: 1) a decrease in the equilibrium CO_2 concentration; and 2) an increase in pH. Pritchard et al. (1962) report that low CO_2 concentrations will induce the extracellular excretion of organic material by algae. Thus, it is proposed that continued reduction of carbon dioxide (via extraction from the alkalinity system) will eventually cause the green algae and diatoms to flocculate and rapidly sink out of the photic zone (King, 1972). Therefore, the buoyant blue-green algae will be more likely to dominate under these biologically lowered CO_2 concentrations. Figure 5 empirically shows that lakes dominated by blue-green algae are generally lower in CO_2 concentration than are lakes dominated by nonblue-green algal types.

As observed in Figure 6, lakes dominated by blue-green algae are generally lower in inorganic nitrogen concentration than are lakes dominated by nonblue-green algal types. Fogg et al. (1973) point out that some blue-greens have the ability to fix elemental nitrogen, thus giving them a competitive advantage in "low nitrogen" or nitrogen limited lakes. An investigation of the blue-green dominated EPA-NES lakes contained in the data set, however, revealed that less than half were dominated by known nitrogen-fixing genera. Therefore, based on empirical evidence derived from this data set, it appears that both nitrogen-fixing and non-nitrogen-fixing blue-greens tend to dominate in lakes with a low nitrogen concentration.

A possible explanation of this "overall" blue-green affinity for EPA-NES lakes with low nitrogen concentration may involve the stress placed on nonblue-green algal cells under nitrogen limited conditions. That is, an inadequate supply of nitrogen (relative to algal demand) may, as in the case of carbon, cause leakage of organic material from the cells, followed by the flocculation and sinking of non-bouyant green and diatom populations. This phenomenon would then favor the dominance of the bouyant blue-greens (both nitrogen-fixers and non-nitrogen-fixers).

Low or reduced inorganic nitrogen concentrations in lakes may be due to many factors and therefore it is beneficial to examine portions of the nitrogen cycle closer. Wetzel (1975) states that nitrogen may be added to lakes via: 1) overland runoff; 2) precipitation; 3) groundwater drainage; and 4) nitrogen-fixing organisms in the water and sediments. Loss of nitrogen from lakes can occur via: 1) the outflow; 2) settling, to the lake bottom, of nitrogen-containing organic and inorganic particles; 3) conversion of the ammonium ion to ammonia gas at high pH levels and its subsequent loss to the atmosphere; and 4) reduction of nitrate and nitrite to nitrogen gas and its subsequent loss to the atmosphere.

The low inorganic nitrogen concentrations that appear to favor blue-green algal dominance may be, in fact, due to the low input of nitrogen in some EPA-NES lakes. For the majority of the EPA-NES lakes, however, low nitrogen concentrations are more probably a result of the various nitrogen-loss mechanisms functioning during the growing season. Specifically, high rates of algal growth and the incorporation of CO_2 from the alkalinity system will increase the pH. This increase can result in a significant loss of ammonia gas to the atmosphere (King, 1978, 1979). Another potentially significant nitrogen loss from the EPA-NES lakes may occur via the denitrification reaction which takes place rapidly under anaerobic conditions. Decay of excessive algal and plant material can cause anoxic environments and thus promote denitrification and nitrogen gas loss.

Figure 8 indicates that pH is another variable that statistically separates the two algal groups. The box plots show that lakes dominated by blue-greens are generally higher in pH than are lakes dominated by nonblue-green algal-types. Based on the above explanations, however, it can be assumed that the reasons the algal groups are discriminated by pH relates to: 1) plant extraction of CO_2 from the alkalinity (effecting a rise in pH) resulting in carbon limiting conditions; and 2) increasing nitrogen loss via the conversion of the ammonium ion to ammonia gas at the higher pH, resulting in nitrogen limiting conditions.

Figure 9 presents box plots of the detention time variable and shows that the algal group medians are also significantly different. This is not surprising since detention time is known to be an important factor in lake dynamics. King (1978) observed that the pH in lakes impacted by waste water increases as a direct function of detention time. Thus, a eutrophic lake with a long detention time is more likely to face a carbon or nitrogen limit and qualitative changes in dominance than a lake with a short detention time. In addition, lakes with a long detention time are also apt to have less turbulent waters than lakes with a short detention time. Longer detention times would therefore favor the selective advantage of bouyancy, possessed by blue-greens. In Figure 9, the blue-greens, as expected, tend to dominate in lakes with a longer detention time than nonblue-green dominated lakes.

As mentioned in Chapter IV, Snedecor and Cochran (1967) provide a method of estimating the discriminating power of a variable given two pre-defined groups. They propose the use of the following function to "measure" the classification effectiveness of a given variable:

$$\frac{u_2 - u_1}{2\sigma} \tag{5.1}$$

where:

u_1 and u_2 = variable means for group 1 and group 2 respectively

σ = variable standard deviation (both groups combined)

Table 7 presents the Snedecor-Cochran statistic for some of the independent variables defined in this research. Recall (from Chapter IV), that the discriminating power of a variable increases as the Snedecor-Cochran statistic increases. Nitrogen concentration, detention time, pH and CO_2 respectively have the largest numerical values

Table 7: Snedecor-Cochran statistic for select EPA-NES variables

Variable	$\frac{u_2 - u_1}{2\sigma}$
log A_L	.074
log Z	.179
log V	.130
log T	.433
log A_B	.261
log Temp	.093
log DO	.016
log pH	.432
log AlK	.303
log P	.041
log CO_2	.378
log N	.437
log INP	.197
log INN	.167

and therefore may be considered to be the best discriminating variables for the algal groups. This statistic, in fact, confirms the information derived from the examination of box plots.

At this point, the EPA-NES's questionable practice of measuring the pH of a water sample at the end of the sampling day should be discussed (see Chapter III). The pH represents the instantaneous hydrogen ion activity in moles/l. Standard Methods (1971) states that pH can change significantly in a matter of minutes, hence its determination should always be carried out promptly in the field. "The ionization (in a sample), dependent upon the values of K_1 and K_2 for H_2CO_3 as well as on K_w for H_2O at the various temperatures, is to a significant extent related to the alkalinity of the sample. Increasing alkalinity reduces the effect of temperature change on the pH" (Standard Methods, 1971).

The uncertainty surrounding the measurement of pH by the EPA- NES caused this investigator to exclude this variable from the discriminant analysis (model-building) portion of this research. Since the sampling bias should be relatively proportional, however, qualitative conclusions derived from the single variable analysis (e.g., box plots) of pH would still be valid (i.e., lakes dominated by blue-green algae generally have a higher pH than lakes dominated by nonblue-green algae). However, this bias would obviously hamper the validity of any predictive model derived using the uncertain EPA-NES pH values. Further, the CO_2 values are also necessarily excluded from model-building analysis since they were calculated using EPA-NES pH (and alkalinity) data (according to Park, 1969).

The potential importance of the pH and CO_2 concentration to algal dominance is documented in the literature (King, 1978; King and Hill, 1978) and also herein. For the reasons stated above, however, the focus of the following analysis is not on these two variables.

Analysis of the EPA-NES Data Using Bivariate-Discriminant Plots

Excluding CO_2 and pH, single variable analysis points to the inorganic nitrogen concentration and the hydraulic detention time as the two most important variables in determining algal dominance. However, the conditions and factors leading to the dominance of a certain algal type are dynamic and interactive and certainly involve more parameters than detention time and nitrogen concentration alone. The investigation of other EPA-NES variables that enter into the determination of dominance requires a closer examination of variable relationships. As noted in Chapter IV, correlation coefficients are useful in this regard, however, they can also be misleading since they measure the strength of a linear association and since they can also be favorably or unfavorably influenced by outliers. Bivariate plots, on the other hand, visably display correlation information and may suggest the need for variable transformation.[1] As mentioned in Chapter IV, it is often valuable to identify dependent variable groups within the plot. The resultant bivariate-discriminant plot allows for the

---

[1]This was the case herein. The majority of the variables used in this study had data that were log-normal in distribution. Therefore, a logarithmic transformation of all variables was performed prior to all analysis.

examination of variable relationships and trends within the groups, between the groups and within the total data set.

Figures 10 through 15 present bivariate-discriminant plots of select variables. Some observations from these plots are listed below:

1) Based on Figure 10 nonblue-green dominated lakes tend to have higher nitrogen concentrations and lower detention times than blue-green dominated lakes.

2) Phosphorus concentration does not appear to influence dominance.

3) Lakes dominated by blue-green algae generally have a higher influent phosphorus concentration than lakes dominated by nonblue-green algae.

The most interesting bivariate-discriminant relationship is inorganic nitrogen concentration versus alkalinity (Figure 14). This plot shows a distinct separation of the algal groups in the higher alkalinity lakes. Specifically, it can be observed that:

1) Lakes with high alkalinity (greater than approximately 56 mg $CaCo_3$/l) tend to be dominated by blue-green algae when the inorganic nitrogen concentration is less than about 0.7 mg/l.

2) Lakes with a high alkalinity (greater than approximately 56 mg $CaCO_3$/l) tend to be dominated by nonblue-green algae when the inorganic nitrogen concentration is greater than about 0.7 mg/l.

3) Lakes with low alkalinity (less than approximately 56 mg $CaCO_3$/l) do not display the separation described in 1 and 2 above, however, there are only a few lakes on which to base this judgement.

Nitrogen has been suggested to be an important factor in algal dominance, especially in culturally-impacted lakes (see Chapter II).

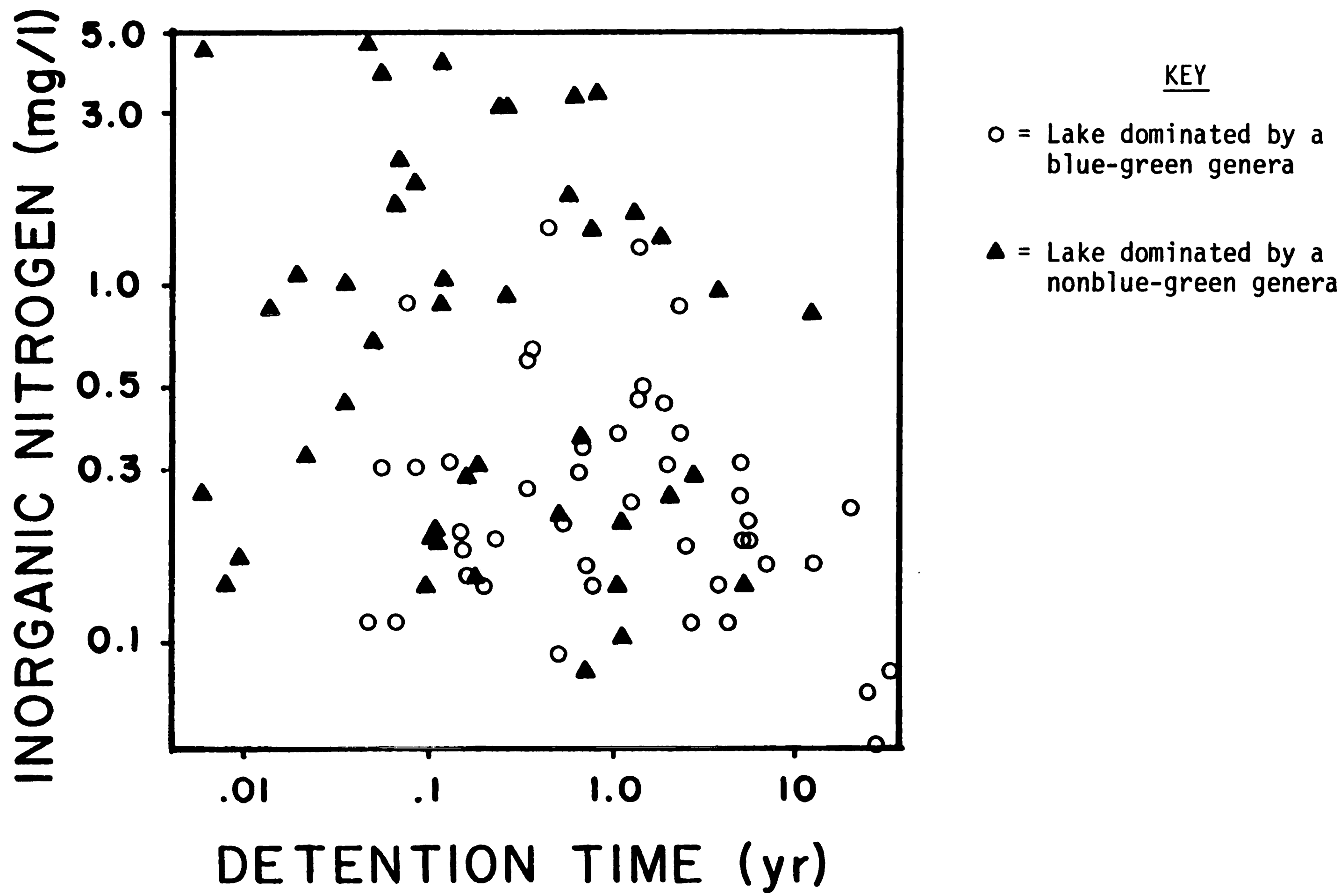

Figure 10.--Bivariate-discriminant plot of inorganic nitrogen versus detention time by algal-type dominance

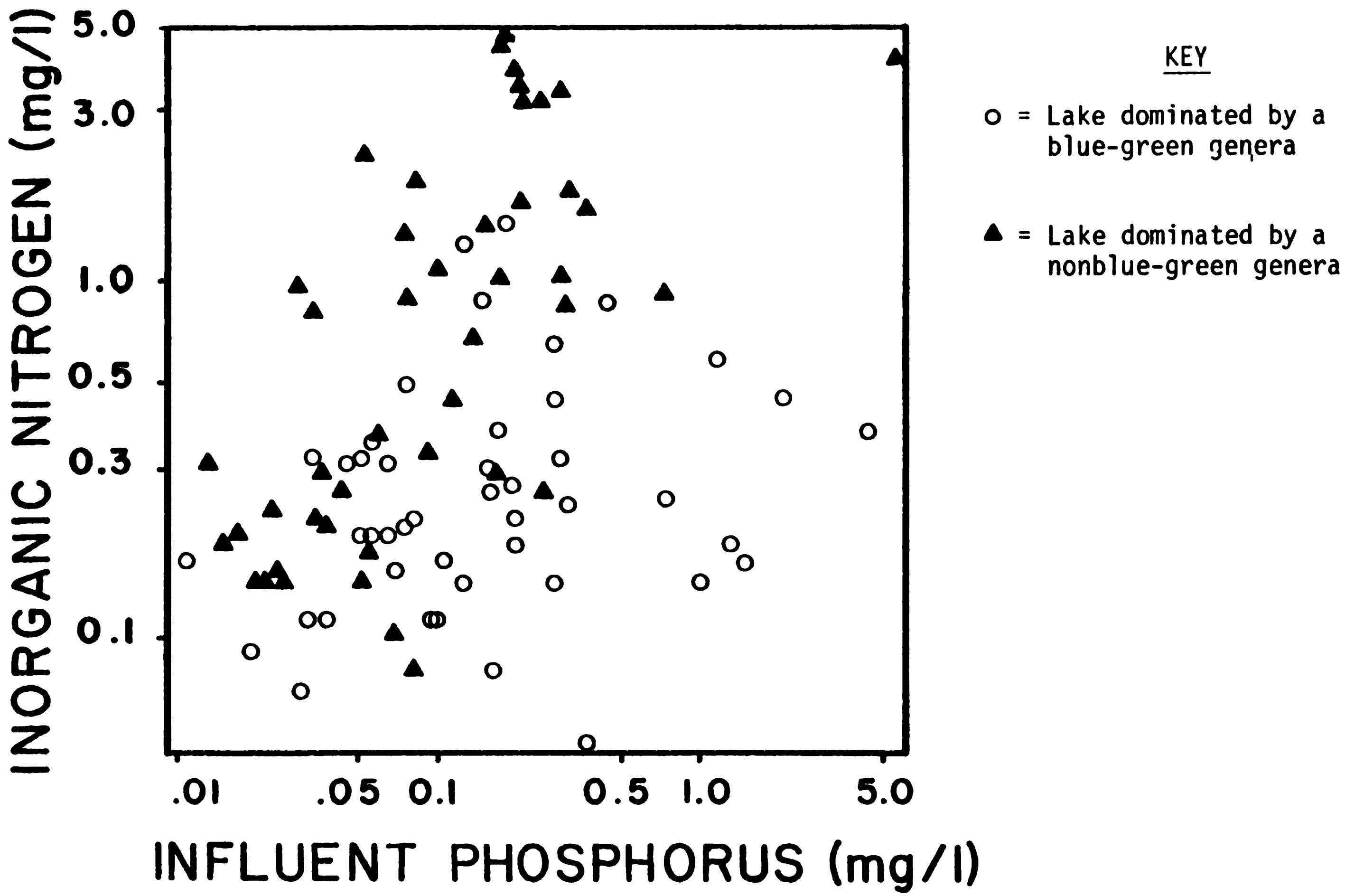

Figure 11.--Bivariate-discriminant plot of inorganic nitrogen versus influent phosphorus by algal-type dominance

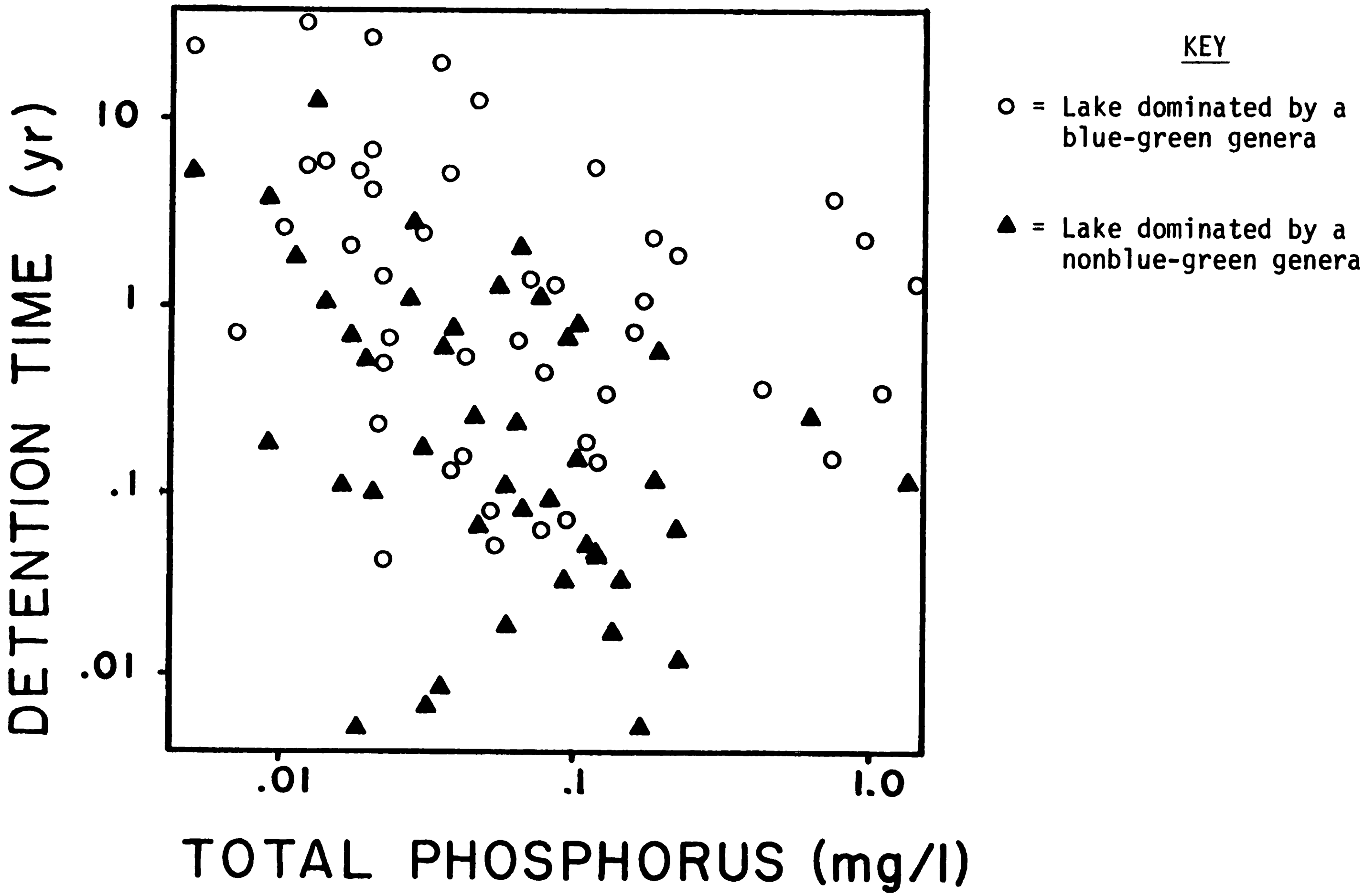

Figure 12.--Bivariate-discriminant plot of detention time versus total phosphorus by algal-type dominance

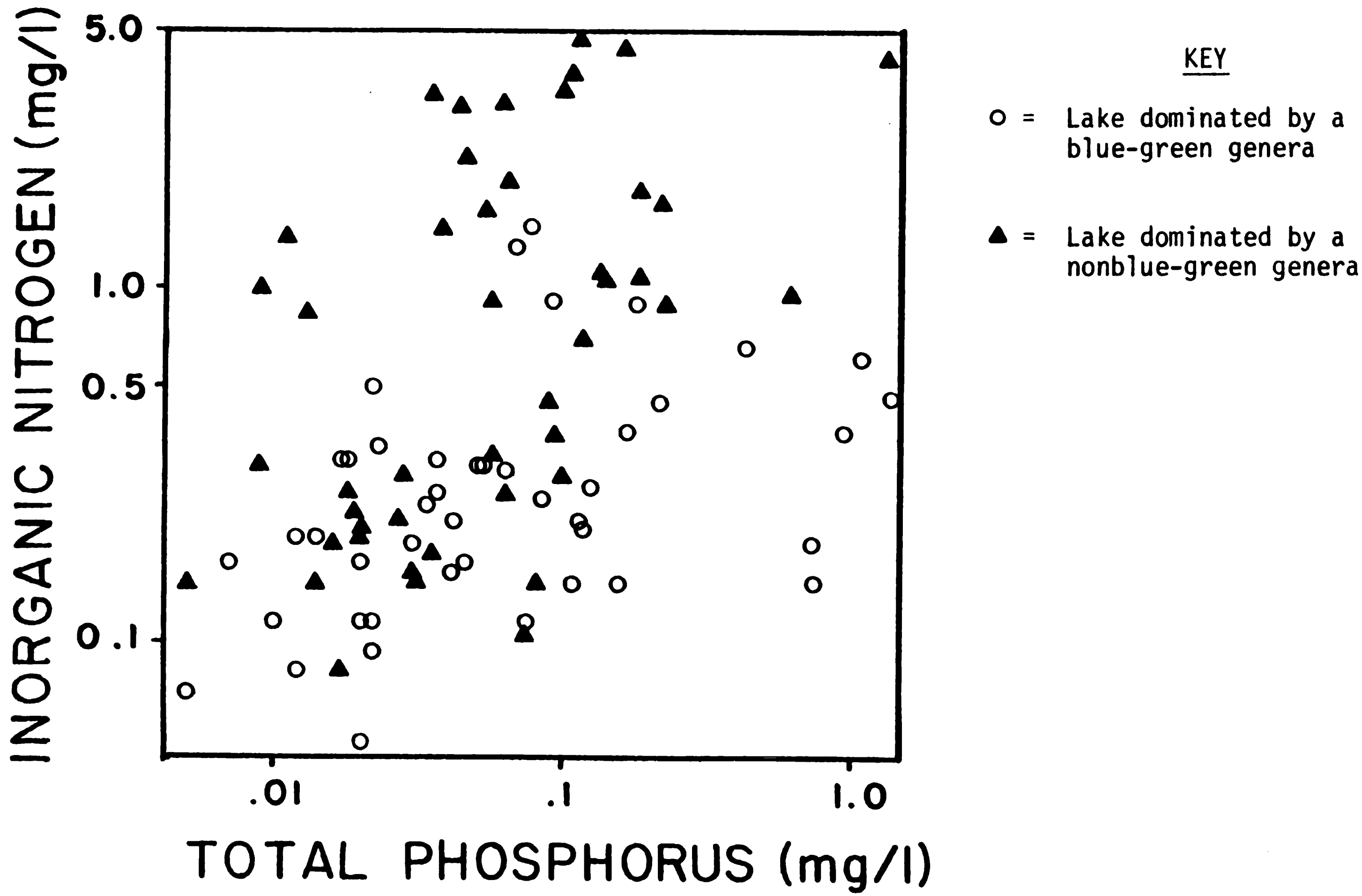

Figure 13.--Bivariate-discriminant plot of inorganic nitrogen versus total phosphorus by algal-type dominance

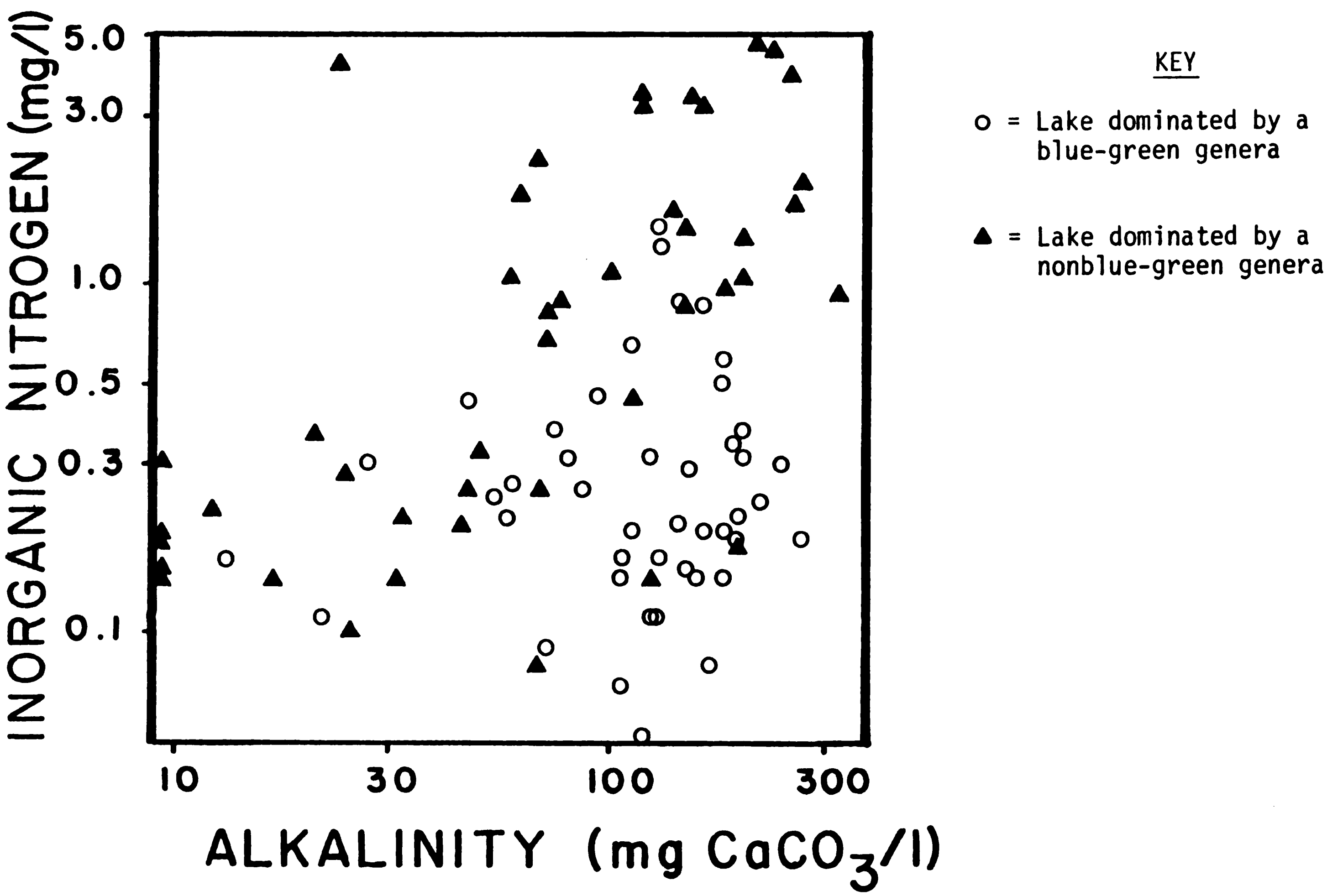

Figure 14.--Bivariate-discriminant plot of inorganic nitrogen versus alkalinity by algal-type dominance

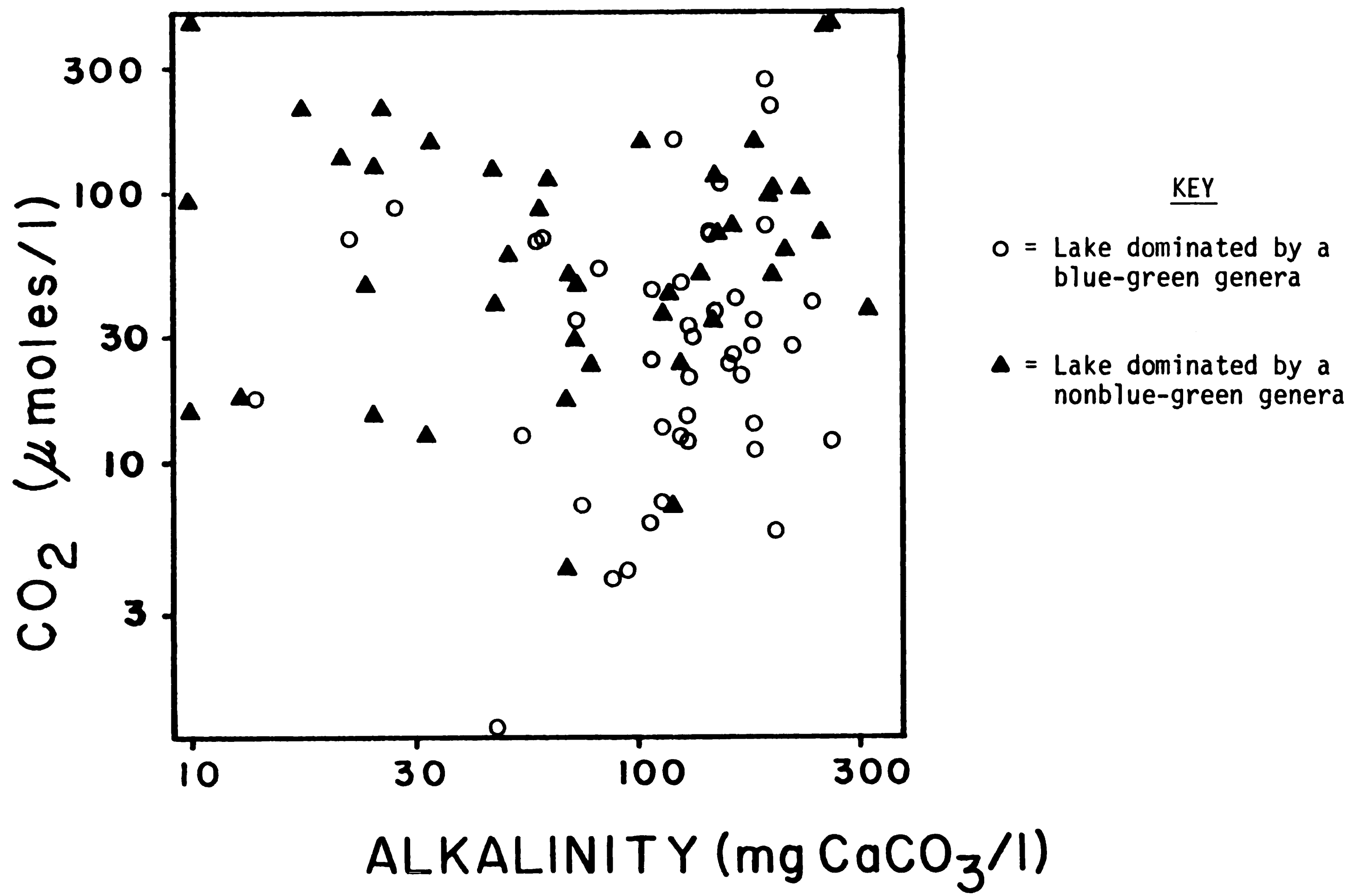

Figure 15.--Bivariate-discriminant plot of carbon dioxide versus alkalinity by algal-type dominance

Empirical analysis of the EPA-NES data set appears to confirm this conclusion, although, based on Figure 14, inorganic nitrogen is important only in the higher alkalinity lakes.

One theory that would explain this phenomenon is presented by King (1979). Assuming a natural phosphorus limit at the onset of cultural eutrophication, King (1970, 1972) states that the alkalinity and the phosphorus concentration are the primary determinates of whether or not aquatic growth progresses to the point where carbon or nitrogen becomes the most important limiting factor of growth. Recall that the alkalinity can serve as a reserve carbon source for photosynthesis. A lake with low alkalinity (i.e., a low carbon reserve) would require little phosphorus to initiate carbon limiting conditions and stress the algae causing a dominance transfer to bouyant blue-greens. Lakes with a high alkalinity (i.e., a large carbon reserve) would not be expected to reach the carbon limit so easily, therefore, nitrogen would be more likely to be of primary qualitative importance when phosphorus limits are nullified (via cultural loading).

Figure 14 appears to empirically confirm that the nitrogen concentration is an important variable in determining algal dominance in the high alkalinity EPA-NES lakes. Further investigation of bivariate-discriminant plots failed to define any strong independent/dependent variable relationships in the low alkalinity EPA-NES lakes, however. King (1970, 1972) points to the CO_2 concentration as being very important in these lakes, but examination of the bivariate-discriminant plot of CO_2 versus alkalinity (Figure 15) reveals no strong relationship.

Note, however, that there are only a few low alkalinity lakes on which to base this conclusion and recall also the uncertainty surrounding the calculation of the CO_2 value. These may be reasons why the CO_2 variable was not found to be important in the low alkalinity EPA-NES lakes.

Since further investigation was unable to identify strong discriminating variables for the few low alkalinity lakes, only high alkalinity lakes were used in the discriminant analysis procedure described below (i.e., the model-building phase).

Discriminant Analysis of High Alkalinity EPA-NES Lakes

The step-wise discriminant analysis program used in this investigation of the 68 high alkalinity EPA-NES lakes is from the Statistical Package for the Social Sciences (SPSS). The linear discriminant function was constructed using the overall multivariate F ratio for the test of difference between group centroids. The variables were chosen to maximize the statistical effect of separation of the pre-defined algal groups; blue-green dominated lakes and nonblue-green dominated lakes. Addition details regarding the step-wise procedure and function calculation may be found in the SPSS user's manual (Nie et al., 1974).

The results of the discriminant analysis is presented in Table 8. Inorganic nitrogen (N), as expected, was the first variable to enter the function (i.e., it is the variable that best discriminates among the pre-defined groups). Hydraulic detention time (T) and influent phosphorus concentration (INP) entered next but displayed must less discriminatory power as observed via the standardized discriminant coefficients.[1] The discriminatory effect displayed by the entry of subsequent variables was deemed insignificant by this investigator and eliminated from the function.

Table 8: Discriminant analysis results

Variable	Step-Wise Analysis		Standardized Coefficients
	Step	Wilk's Lambda	
log N	1	.572	1.240
log INP	2	.538	-0.338
log T	3	.524	-0.262

Discriminant analysis of the variables (log) N, (log) T, and (log) INP yielded the following discriminant function:

$$\text{d.f.} = \frac{10^{.129}\, N^{2.463}}{T^{.284}\, INP^{.726}} \tag{5.2}$$

$$\text{d.s.} = \log \text{d.f.} \tag{5.3}$$

where:

d.f. = the discriminant function

d.s. = the discriminant score

The function is the first principal component and maximizes the group separation on a single axis (Reckhow, 1979).

[1] The examination of standardized discriminant coefficients was considered a valid method of assessing discriminatory power because of the relatively low correlation between the independent variables contained in the function.

The discriminant analysis information may also be expressed in terms of classification functions, one for each group (Nie et al., 1974). The following are the SPSS derived classification functions (c.f.) for the two algal groups.

1) Classification function for the blue-green group:

$$c.f._{(1)} = -3.85(\log N) - .56(\log T) - 2.7(\log INP) - 2.21 \quad (5.4)$$

2) Classification function for the nonblue-green group:

$$c.f._{(2)} = 1.12(\log N) - 1.10(\log T) - 4.08(\log INP) - 2.29 \quad (5.5)$$

A case is classified into the group that yields the highest classification function score.

Figure 16 is a plot of the discriminant function (Equation 5.2) with the log of the numerator on the x-axis and the log of the denominator on the y-axis. The classification boundary occurs approximately where d.s. = .17. For example, if d.s. < .17 a case is classified in the "blue-green dominated" category. Alternatively, if d.s. > .17, a case is classified in the "nonblue-green dominated" category. The classification phase of the SPSS discriminant analysis program resulted in 86.8% of the EPA-NES high alkalinity data lakes being correctly classified.[1] As observed in Figure 16, group separation is distinct, however, misclassification do exist. Investigation of these misclassified lakes revealed no apparent characteristics that would indicate

---

[1]Most likely there is an upward bias of classification results since the discriminant function is optimumly fitted to the data set when the function was constructed (Morrison, 1969). The jackknife technique, a method to assess classification bias, is not available with SPSS, and thus could not be used.

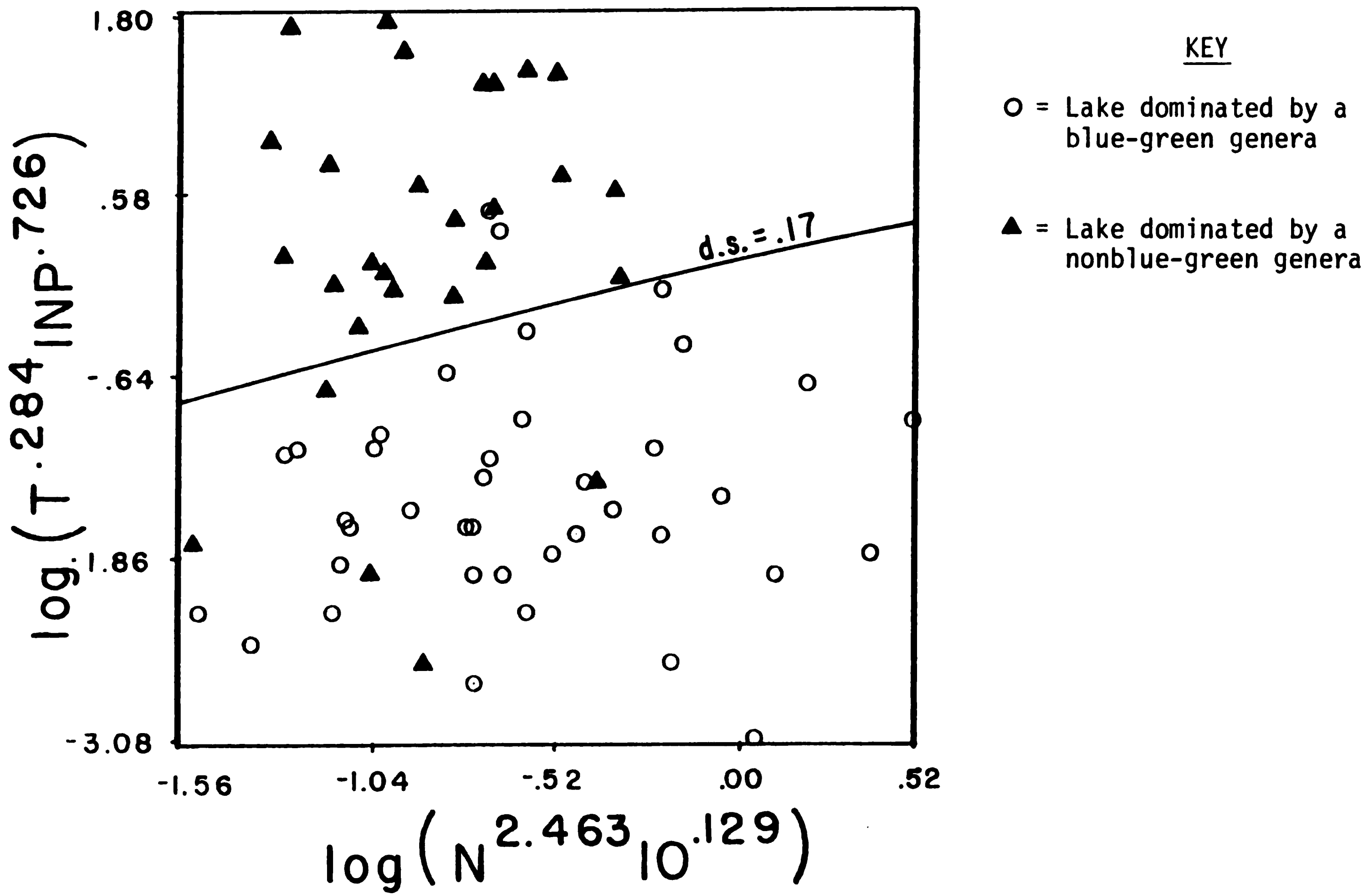

Figure 16.--The algal-type dominance discriminant function

non-representativeness in either the data collection or assessment.

The presence of misclassified lakes is not really surprising since the discriminant function represents a simplification of the real world and cannot be expected to account for all the variability found in multi-lake analyses. Recall too, the dynamic character of algal populations and that an EPA-NES lake was originally classified into the pre-defined algal groups based on the dominant genera occupying the water column at the time the sample was taken. Importantly, successional dynamics are unique to each lake and an algal type that dominates at any particular time may not be dominant two weeks later. The waxing and waning of populations occurs continuously but at different rates in individual lakes (Fogg, 1965).

Since an equation (or model) such as Equation 5.2 represents a simplification of the real world, the practical application of the discriminant function (or the classification functions) requires an evaluation of the uncertainty of the results. A presentation of this uncertainty for a given prediction may be considered as a measure of the value of the information provided by the model (Reckhow, 1979). It is possible to express the uncertainty inherent in the model probabilistically, and then present the probabilities graphically. The SPSS discriminant analysis program calculates classification probabilities using the following function:

$$P(G_j/X) = \frac{P_j|D_j|^{-1/2} e^{-X_j^2/2}}{\sum_{i=1}^{g} P_j|D_j|^{-1/2} e^{-X_j^2/2}} \qquad (5.6)$$

where:

P_j = the prior probability for group j (.50 in this study)

D_j = the group covariance matrix for group j

g = the number of groups

X_j^2 = the chi-square distance from each group centroid

Classification probabilities for the EPA-NES lakes were calculated and matched with the associated discriminant scores (calculated from Equation 5.2) to construct Figure 17. Because the discriminant function incorporates the uncertainties in the model development data set into the probability estimates, Figure 17 can be used without additional uncertainty estimates if the techniques used to survey an application lake are similar to the methods used by the EPA-NES (Walker, 1977; Reckhow, 1979).

Since the discriminant function can be used as a predictive model, a few limitations should be mentioned. First, a model should not be applied to a lake with variable values more than the maximum values or less than the minimum values contained in the data set used to construct the model. Table 9 presents the ranges of the independent variable values used to construct the discriminant function. Further, the functions were constructed only from lakes within the north temperate climatic zone and thus should only be applied to lakes within this zone. Also, recall that the EPA and this investigator placed selection criteria for lake inclusion in the EPA-NES and the model-building data set respectively (see Chapter III). Each of these criteria may also

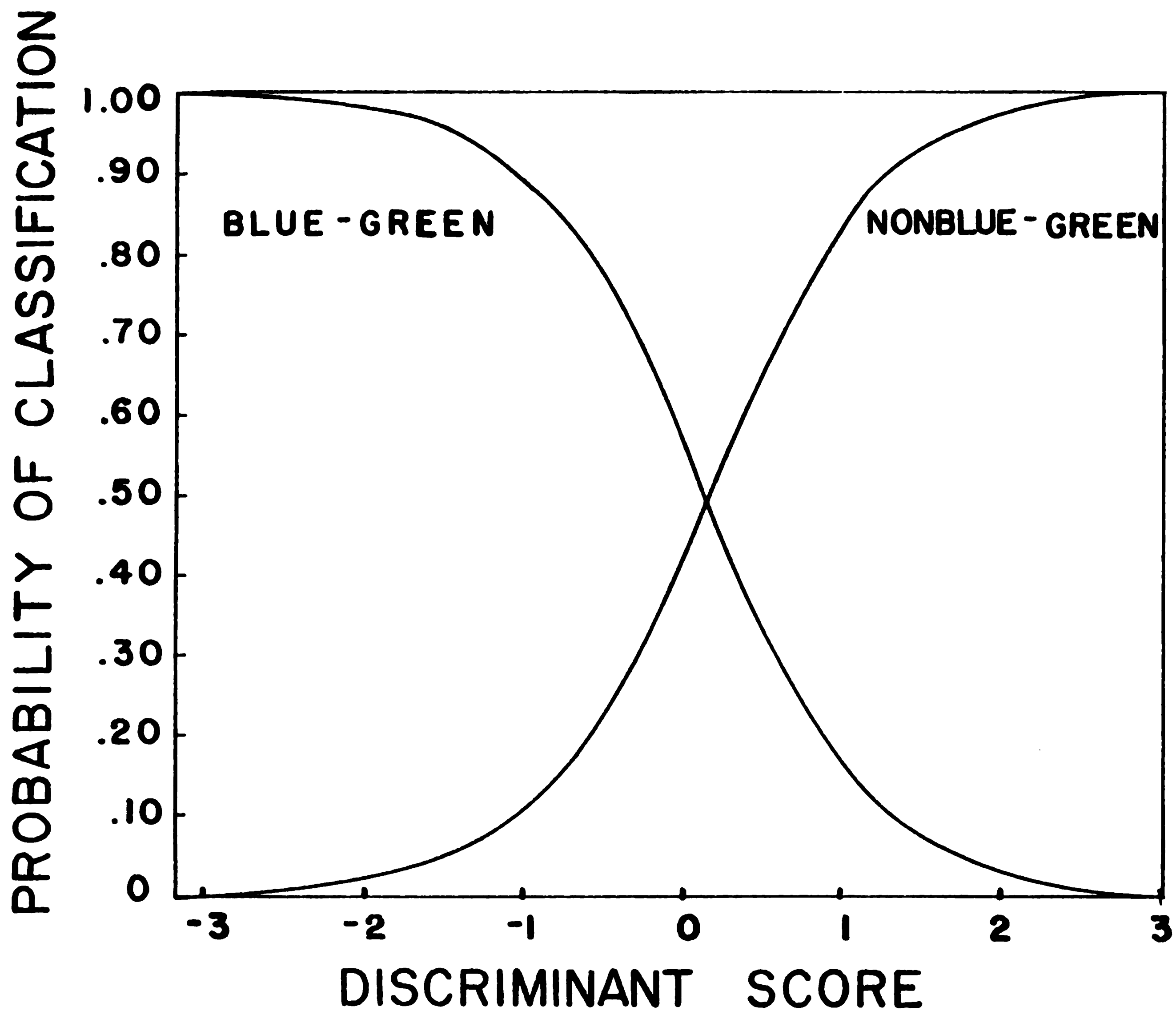

Figure 17.--Discriminant score assessing the probabilities of algal-type dominance classification

be considered a limit of model applicability. In summary, the empirical model should not be used on lakes different from those used to develop the model, without prior testing.

Table 9: Minimum and maximum values for the data set used to develop the discriminant models.

Variable	Minimum	Maximum
N (summer)	.05 mg/l	4.62 mg/l
INP	.020 g/m^2-yr	4.308 g/m^2-yr
T	.008 yr	21.0 yr
Alk (summer)	60 mg $CaCO_3$/l	334 mg $CaCO_3$/l

A Case Study and Summary

One objective for recreational and water quality planners may be, in the future, the direct management of algal dominance in lakes (Shapiro et al., 1977). A mathematical model that predicts whether or nor obnoxious blue-green algae are likely to dominate can be very valuable in this regard, especially in lake basins faced with changing land use or water budgets. The discriminant function presented in the preceeding section represents an heuristic attempt to model algal-type dominance. In order to facilitate the understanding of the use of the discriminant function as a model, a brief case study is presented below.

Higgins Lake, located in the northern lower peninsula of Michigan, was selected from the EPA-NES data set and will be used to demonstrate the model. Table 10 provides the lake statistics necessary to apply the model. Substituting the variable values from Table 10 into Equation 5.2 and 5.3, one may estimate the discriminant score for Higgins Lake:

$$\text{d.f.} = \frac{10^{.129} \times .07^{2.463}}{15.6^{.284} \times .031^{.726}} \tag{5.7}$$

$$= .011$$

$$\text{d.s.} = \log .011 \tag{5.8}$$

$$= -2.0$$

Table 10: Higgins Lake data

Variable	
Summer N (mg/l)	= .07
T (yr)	= 15.6
INP (mg/l)	= .031
Summer Alkalinity (mg $CaCO_3$/l)	= 109

Figure 17 graphically presents the probabilities associated with algal group classification given the discriminant score. The d.s. for Higgins Lake equals -2.0, thus (according to Figure 17):

.98 = the probability of the lake being dominated by blue-green algae.

.02 = the probability of the lake being dominated by nonblue-green algae.

Higgins Lake was selected not only to demonstrate the use of the model, but to also point out the fact that only algal dominance is addressed, in terms of prediction results. The quantity of algae, however, is not. Based on EPA-NES summer phytoplankton sampling, Higgins Lake was indeed dominated by the blue-green algae genera *Lyngbya* (USEPA, 1975). Traditionally, blue-green algal dominance is equated with eutrophic conditions. However, Higgins Lake is regarded as an oligotrophic lake. In fact, of all the forty-one Michigan lakes sampled by the EPA-NES, none had greater mean Secchi disc transparency and none had less mean chlorophyll *a* concentration than Higgins Lake (USEPA, 1975).

Of course, both algal quantity and quality are important in determining the use and enjoyment of a lake. Thus, when applying the model, it is important to be aware that the quantitative aspect of algal populations is not addressed in this research nor in the discriminant function. In the case of Higgins Lake, a planner's management objective would not likely focus on blue-green algal dominance. Rather, management efforts would be more likely (and logically) focused on maintaining low phosphorus input to the lake and therefore limit algal growth to low levels without regard to qualitative dominance. Many other lakes, nevertheless, are at least potentially faced with accelerated cultural eutrophication and it is for these lakes that the algal-type dominance models best apply.

CHAPTER VI

CONCLUSIONS

Cultural investment, development and perturbation within a lake drainage basin disturbs the natural watershed/lake equilibrium; however, in time, lake biodynamics tend to re-equilibrate with the new flux of particulate and dissolved material and nutrients from the watershed. Unfortunately, this new equilibrium can result in eutrophic symptoms and effects that may be highly irritating to landowners and users of the lake. Faced with this situation they often demand that the lake be restored to its "natural" state. The resultant lake restoration efforts can be classified into four basic categories:

1) mechanical (e.g., harvesting or underwater mowing);
2) chemical (e.g., algaecides);
3) biological (e.g., virus introduction or fish grazing);
4) ecological (e.g., nutrient diversion or waste water treatment).

Each approach has met with varying degrees of success in practical application. Mechanical and chemical efforts function primarily as "cosmetic" treatments and are not usually initiated until the water quality has been severely impaired. This is also true for most biological approaches which require, in addition, an in-depth understanding of long-term biological systems and population cycles. For these reasons, the ecological approach towards controlling cultural eutrophication is frequently the most logical, and in the long term, the most mitigative strategy. This approach usually involves curtailing excessive nutrient inputs to a lake through conscientious basin planning and management, and therefore, unlike the "cosmetic" treatments, attacks the causes rather than the symptoms of cultural eutrophication.

In developing a long-term ecological approach to water quality, a careful examination is needed of all the factors that contribute to, or are associated with, cultural eutrophication. For example, although algae are a natural and important component of the lake ecosystem, many lakes in Michigan and the nation face a water quality problem in the form of the dominance of undesirable blue-green algae.

A comprehensive approach towards the complete understanding of algal dominance as a nuisance condition requires examination of such factors as; essential and limiting growth requirements, growth rates, environmental niches and relative competitive abilities among the algae. Obviously, it is only through complete knowledge that we will be able to manage aquatic plant life successfully and as an asset to man.

This research presents a new approach to the study of the blue-green dominance problem. Through the use of a large set of multivariate data, chemical and physical parameters and algal-type dominance were studied empirically using a variety of statistical techniques. Other goals of this study were to; provide a working example of how water quality problems may be approached using empirical data analysis and, to show how empirical research can be used in the support (or non-support) of limnological theory.

Conclusions based on this empirical research are listed below:

1) Various techniques of empirical data analysis such as box plots, bivariate-discriminant plots and discriminant analysis can be used to identify which independent variables appear to be most important in determining whether or not a case is likely to belong in some dependent variable grouping.

2) Examination of box plots and the "Snedecor-Cochran statistic" revealed that the inorganic nitrogen concentration, the hydraulic detention time, the free carbon dioxide concentration and the pH are statistically the most important independent variables in determining whether a lake is dominated by blue-green or nonblue-green algae. Limnological theory indicates, however, that these variables are both dynamic and interactive. Because of the uncertainty in pH sampling methods, the pH variable and the subsequently calculated CO_2 variable were excluded from further analysis.

3) Examination of bivariate-discriminant plots revealed that the inorganic nitrogen concentration appeared to be important only in the high alkalinity lakes (lakes with greater than 56 mg $CaCO_3$/l). Specifically, blue-green algae tended to dominate in high alkalinity lakes with an inorganic nitrogen concentration less than .7 mg/l. Nonblue-green algae tended to dominate in high

alkalinity lakes with an inorganic nitrogen concentration greater than .7 mg/l.

4) No strong independent/dependent variable relationships were identified in low alkalinity lakes (lakes with less than 56 mg $CaCO_3$/l).

5) The identification of a strong independent/dependent variable relationship in the high alkalinity lakes justified the development of a predictive model using discriminant analysis. As expected, the SPSS step-wise discriminant analysis program selected inorganic nitrogen as the best discriminating variable. Hydraulic detention time and influent phosphorus concentration were also subsequently added but displayed much less discriminatory power (as observed via standardized discriminant coefficients). Remaining variables did not add significant discriminating power to the function and were eliminated. The proposed model equation derived from the discriminant analysis is presented below:

$$\text{d.f.} = \frac{10^{.129}\, N^{2.463}}{T^{.284}\, INP^{.726}}$$

where:

N = the median summer inorganic nitrogen concentration (mg/l)

T = the hydraulic detention time (yr)

INP = the influent phosphorus concentration (mg/l)

> Classification probabilities were also a calculated and presented graphically. The SPSS classification phase of the discriminant analysis program correctly classified 86.8% of the 68 EPA-NES lakes used to derive the function.

Based on this research the inorganic nitrogen concentration appears to be the most important "control" variable in determining the dominance of blue-green algae in high alkalinity, eutrophic lakes. Thus, it can be theorized that planning and management attempts to reduce nitrogen loading to these lakes (and therefore reduce the lake inorganic nitrogen concentration) may result in a shift to blue-green algal dominance earlier in the growing season. Nitrogen fertilization, on the other hand, may allow green algae and diatoms to more effectively compete with the blue-greens during the summer.

The inorganic nitrogen concentration would be a very difficult variable to manipulate, however, because so many natural and cultural sources contribute to the total nitrogen loading of a lake. In addition, since nitrogen is very active in lake biogeochemical cycles, the acceleration of these cycles (via cultural eutrophication) increases the impact of the nitrogen loss mechanisms. In particular, the ammonia-nitrogen loss is now thought to have substantial effect on the total nitrogen budget of a lake, even at a moderate pH (King, personal communication).

Thus, the complexity of eutrophic lake systems generally precludes the use of inorganic nitrogen as a "control" variable for algal-type dominance, at the present time. Therefore, permanent success in the restoration of lakes must still focus on the ability to reduce the amount of phosphorus available for primary production (King, 1979).

In the majority of lake basins, however, present-day economic, social and institutional constraints make it naive to assume that lake basin development and management will progress, remain, or "backtrack" to practices that promote phosphorus control to a level conducive to proper lake management. Therefore, in many lake systems, relief from the consequences of artificial phosphorus loading and cultural eutrophy lies with repetitive (and expensive) in-lake treatments. For these reasons, it may become desirable to increase our knowledge of the nitrogen input/loss system since nitrogen feeding could become a way to control blue-green algae infestations. This management objective, of course, assumes that the resultant expansion of green algae could also be controlled. It is possible that further empirical, field and laboratory research could develop additional relationships involving algal interactions and life cycles. Assuming that many lakes are resigned to excessive phosphorus loading, such research may perhaps lead to the development of successful non-phosphorus related control methods. Such methods may someday prove to be more practical, cost-effective and long-term than current "cosmetic" mechanical or chemical treatments.

BIBLIOGRAPHY

BIBLIOGRAPHY

Baker, A. L. and Brook, A. J. 1971. Optical Density Profiles as an Aid to the Study of Microstratified Phytoplankton Populations in Lakes. Arch. Hydrobiol. 69, 214-233.

Barlowe, R. 1976. A Framework for Resource Development Policy. Ocassional Paper. Department of Resource Development, M.S.U., E. Lansing, MI.

Bernhardt, J. J. 1967. Aeration of Wahnback Reservoir Without Changing the Temperature Profile. Am. Water Works Assoc. 59, 943.

Birge, E. A. and Juday, C. 1922. The Inland Lakes of Wisconsin. The Plankton-1. Its Quantity and Chemical Composition. Bull. Wisc. Geol. Hist. Surv. 64 (Sci. Ser. 13), 222 pp.

Boyde, C. E. 1973. Biotic Interaction Between Different Species of Algae. Weed Science 21, 31-37.

Canter, H. M. and Lund, J. W. G. 1948. Studies on Plankton Parasites I. Fluctuations in the Numbers of *Asterionella formosa* (Hass.) in Relation to Fungal Epidemics. New Phytol. 47, 238-261.

Dillon, P. J. and Rigler, F. H. 1975. A Simple Method for Predicting the Capacity of a Lake for Development Based of Lake Trophic Status. J. Fish. Res. Board Can. 32 (9), 1519-1531.

Dugdale, R. C. and Neess, A. C. 1961. Recent Observations on Nitrogen Fixation in Blue-green Algae: in Algae and Metropolitan Wastes. R. A. Taft Sanitary Eng. Center, Cincinnati, Ohio, 103-106.

Edmondson, W. T. 1972. Nutrients and Phytoplankton in Lake Washington: in G. Likens (ed.), Nutrients and Eutrophication. ASLO Special Symp. 1, 1172-1188.

Fogg, G. E. 1965. Algal Cultures and Phytoplankton Ecology. The Univ. of Wisc. Press, 175 pp.

Fogg, G. E., Stewart, W. D. P., Fay, P., and Walsby, A. E. 1973. The Blue-green Algae. New York, Academic Press, 459 pp.

Frank, R. E., Massy, W. F., and Morrison, D. G. 1965. Bias in Multiple Discriminant Analysis. J. of Marketing Research. 2, 250-258.

Helfrich, L. A. 1976. Effects of Predation by Fathead Minnows Pimephales promelas, on Planktonic Communities in Small Eutrophic Ponds. Ph.D. dissertation. M.S.U., East Lansing, MI 56 pp.

Horne, A. J. and Fogg, G. E. 1970. Nitrogen Fixation in Some English Lakes. Proc. Roy. Soc. London (Ser. B). 175, 351-366.

Hutchinson, G. E. 1961. The Paradox of the Planton. Amer. Nat. 95, 137-146.

Hutchinson, G. E. 1967. A Treatise on Limnology. II. Introduction to Lake Biology and the Limnoplankton. New York, John Wiley and Sons, Inc. 1115 pp.

Hutchinson, G. E. 1973. Eutrophication: The Scientific Background of a Contemporary Practical Problem. Amer. Sci. 61, 269-279.

Jackson, D. F. 1965. Ecological Factors Governing Blue-green Algae Blooms: in Proc. of the 19th Industrail Waste Conf. Eng. Ext. Ser. 117, Purdue Univ. 402-420.

Keenan, J. D. 1973. Response of Anabaena to pH, Carbon and Phosphorus. J. of Envir. Eng. Div. EE5, 607-620.

King, D. L. 1970. The Role of Carbon in Eutrophication. J. Water Poll. Control Fed. 42 (12), 2035-2051.

King, D. L. 1972. Carbon Limitations in Sewage Lagoons: in G. Likens (ed.), Nutrients and Eutrophication. ASLO Special Symp. 1, 98-110.

King, D. L. 1978. The Roles of Ponds in Land Treatment of Wastewater: in Int. Symp. on Land Treatment of Wastewater. Vol. 2. U.S. Army Cold Regions. Res. Eng. Lab., Hanover, N.H.

King, D. L. 1979. Lake Measurements: in Lake Restoration: Proc. of a National Conference. USEPA 440/5-79-001, 71-76.

King, D. L. and Hill, M. T. 1978. Interacting Environmental Factors Which Control the Sinking Rate of Planktonic Algae. Final Rep. OWRT/A-090-Mich. Natl. Tech. Inf. Serv. Springfield, VA 40 pp.

Lund, J. W. G. 1964. Primary Production and Periodicity of Phytoplankton. Verh. Int. Ver. Limnol. 15, 37-56.

Lund, J. W. G. 1965. The Ecology of the Freshwater Phytoplankton. Biol. Rev. 40, 231-293.

McGill, R., Tukey, J. W., and Larsen, W. A. 1978. Variations of Box Plots. The Am. Stat. 32 (1), 12-16.

Morrison, D. G. 1969. On the Interpretation of Discriminant Analysis. J. of Marketing Res. 6, 156-163.

Morrison, D. G. 1974. Discriminant Analysis: in R. Ferber, ed. Handbook of Marketing Research. New York. McGraw-Hill. 442-457.

Morton, S. D. and Lee, T. H. 1974. Algal Blooms-Possible Effects of Iron. Env. Sci. and Tech. 8, 673-674.

Moss, B. 1969. Vertical Heterogeneity in the Water Column of Abbot's Pond. II. The Influence of Physical and Chemical Conditions on the Spatial and Temporal Distribution of the Phytoplankton and on a Community of Epipelic Algae. J. Ecol. 57, 397-414.

Moss, B. 1972. Studies on Gull Lake,Michigan. I. Seasonal and Depth Distribution of Phytoplankton. Freshwat. Biol. 2, 289-307.

Mosteller, F. and Tukey, J. W. 1977. Data Analysis and Regression. Addison-Wesley Pub. Co. Reading, Mass. 588 pp.

Munk, W. H. and Riley, G. A. 1952. Absorption of Nutrients by Aquatic Plants. J. Mar. Res. 11, 215-240.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. (eds). 1974. Statistical Package for the Social Sciences. 2nd edition, McGraw-Hill. 675 pp.

Palmer, C. M. 1962. Algae in Water Supplies. Public Health Serv. Publ. No. 657. Washington, D.C. 88 pp.

Park, K. 1969. Oceanic CO_2 Systems: An Evaluation of Ten Methods of Investigation. Limnol and Oceanogr. 14, 178-186.

Pavonii, J. L., Tenney, M. W., and Echelberger, W. F., Jr. 1971. The Relationship of Algal Exocellular Polymers to Biological Flocculation. Eng. Bull. Purdue Univ. Proc. 26th Indust. Waste Conf. Eng. Ext. Series No. 140, Part II, 947-974.

Pritchard, G. G., Griffin, W. J., and Whittingham, C. P. 1962. The Effect of Carbon Dioxide Concentration, Light Intensity and Isonicotinyl Hydrazide on the Photosynthetic Production of Clycolic Acid by Chlorella. J. Exp. Botany, 13 (38), 176-184.

Reckhow, K. H. 1977. Phosphorus Models for Lake Management. Ph.D. thesis. Harvard Univ., Cambridge, Mass. 210 pp.

Reckhow, K. H. 1978. Lake Quality Discriminant Analysis. Water Res. Bull. 14, 856-867.

Reckhow, K. H. 1979. Lake Data Analysis and Phosphorus Variability. Paper presented at the N. Amer. Lake Management Conf. M.S.U., East Lansing, MI April 16-18, 1979.

Reynolds, C. S. 1972. Growth, Gas Vacuolation and Buoyancy in a Natural Population of Planktonic Blue-green Algae. Freshwater Biol. 2, 87-106.

Richerson, P., Armstrong, R., and Goldman, C. R. 1970. Contemporaneous Disequilibrium, A New Hypothesis to Explain the "Paradox of the Plankton." Proc. Nat. Acad. Sci. 67, 1710-1714.

Rodhe, W., Vollenweider, R. A., and Nauwerck, A. 1958. The Primary Production and Standing Crop of Phytoplankton: in Perspectives in Marine Biology. (A. A. Buzzati-Traverso ed.) Univ. of Calif. (Berkley) Press. 299-322.

Sawyer, D. 1947. Fertilization of Lakes by Agricultural and Urban Drainage. J. New Eng. Water Works Assoc. 61, 650.

Schindler, D. 1978. Factor Regulating Phytoplankton Production and Standing Crop in the World's Freshwaters. Limnol. Oceanogr. 23, 478.

Shapiro, J. 1979. The Need for more Biology in Lake Restoration: in Lake Restoration: Proc. of a National Conference. USEPA 440/5-79-001, 161-168.

Shapiro, J., Zoto, G., Lamarra, V. 1977. Experimental Studies on Changing Algal Population from Blue-greens to Greens. Limnol. Res. Center. Univ. of Minn. Contribution No. 168.

Shilo, M. 1971. Biological Agents Which Cause Lysis of Blue-green Algae. Mitt. Int. Ver. Limnol. 19, 206-213.

Snedecor, G. W. and Cochran, W. G. 1967. Statistical Methods. Iowa State Univ. Press. 414 pp.

Standard Methods for the Examination of Water and Wastewater. 1971. 13th ed. Publ. jointly by the Amer. Public Health Assoc., Amer. Water Works Assoc. and Water Poll. Control Fed., Washington, D.C. 874 pp.

Stewart, A. G., Barnum, D. A., and Henderson, J.A. 1950. Algal Poisoning in Ontario. Can. J. Comparative Med. 14, 197-202.

Symons, J. M. 1969. Water Quality Behavior in Reservoirs. U.S. Public Health Serv. Publ. No. 1930.

Thomas, E. 1969. The Process of Eutrophication in Central European Lakes: in Eutrophication: Causes, Consequences, Correctives. Natl. Acad. Sci. (Natl. Res. Conn. Pub.). 1700, 20-49.

USEPA. 1975. National Eutrophication Survey Methods 1973-1976. Working Paper No. 175, Corvallis, Oregon.

USEPA. 1975. NES Working Paper on Higgins Lake, Michigan. No. 195. Corvallis, Oregon.

Vallentyne, J. 1974. The Algal Bowl. Misc. Spec. Publ. 22. Dept. Environ. Fish. Mar. Serv., Ottawa, Can. 186 pp.

Vollenweider, R. A. 1968. The Scientific Basis of Lake and Stream Eutrophication, with Particular Reference to Phosphorus and Nitrogen as Factors in Eutrophication. Technical Report to O.E.C.D. Paris, DAS/CSI/68. 27. 182 pp.

Vollenweider, R. A. 1975. Input-Output Models with Special Reference to the Phosphorus Loading Concept in Limnology. Schweis. Z. Hydrol. 37. No. 1. 53-84.

Walker, W. W., Jr. 1977. Some Analytical Methods Applied to Lake Water Quality Probelms. Ph.D. thesis. Harvard Univ., Cambridge, Mass.

Ward, C. H. and King, J. M. 1976. Fate of Algae in Laboratory Cultures: in Ponds as a Wastewater Treatment Alternative (E.F. Gloyna, J. F. Malina and D. M. Davis, eds.). Water Res. Symp. No. 9. Center for Res. in Water Res., Univ. of Texas (Austin).

Wetzel, R. G. 1975. Limnology. W. B. Saunders Co. Phil., London and Toronto. 743 pp.

Wirth, T., Dunst, R., Uttormark, P., Hilsenhoff, W. 1970. State of Wisc. DNR Res. Rept. No. 62.

Yentsch, C. S. and Menzel, D. W. 1963. A Method for the Determination of Phytoplankton Chlorophyll and Phaeophytin by Flourescence. Deep Sea Research. 10, 221-231.