A WATERSHED-BASED CLASSIFICATION SYSTEM FOR LAKES IN AGRICULTURALLY-DOMINATED ECOSYSTEMS: A CASE STUDY OF NEBRASKA RESERVOIRS

By

Henry Nii Nmai Bulley

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Geography

Under the Supervision of Professor James W. Merchant

Lincoln, Nebraska

December, 2004

DISSERTATION TITLE

A watershed-based classification system for lakes in agriculturally-dominated ecosystems: a case study of Nebraska reservoirs

BY

Henry N. N. Bulley

SUPERVISORY COMMITTEE:

Approved	Date
Signature	12/8/04
Dr. James W. Merchant Typed Name	
Signature	11-30-04
Dr. Steven J. Lavin Typed Name	
Signature	11/30/04
Dr. Kyle D. Hoagland Typed Name	
Signature	11/23/04
Dr. David B. Marx Typed Name	
Signature	10DEC04
Dr. Geoffrey M. Henebry Typed Name	
Signature	
Typed Name	

University of Nebraska Lincoln

A WATERSHED-BASED CLASSIFICATION SYSTEM FOR LAKES IN AGRICULTURALLY-DOMINATED ECOSYSTEMS: A CASE STUDY OF NEBRASKA RESERVOIRS

Henry Nii Nmai Bulley, Ph.D.
University of Nebraska, 2004

Adviser: James W. Merchant

The U.S. Environmental Protection Agency is charged with establishing national standards and criteria for assessing lake water quality. It is increasingly evident that a single set of national water quality standards that do not take into account regional hydrogeologic and ecological differences will not be viable. Lakes clearly have different inherent capacities to meet such standards. The principal objective of this study was to define and test a watershed-based classification procedure for identifying groups of lakes that have similar potential capacity to meet proposed water quality standards. The strategy employs variables such as watershed area, mean watershed slope, soil organic matter, soil pH, and soil erodibility. This study focused on reservoirs in Nebraska, an agriculturally-dominated area of the United States. A preliminary cluster analysis of 78 reservoirs was performed to determine the optimal number of Nebraska reservoir groups. Subsequently, a Classification Trees method was used to describe the structure of reservoir watershed classes and to develop a predictive model that relates watershed conditions to reservoir classes. Results suggest that Nebraska reservoirs can be represented by nine optimal classes, and that soil organic matter content in the watershed is the most important single variable for segregating the reservoirs. The cross-validation prediction error rate of the Classification Tree model was 26.33 percent. The Classification Tree-based watershed classification was then compared with discriminant

function analysis (DFA)-based reservoir watershed classification, as well as ecoregions-derived reservoir classes. Overall, both watershed-based classifications were more effective than ecoregions in accounting for variations in lake water quality characteristics. Furthermore, Classification Trees approach was more suited for the ecologically complex datasets than the DFA classification method. Because all geospatial data used in this study are available nationally, the procedure can be adopted throughout the United States. Through model refinement, the Classification Tree interface for watershed-based reservoir classification promises to provide water resources managers an effective decision-support tool in the management of reservoir water quality.

Keywords: Reservoirs; Water Quality; Lake Classification; Classification Trees; Decision Trees; Geographic Information Systems; Watershed Management.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my advisory committee, especially my advisor, Dr. James Merchant, for his guidance and direction. I will also like to thank Dr. Dave Marx and Dr. Kyle Hoagland for their keen interest in my progress, encouragement and constructive suggestions that helped shed light on aspects of my dissertation research; to Dr. Steve Lavin and Geoff Henebry for reviewing this dissertation.

This research was partially supported by the U.S. Environmental Protection Agency (EPA) Science to Achieve Results (STAR) program (Grant # R828635). I am also grateful to Dr. John Holz and Aris Holz with whom I have worked closely on the E.P.A project; and have provided assistance with water quality data. Kris Verdin and Norman Bliss, USGS EROS Data Center, provided critical assistance in using EDNA and STATSGO datasets respectively and their efforts are gratefully acknowledged. Josh Lear, Nebraska Department of Natural Resources, provided initial GIS water coverage from SSURGO database, and his effort is also appreciatively recognized.

I will especially like to express my heartfelt thanks and appreciation to my wife (My Jewel of the Nile), Michelle Bernadine Bulley, for her love, support, encouragement; and without whom this dissertation would not have been possible. Special thanks to my son, Kreistan Akwei Bulley, for his patience and understanding throughout my dissertation research.

"I am all that I could be because of your unfailing love and selflessness"

I also wish to recognize my colleagues, graduate students and staff of the Center for Advanced Land Management Information Technologies, as well as Conservation and

Survey Division for the collegial and supportive environment I have enjoyed during the course of my graduate program. I also wish to express my gratitude to the directors, and staff of Student Involvement and the Midwest Consortium for Service Learning in Higher Education at the Nebraska Union, for the warm and friendly working environment I enjoyed over the past year. Many thanks also to the faculty, staff and students of the Department of Geography for their support and friendship. Finally, I wish to thank the Director, faculty and staff of the School of Natural Resources with whom I interacted in various ways and capacities. You all made my graduate experience at University of Nebraska-Lincoln, a worthwhile experience.

Finally, I wish to express my heartfelt gratitude to my father (Mark Akwei Bulley) and siblings (Andrew, Rebecca, Gabriel and Richard) for the nurturing and joyful environment that shaped my life and career choices. Thank you all.

DEDICATION

This dissertation is dedicated to the memory of my son Tjaden Adotei Bulley. Furthermore, I wish to dedicate the dissertation to the memory of my beloved mother, Juliana Amkpa Clegg-Bulley, for her indomitable spirit of courage, strength and generosity that has profoundly influenced my life. Mom, you represent the great spirit of African Woman.

"... Accra Hearts of Oak! Never Say Die Unless The Bones Are Gone!"

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.0. General background

In recent decades substantial progress has been made in improving the quality of surface waters in the United States (Hawkins *et al.,* 2000; EPA, 2000; EPA, 2001); nevertheless, much work remains to be done in assessing the state of impairment of lake waters. Impairment implies that the existing water quality of a lake, as measured by selected criteria (e.g., nitrogen, phosphorus, chlorophyll-a, Secchi depth), exceeds a threshold value or standard that presumably reflects optimal attainable lake water quality conditions (or "reference" conditions) (Hughes, 1995; EPA, 2000; EPA, 2001). Such impaired waters are not suitable for designated uses such as drinking, irrigation, recreation or fishery (Carpenter *et al.*, 1998). The management of lake water quality requires an effective means to establish which lakes are most impaired (and, hence, may require restoration) and which lakes are least impaired.

It is estimated that about 43 percent of the 16.4 million hectares comprising the United States' lake area have been adequately assessed for water quality (EPA, 2000). Of the lakes that have been assessed, 45 percent are "impaired" and 9 percent of the impaired lakes are listed as threatened. Nutrients exported from agricultural lands contribute about 50 percent of water quality problems in impaired lakes (Figure 1.1) (EPA, 2000). Water quality standards are particularly difficult to establish for lakes located in areas highly modified by humans, such as agricultural landscapes of the Midwest. In these areas (a) few, if any, lakes may represent pre-settlement "reference"

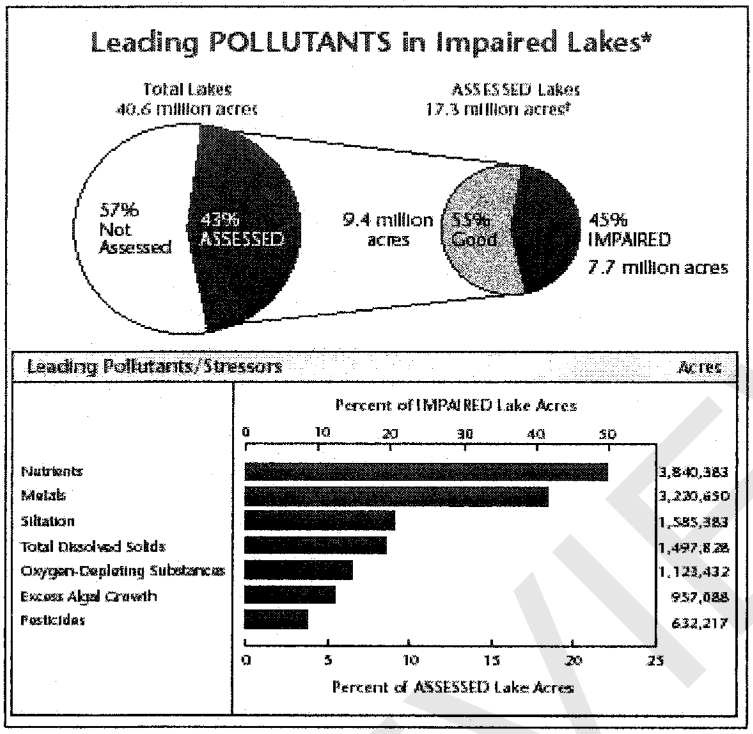

Figure 1.1. National Water Quality Inventory. Agricultural nutrients are the most common pollutants affecting assessed lakes; are found in 22% of assessed lakes; Contribute to 50% of reported water quality problems in impaired lakes.
Source: EPA (2000)

conditions, and (b) many lakes are human constructed (e.g., reservoirs). The principal objective of this research is to develop and evaluate an approach for establishing lake water quality standards using watershed-based classification of lakes.

Lakes are inland water bodies that serve as sources of drinking water, flood control, and outdoor recreation in addition to providing habitat for many wildlife species. The different types of lakes include natural lakes, reservoirs, and sand pits (or gravel pits) (Whittier *et al.,* 2001). Natural lakes were created as a result of geologic processes like glacial movement, while reservoirs in Nebraska were created by communities for flood

control, drinking water, irrigation, hydroelectric power, and recreation. Sand pits are generally by-products of road construction activities where the sand or gravel was removed to provide aggregate materials. Natural lakes and sand pits are fed primarily by lower order stream and groundwater respectively, so both natural lakes and sand pits usually have very small or negligible watersheds. On the other, the primary source of water for reservoirs is high order streams. The response of reservoirs to climatic conditions is intricately linked to the lakes' morphology and watershed characteristics. As such, a watershed approach to lake classification seems more applicable to reservoirs than natural lakes and sand pits. Consequently, the focus of this research is on the watersheds of Nebraska reservoirs. There are about 6796 reservoirs in Nebraska. While each lake is unique, it is impossible to manage all of these lakes individually. Moreover, the term lakes and reservoirs will be used interchangeably in the following paragraphs.

The U.S. Environmental Protection Agency (EPA) is charged with establishing national standards and criteria for assessing lake water quality. However, it is increasingly evident that a single set of national water quality standards that does not take into account the hydrogeologic and ecological differences among lakes will not be viable, since lakes have different inherent capacities to meet such standards (EPA, 2000; EPA, 2001). For example, in Nebraska, the EPA suggested criteria for the management of lake phosphorus (30 μgL^{-1}) has likely never been met in Nebraska lakes even under natural (pre-settlement) conditions. A more realistic standard might be about 60μgL^{-1} (John Holz, *pers. comm.*). This inconsistency is partly due to the fact that Nebraska lakes are typically assessed in the same manner as lakes in nearby regions, such as the Ozarks of

Missouri, which have very different hydrogeologic settings and relatively undisturbed environmental conditions.

A more tenable approach would be to define different standards for groups ("classes") of lakes determined to be similar to one another in terms of their potential to attain a certain level of water quality. Standards could then be established independently for lakes in different classes according to a set of "reference" target conditions unique for each class. Lake classification is used to group lakes into ecologically relevant or environmentally similar classes, enhance our understanding of complex systems, and to improve management and decision-making processes (Conquest *et al.,* 1994; Hawkins *et al.,* 2000). To be effective, a lake classification system designed to assess potential lake conditions must be based on environmental variables that underlie, determine, and explain the patterns of change in physical, chemical or biological water quality performances over seasonal or annual cycles (Warren, 1979). It is therefore important to differentiate between the natural or potential capacity of a lake to meet a certain level of water quality from actual water quality conditions that exist at a specific time of sampling.

The watershed classification approach that is proposed here is based on the premise that in the absence of human interference, lake ecosystems evolve in response to physical, chemical, and biological processes in their watersheds. It reflects an emerging emphasis on the watershed framework for water resource management (e.g., Warren, 1979; Satterlund and Adams, 1992; EPA, 1993; EPA, 1997; National Research Council, 1999; Mehan 2002; Bohn and Kershner, 2002). The lake watershed provides an important spatial framework to develop a classification system because it is the source of

runoff water, sediments and nutrients for lakes. In general, lake watersheds integrate the effects of all the natural and anthropogenic processes on water quality.

A watershed is a topographically defined area that collects all surface runoff and groundwater and discharges them into the lake up to the furthest downstream point (Ponce, 1989; Satterland and Adams, 1992). The term watershed has been used synonymously with drainage basin or catchments (Viessman *et al.,* 1977; Ponce, 1989; Satterland and Adams, 1992). Watersheds influence lake water quantity (e.g., peak flows and seasonal low flows) and quality (e.g., sedimentation rate and nutrient enrichment or eutrophication) (Welch, 1978; Warren 1979; Wetzel, 1983; Frissell *et al.,* 1986; Ponce, 1989; Satterlund and Adams, 1992; Omernik, 2003). This makes watershed boundaries the most appropriate spatial and topographic units for lake classification, assessment, and management.

1.1. Previous Research

Previous attempts to classify lakes have been based either on actual, measurable biochemical conditions of lakes, or on biogeographic characteristics of ecological regions or zones (Vollenweider 1968; Carlson 1977; Schindler 1971; Jensen and Van der Maarel, 1980; Omernik, 1987; Omernik *et al.,* 1991; Lomnicky, 1995; Niles *et al.,* 1996; Heiskary, 2000; Winter, 2001, EPA, 2002a; Jenerette *et al.* 2002; Moss *et al.,* 2003; Detenbeck *et al.,* 2004). Schindler (1971), Carlson (1977), and Heiskary (2000) for example, classified lakes based on indices of lake performance that required extensive or repeated sampling of lake water quality parameters, e.g. nitrogen, phosphorus, chlorophyll-a and Secchi depth. On the other hand, Omernik *et al.* (1991), Maxwell *et al.* (1995), Hargrove and Luxmoore (1998), Winter (2001), McMahon *et al.* (2001), EPA

(2002a) and Moss *et al.*,(2003) have developed landscape classification systems that may represent potential conditions of lakes and other water bodies, based on the characteristics of biogeographic or hydrogeologic regions, i.e., ecoregions and hydrologic landscapes.

Existing watershed-based classification systems for lakes and other water bodies have most often used actual water quality conditions in combination with topographic, soils, land use, and other data (Heywood *et al.,* 1980; Paulsen *et al.,* 1998; Momen and Zehr, 1998; Emmons *et al.,* 1999; Detenbeck *et al.,* 2000; Hawkins *et al.*, 2000; Johnson *et al.*, 2001; Lu and Lo, 2002; Bryd and Kelly, 2003; DeNicola *et al.,* 2004). Momen and Zehr (1998), for example, used discriminant function analysis (DFA) of lake water chemistry and land use data in a watershed-based lake classification. Emmons *et al.* (1999) compared DFA with a non-parametric statistical method, i.e. a decision tree model, in classifying northern Wisconsin lakes based on actual lake water quality data. They found that the decision tree method resulted in lower-rates of misclassification and more interpretable lake classes than those derived from DFA. Also, decision tree models can account for non-linear relationships, variable interactions and missing values in a given dataset (Breiman *et al.,* 1984; Verbyla, 1987; De'ath and Fabricius, 2000). Even though decision tree is a promising new tool for lake classifications it has not been applied extensively.

Other watershed-based classification systems and watershed assessments have been developed using the smallest (or fourth level) division of U.S. Geological Survey hydrologic units, i.e. hydrologic cataloging units (e.g., Smith *et al.* 1997, Griffiths *et al.,* 1999; EPA, 2002b; Bryd and Kelly, 2003, Papahicolau *et al.*, 2003). However, these

hydrologic units are not topographic watersheds and limitations of their use as surrogates for watersheds have been documented (e.g., Verdin and Verdin, 1999; Gesch *et al.*, 2002; Omernik, 2003).

According to Grigg (1965), "Classifications should be designed for a specific purpose since they rarely serve two purposes equally well". The purpose for classifying lakes in this research is in part to help the U.S. Environmental Protection Agency to establish reasonable attainable water quality standards ("targets") for groups of lakes that are considered to share similar potential capacity to meet these standards. Classification frameworks such as those cited above, while quite effective for a number of applications, do exhibit several major shortcomings for setting lake water quality standards. For example:

1. Lake classification based on observed water quality does not provide adequate insights into the potential of lakes to meet water quality standards for the following reasons:

- Human activities, such as land use, impact water quality.
- Water quality data represent observed water quality conditions, not the potential to meet a water quality standard.
- Extensive and frequent sampling of lakes in a given region is required, and lake sampling campaigns can be costly, in terms of personnel and equipment.
- Sampled lakes may not adequately represent the lakes in a given region.
- Lake water quality parameters are sometimes so variable that one lake may change classes over seasonal or annual cycles.

2. Omernik's ecoregions are inappropriate because they were based on subjective criteria of perceived patterns of land surface form, climate, vegetation, soils and land use.

Hence, these ecoregions can not be easily replicated. Also, ecoregion boundaries do not coincide with watershed boundaries, and the inclusion of land use data reflects the impact of human activities.

3. Attempts to delineate ecoregions via quantitative and objective methods (e.g. Hargrove and Luxmoore, 1998; Zhou *et al.,* 2003) are not appropriate because:

- These ecoregions include aspects of human influence, such as land use.
- The unit of analysis, e.g. 1 kilometer pixel of satellite data, does not take into consideration the terrain effect of watershed boundaries.

4. Existing watershed-based classifications are not appropriate because:

- They include aspects of human influence, such as measured water quality condition and land use.
- They are sometimes based on hydrologic cataloging units which do not conform to the natural hydrologic boundaries of the terrain.
- They are usually based on parametric statistical approaches such as discriminant function analysis (DFA) and regression analysis, which presume the use of normally distributed data, although most watershed data are multimodal and not normally distributed.
- Lake classes as well as some watershed data are categorical, and these types of data usually require transformations, when using traditional statistical approaches.

In summary, most lake classifications are based on observed, extant water quality data or on environmental variables that are often impacted by human activities and, thus, usually cannot be used directly for determining lake classes and subsequently setting lake reference conditions; data collection is expensive and time consuming. Regionalization

schemes, on the other hand, generally use subjective criteria for delineating boundaries (e.g., ecoregions) which do not coincide with watersheds. In both cases, there is an apparent arbitrary and often subjective choice of the number of classes.

This research focuses on the development of a watershed-based lake classification system that is based on: topographic boundaries that represent the lake watersheds; watershed characteristics that underlie, determine and explain the patterns of change in physical, chemical or biological water quality performances of lakes; and non-parametric statistical approaches that can account for the multimodal and categorical nature of watershed variables and lake classes.

1.2. Objectives and research questions

The primary objective of this dissertation is to develop a watershed-based approach to classify reservoir watersheds and to evaluate the effectiveness of the classification method to account for variations in water quality data that are pertinent to reservoir water quality management. The utility of Geographic Information Systems (GIS) and decision tree algorithms in developing a watershed-based approach to reservoir classification is also evaluated. This work is based on Nebraska reservoirs because most of the lakes in the state are constructed and located in agriculturally-dominated landscapes. Nebraska has a broad diversity of environments and landscapes, and is representative of many mid-latitude regions of the United States.

The specific research objectives are to:

1. *Determine the optimal number and characteristics of classes of Nebraska reservoirs based on their watershed characteristics. The research question that was addressed with*