Development of An Integrated GIS-Based System for Surface Water Quality Assessment and Management (GIS-SWQAM)

FANG LU

A Thesis
In the Department
of
Building, Civil and Environmental Engineering
Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Civil Engineering) at
Concordia University
Montreal, Quebec, Canada

October 2015

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Fang Lu

Entitled: Development of An Integrated GIS-Based System for Surface Water Quality Assessment and Management (GIS-SWQAM)

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Civil Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Marius Paraschivoiu — Chair

Dr. Chang-Hui Peng — External Examiner

Dr. Wen-Fang Xie — External to Program

Dr. S. Samuel Li — Examiner

Dr. Amruthur S. Ramamurthy — Examiner

Dr. Zhi Chen — Thesis Supervisor

Approved by

Chair of Department or Graduate Program Director

October 27, 2015

Dean of Faculty

ABSTRACT

Development of An Integrated GIS-Based System for Surface Water Quality Assessment and Management (GIS-SWQAM)

Fang Lu, Ph.D.

Concordia University, 2015

It is an fact that surface water receives a large volume of pollutants from industrial, agricultural, and municipal sources. The adverse health and environmental effects of surface water pollution have been a major concern in environmental management. Water quality models are useful tools to simulate the complex transport and fate of pollutants in a water body and predict the short-term and long-term effects on water quality variation. The emergence of spatial information technologies, such as Geographic Information System (GIS) make it possible to assess and predict surface water quality with more details with respect to spatial information.

The focuses of this thesis is to develop a comprehensive system named as GIS-SWQAM, which includes: (1) the development of a GIS-based water quality assessment system to assess the water quality and provide spatial distribution of water quality variables; (2) the development of an artificial neural network model to predict the change of water quality variables; (3) the development of a user interface that integrates the above models and functions; furthermore, a comparative analysis of the modeling approach developed in the GIS-SWQAM and the commercial model MIKE 21 was performed through field case studies.

The GIS-based water quality and ecological risk assessment models (MWQ module for marine water quality assessment and LWQ module for lake water quality assessment) are developed by integrating a fuzzy risk assessment model, a eutrophication risk assessment model, a heavy metal risk assessment model, a

dynamic database, the ArcGIS Engine, and a graphical user interface (GUI). The assessment results are both spatially and visually presented in the form of contour maps and color-coded maps that indicate risk levels. A large amount of data with both spatial and temporal distributions is managed by the developed system and analyzed by the assessment modules. The developed MWQ and LWQ modules are respectively applied in the Liaodong Bay of China and Lake Champlain. The MWQ and LWQ produce risk maps that depict the spatial distribution of integrated water quality index values, eutrophication risk levels and heavy metal risk levels in the study area. The maps generated can provide a better understanding of the distribution of the water quality and ecological risk levels. The primary factors that affect the water quality are subsequently examined using the visualized results.

An artificial neural network model with the back-propagation algorithm (BPANN) is first developed using Matlab to predict the chlorophyll-a concentration in Lake Champlain. Then, the algorithm of the BPANN model is built using the C# programing language and integrated with GIS and the database to build the ANN module, which is applied to predict the total phosphorus concentration in Lake Champlain. The best performing model is determined among the results of models built with different combination of input variables, which are preliminarily selected by linear correlation analysis and domain knowledge. Subsequently, the performances of the BPANN models are validated by a new set of field data. Similar to the MWQ and LWQ modules, the ANN module also produces the spatial distribution maps of the predicted concentrations; errors made during the prediction are presented in the user interface. The results indicate that the developed BPANN models can provide acceptable prediction results and can be used to provide a quick modeling assessment of water quality variation for managers.

In this thesis, the MIKE 21 FM software is also used to establish a hydrodynamic model coupled with a transport model to simulate the total phosphorus concentration in Lake Champlain. A comparative analysis is performed between the results of the

MIKE 21 model and the BPANN model. The results of the MIKE 21 model are acceptable, but not as good as that of the BPANN model. This further verifies that the developed BPANN model is a reliable tool to assess the lake eutrophication and to help managing lake water quality. The developed system can be also applied to surface water management in other area.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep and sincere gratitude and acknowledge to my supervisor, Dr. Zhi Chen, for his guidance and patience throughout my Ph.D. study. His academic advices have helped me establish a solid foundation for my research and encouraged me to furnish my best effort, leading to the successful completion of the present study and submission of the thesis in the present form.

I would also like to express my sincere thanks to my friends and colleagues: Yuan Yuan, Wen Ma, Dr. Baozhen Wang, Guan Wang, Likai Geng and all of them who have directly or indirectly helped me to complete this work.

Last but not the least, I wish to extend my deepest gratitude and appreciation to my families for their unconditional support and encouragement far away in China. In particular, my husband Wenquan Liu, I have been supported by him who made this long journey much easier. This thesis is simply impossible without his support and help. I must say a thank-you to him from the bottom of my heart for everything he has done to me that no word could explain.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Section 3.2.

$r(x)$	membership function
i	number of assessment factors
j	number of assessment criteria levels
x_i	analyzed value of assessment factor I (M/L^3)
S_{ij}	assessment threshold of the ith assessment factor at level j (M/L^3)
R	fuzzy matrix
W	weight set
w_{oi}	weight after normalization
S_{oi}	average value of all the assessment thresholds of factor I (M/L^3)
b_j	weight of the value of ith assessment factor at level j
r_{ij}	membership degree of the ith assessment factor to the assessment criteria at level j
B	fuzzy synthetic assessment vector
k	coefficient to control the role of maximum value
B^*	fuzzy marine water quality index

Section 3.3

E	Eutro phication index
C_{COD}	concentration of chemical oxygen demand (M/L^3)
C_{DIN}	concentration of inorganic nitrogen (M/L^3)
C_{DIP}	concentration of phosphate (M/L^3)
SD	Secchi Depth (M)
T_r^i	toxic-response factor for a given substance
C_s^i	concentration of heavy metal in the water samples (M/L^3)
C_n^i	reference concentration of heavy metals to which the risk is considered (M/L^3)

Section 3.4

i	number of neurons in the input layers

j	number of neurons in the hidden layers
k	number of neurons in the output layers
x_i	value of the input variable (M/L^3)
X_j	results obtained by the active function from the input neurons
w'_{ij}	connection weight between the input and hidden neuron
w''_j	connection weight between the hidden neuron and output neuron
b_{jh}	bias for the j_{th} neuron
b_{ko}	bias for the k_{th} neuron
Y_N	data value after normalization
x_{max}	maximum values of data (M/L^3)
x_{min}	minimum values of data (M/L^3)
y_{max}	maximum value of the specific range determined by the user
y_{min}	minimum value of the specific range determined by the user
$\hat{y}_i$	estimated response of the BPANN model (M/L^3)
y_i	target output of the BPANN model (M/L^3)
N_h	neuron numbers in the hidden layer
R^2	coefficient of determination
y_{mean}	mean value of the observed data (M/L^3)
n	total number of observations in each data set

Section 3.6

t	Time (t)
x, y, z	Cartesian coordinates
η	surface elevation (L)
d	equilibrium water depth (L)
$h = \eta + d$	total water depth (L)
u	velocity components in the x direction (L/t)
v	velocity components in the y direction (L/t)
$f = 2\omega \sin\phi$	Coriolis parameter (/t)
ω	angular rate of revolution (rpm)
ϕ	latitude

g	gravitational acceleration (L/t^2)
ρ	density of water (M/L^3)
ρ_0	reference density of water
s_{xx}	component of the radiation stress tensor
s_{xy}	component of the radiation stress tensor
s_{yx}	component of the radiation stress tensor
s_{yy}	component of the radiation stress tensor
P_a	atmospheric pressure (Pa)
S	magnitude of the discharge due to point sources (M/t)
(u_s, v_s)	velocity that the water is discharged into the ambient water (L/t)
T_{ij}	lateral stresses ((N/L^2))
c_f	drag coefficient (dimensionless)
$\vec{u}_b = (u_b, v_b)$	flow velocity above the bottom (L/t)
$\vec{u}_b$	depth-average velocity (L/t)
C	Chezy coefficient ($L^{1/2}/t$)
M	Manning roughness coefficient ($t/L^{1/3}$)
$\bar{\tau}_s$	Wind stress (N/L^2)
ρ_a	density of air (M/L^3)
c_d	drag coefficient of air (dimensionless)
$\bar{u}_w = (u_w, v_w)$	the wind speed 10 m above the water surface (L/t)
c_a	empirical factors
c_b	empirical factors
w_a	empirical factors
w_b	empirical factors
w_{10}	wind velocity 10 m above the water surface (L/t)
I	inviscid fluxes (L^2/t^2)

V	viscous fluxes (L^2/t^2)
U	vector of conserved variables
F	flux vector function
S	vector of source terms
A	area of the cell (L^2)
Δt	time step interval (t)
h_{wet}	wetting depth (L)
h_{dry}	drying depth (L)
h_{flood}	flooding depth (L)
D_v	vertical turbulent diffusion coefficient (L^2/t)
S	magnitude of discharge due to point sources (M/t)
F_C	horizontal diffusion term (L^2/t^2)
D_h	horizontal diffusion coefficient (L^2/t)
h	water depth (L)
$\overline{u}, \overline{v}$	depth-averaged velocity component (L/t)
C	concentration of scalar quantity (M/L^3)
k_p	linear decay rate of scalar quantity (/t)
C_S	concentration of scalar quantity in source (M/L^3)

Section 4.5

I_{i}	standard index of each parameter
C_i	measured concentration of each parameter (M/L^3)
S_i	evaluation criteria of each parameter (M/L^3)
DO_f	saturated concentration of dissolved oxygen in water (M/L^3)
DO_s	dissolved oxygen criteria (M/L^3)
t	water temperature (°C)
$I_{\max}$	maximum values of all standard indices
I_j	average values of all standard indices

Section 5.2

$Q_{ungaged}$	flow at the ungaged site (L^3/t)
Q_{gaged}	flow at the gaged site (L^3/t)
$A_{ungaged}$	drainage area of the ungaged station (L^2)
A_{gaged}	drainage area of the gaged station (L^2)

Section 7.4

D_{50}	the median grain diameter (L)
φ	latitude

LIST OF ACRONYMS

AD	Advection-Dispersion
AI	Artificial Intelligence
AML	ESRI ArcMacro Language
ANNs	Artificial Neural Networks
BOD	Biological Oxygen Demand
BPANN	Back-Propagation Artificial Neural Network
CEHQ	The Centre d'expertise Hydrique du Québec
Chl a	Chlorophyll-a
Cl	Chloride
COD	Chemical Oxygen Demand
CTD	An instrument used to determine the conductivity, temperature, and depth of the water body
DEM	Digital Elevation Model
DLL	Dynamic Link Library
DO	Dissolved Oxygen
DP	Dissolved Phosphorus
ECO lab	Environment Water Quality Model
EFDC	Environmental Fluid Dynamics Code
GIS	Geographic Information System
GUI	Graphical User Interface
HD	Hydrodynamic
IN	Inorganic Nitrogen
LMA	Levenberg-Marquardt Algorithm
LOAD	Non-point Load Evaluation Model
LWQ	Lake Water Quality Assessment Module
MDDELCC	The Ministry of Sustainable Development, Environment and the Fight against Climate Change
MIKE 11 HD	MIKE 11 Hydrodynamic Model
MSE	Mean Square Error
MWQ	Marine Water Quality Assessment Module

NAM	Rainfall-runoff Model
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
OECD	Organization for Economic Cooperation and Development
P	Phosphate
PCB	Polychlorinated Biphenyls
PT	Particle Tracking
RMSE	Root Mean Square Error
RSR	RMSE-Observations Standard Deviation ratio
SD	Secchi Depth
ST	Sediment Transport
SWAT	The Soil and Water Assessment Tool
T	Temperature
TMDL	Total Maximum Daily Loads
TN	Total Nitrogen
TP	Total Phosphorus
TSI	Trophic State Index
TSI-AVG	The Average Trophic State Index
U.S. EPA	United States Environmental Protection Agency
USGS	U.S. Geological Survey
WASP	Water Quality Analysis Simulation Program
WWTPs	Wastewater Treatment Plants

CHAPTER 1 INTRODUCTION

1.1 Problem Statement

Surface water quality encompasses a wide range of conditions that are part of the aquatic environment in a water body. In turn, the aquatic environment provides diverse habitats and a clean water supply for aquatic life, wildlife and humans; water quality issues influence human and environmental health. The development of industry and human population growth result in water quality degradation. Pollutants are discharged into the surface water directly and indirectly from urban run-off from impervious surfaces, inadequate infrastructure for storm and wastewater treatment, the discharge of treated wastewater from municipalities and industries, air deposition of contaminants, transportation, industrial production processes, and so forth. (Great Lakes Science Advisory Board et al., 2009, Sánchez-Avila et al., 2009, Olu-Owolabi et al., 2012). Meanwhile, the amount of pollutants discharged into surface water has significantly increased (Zhao et al., 2006). A large number of microbial agents, elements, and compounds may cause water pollution. They can be classified as microbiological organisms, biodegradable organic compounds, suspended matter, nitrates, salts, heavy metals, nutrients, and organic micropollutants (Helmer and Hespanhol, 2011).

Surface water quality assessment and prediction play important roles in environmental management. The managers can make decisions to control and remediate water pollution based on the assessment and prediction results, which provide useful information about water quality status and variation trends. Several methods have been proposed for surface water quality assessment, primarily including the water quality index method (Hurley et al., 2012), the fuzzy logic based method (Gharibi et al., 2012), the multivariate statistical method (Shrestha and Kazama, 2007, Kazi et al., 2009) and water quality assessment using biological indicators (Liao et al., 2011, Almeida et al., 2014). For water quality modeling and prediction, hydraulic and physicochemical water quality models are commonly applied, like CE-QUAL-W2, the Water Quality Analysis Simulation Program (WASP), the Environmental Fluid

Dynamics Code (EFDC), and MIKE models (Warren and Bach, 1992, Deus et al., 2013, Zhao et al., 2013, Bedri et al., 2014). Meanwhile, data driven models like artificial neural networks represent an alternative approach predicting water quality variation (Lee et al., 2003, He et al., 2011, Huo et al., 2013, Karakaya et al., 2013, Kim et al., 2013).

The methods mentioned above have both advantages and limitations. For water quality assessment, traditional water quality index methods are incapable of handling the uncertainties introduced by complex environmental factors; these methods provide assessment results for only a single parameter, rather than providing comprehensive assessment results of water quality in the area. These problems make the use of fuzzy logic methods favorable to the conventional index approaches in handling uncertainties and achieving comprehensive water quality assessments. However, the fuzzy logic-based methods require great computational efforts when the data quantity is large and the results are numbers without visually organized output. For water quality variable prediction, the mechanism models require extensive information that is not easily accessible, like the geometry of the study area and the hydraulic coefficients that are applied in the various hydrological sub-processes. Thus, artificial neural network models will be used as an alternative method in water quality prediction, since the ANN models are capable of reflecting the underlying linear or non-linear relationships amongst water quality data and do not require the specific internal processes (Dogan et al., 2009).

Recent studies have focused on the combination of water quality models with the Geographic Information System (GIS) in order to effectively handle spatial data, pollutant distributions and geophysical features. GIS includes powerful functions to visualize data and analyze results, which are useful in decision-making processes (Debaine and Robin, 2012). The integration of GIS and water quality assessment and prediction models combines their functions and advantages. This could have several benefits for environmental assessment and management processes.

In summary, every existing water quality model has limitations and there is little research regarding the interface system that integrates: 1) the fuzzy-set based methods of water quality assessment; 2) the artificial neural network models for water quality variable prediction; and 3) spatial output in the form of maps representing the degree of contamination.

Consequently, an integrated GIS-based surface water quality assessment and management approach is proposed in this study, where a fuzzy-set based water quality assessment model, eutrophication and heavy metal ecological risk assessment model, and an artificial neural network model for water quality prediction are built and integrated with the ArcGIS Engine. A system with a user-friendly interface is further developed that combines the water quality assessment and prediction models, the ArcGIS Engine, and a database for data storage and management. In order to verify the system developed, it was used to assess water quality conditions in the northern part of Liaodong Bay in China, and Lake Champlain, locates in the northern New York and Vermont and the Southeast Territories of Canada. For the water quality assessment modules, the model results are compared with previous studies in the same area and the standard water quality index method. The artificial neural network model is separately applied to predict the variation of chlorophyll-a concentration in Lake Champlain using Matlab. Sequentially, the algorithm of the BPANN model is represented using the C# programming language and integrated with the ArcGIS Engine and database. The ANN module developed is further applied to predict total phosphorus concentration in Lake Champlain. The prediction results are compared with in situ measurement data.

1.2 Research Objectives

The main objective of this thesis is to develop a geographical information technology based water quality assessment and management system. The system is capable of data processing, data management, fuzzy water quality assessment, eutrophication risk evaluation, heavy metal risk evaluation, water quality variable prediction, and the results are displayed in

the form of contour maps and color-coded maps with spatial information. Specifically, this thesis has the following objectives:

(1) To develop a GIS-based water quality assessment system that integrates the fuzzy synthetic assessment algorithm, the ArcGIS Engine and a database. Two case studies are conducted. One is a water quality assessment of the offshore area in the northern part of Liaodong Bay in China, the other one is a water quality assessment of Lake Champlain.
(2) To develop a back-propagation artificial neural network (BPANN) model for the prediction of chlorophyll a concentration using MATLAB.
(3) To represent the algorithm of the BPANN model in C# language and developed a GIS-based water quality parameter prediction system that integrates the algorithm of BPANN model, the ArcGIS Engine and a database. A case study is conducted for the simulation of total phosphorus concentration in Lake Champlain.
(4) To simulate total phosphorus concentrations in Lake Champlain using a MIKE 21 model and compare results with the BPANN model.

1.3 Thesis Organization

This thesis is organized in eight chapters:

Chapter 1 introduces the research background, states the research problems, specifies the research objectives, and briefly introduces the research methodologies.

Chapter 2 provides an extensive literature review regarding to surface water quality assessment and ecological risk assessment, artificial neural network model, physical surface water quality models, and the integration of Geographical Information System.

Chapter 3 describes the theories and methodologies about developing a GIS integrated surface water assessment and management system, including fuzzy-set based water quality and ecological risk assessment models (MWQ module and LWQ module) and artificial neural

network model for water quality prediction (ANN module). The assessment and prediction models are integrated with the ArcGIS Engine and a database. The scientific introduction of the MIKE 21 Flow Model with the flexible mesh approach is also illustrated.

Chapter 4 presents a case study of applying the MWQ module developed to assess the water quality and ecological risk in the northern part of Liaodong Bay in China. The results depict the spatial variation of water quality conditions and the main pollutants in the study area.

Chapter 5 depicts a case study of the assessment of water quality and ecological risk in Lake Champlain with the developed LWQ module. The temporal and spatial variation of the water quality in Lake Champlain, as well as the primary pollutant is analyzed.

Chapter 6 explores the efficiency of artificial neural network model in the prediction of water quality variables in Lake Champlain. Chlorophyll-a concentrations are predicted through the ANN model provided by Matlab software. Subsequently, the ANN module is developed in C# language and applied to predict the total phosphorus concentrations in Lake Champlain. The prediction results are verified with the lake water monitoring data in terms of R^2 and RSR. Moreover, the generated spatial distribution maps of the predicted results are generated and analyzed.

Chapter 7 provides a case study of simulating the total phosphorus in Lake Champlain using the hydrodynamic model and transport model of the MIKE 21 Flow Model FM. The simulated total phosphorus results of the MIKE 21 model, the predicted total phosphorus concentrations of the ANN model, and the measurement results of total phosphorus concentration in Lake Champlain are verified through regression analysis and error statistics.

Chapter 8 presents the conclusions and contributions as well as the recommendations for further research.

CHAPTER 2 LITERATURE REVIEW

2.1 Fuzzy Set Based Water Quality Assessment

Fuzzy-set based methods have been extensively used to assess surface water quality. To address the additional uncertainty that arises from missing data of hydrology variables, an imprecise fuzzy waste-load-allocation model was developed by Rehana and Mujumdar (2009) for river water quality management. Moreover, Zheng et al. (2007) applied the fuzzy synthetic evaluation method to assess seawater quality in a waste dumping area; the marine water quality level was determined according to the maximum membership principle. The maximum membership principle is disadvantaged as a result of information loss and may cause deflected results. Therefore, Liu et al. (2010) applied an improved fuzzy synthetic evaluation method to assess the water quality in the Three Gorges region; more credible results were obtained by replacing the maximum membership principle with the weighted average principle. Water quality is spatially distributed; however, the results of fuzzy set-based methods used in previous studies are simply numerical index values without visually organized output, which leads to an inadequate representation of the results of water quality assessment models. Meanwhile, the required computational effort for fuzzy-logic based water quality assessment is great, particularly when the amount of sampling data is large. Furthermore, there is a lack of knowledge regarding the interface system that integrates fuzzy-set based methods and the spatial output in the form of maps representing the degree of contamination.

2.2 Artificial Neural Networks

Several studies have recently been conducted on water quality forecast models (Wu and Xu, 2011, Zhao et al., 2012, Chibole, 2013). Traditional water quality modeling approaches, including parametric statistical and deterministic models (Ma et al., 2014), are capable of simulating the internal physical processes of the aquatic

system, but require extensive information that is not easily accessible, such as hydrologic time series, the geometry of rivers or lakes, and hydraulic coefficients that are applied in various hydrological sub-processes (Dogan et al., 2009). Moreover, many traditional water quality models are time-consuming and likely to introduce errors due to their assumption that the relation between the variables is linear (Singh et al., 2009). Since a large number of factors that affect the water quality have a complicated non-linear relationship, an artificial neural network (ANN) tries to simulate the learning processes of human being and has the ability to reflect the underlying linear or non-linear relationships amongst input and target data (Dogan et al., 2009). The advantage of an ANN approach over the traditional computing paradigm is that it avoids the need for a model with a physical basis and is able to rapidly map a given input into the desired output quantities, and the underlying processes causing the response do not need to be known explicitly (Bowden et al., 2006). Therefore, there has been a significant increase in their application in different scientific areas, such as modeling process, pattern recognition, and time series analysis (Li et al., 2007, Patil et al., 2008).

ANN systems have two main types of training process: supervised (e.g. multi-layer feed forward neural network) and unsupervised training (e.g. self-organized map neural network). In supervised training, the desired output is known and the weight coefficients are adjusted to make the calculated and desired outputs as close as possible. In unsupervised training, the desired output is not known and the system is left to itself to produce a stable statement based on a group of provided facts (Svozil et al., 1997). This thesis focuses on the supervised multi-layer feed forward neural network.

Maier and Dandy reviewed the use of ANN models to predict and forecast water resource variables (Maier and Dandy, 2000). The majority of those models used multi-layer feed forward neural networks. Most of these feed forward networks were

trained using the back-propagation algorithm, also referred as back-propagation artificial neural networks (BPANN). The BPANN model has been successfully applied to predict water temperature, salinity, DO, and Chl-a in Singapore coastal water. Results show the great potential of BPANN to simulate water quality variables (Palani et al., 2008). In Sahoo's study, a genetic algorithm-optimized BPANN model, multiple regression analysis, and chaotic non-linear dynamics algorithms were used to predict the stream water temperature and results indicated that the performance efficiency of the optimized BPANN was the highest among all considered algorithms (Sahoo et al., 2009). By comparing the results of the BPANN model and the MIKE 11 physically based hydrodynamic model in the simulation of water levels in a river, the results obtained from the BPANN model were much better than that of the MIKE 11, as indicated by the Nash-Sutcliffe index and root mean square error (Panda et al., 2010). The ANN approach with the vertical 2D and 3D hydrodynamic models were compared and the simulated results revealed that the vertical 2D and 3D hydrodynamic models could not capture the observed water stages during an input of high freshwater discharge from upstream boundaries, while the ANN could match the observed water stage; however, the ANN approach was inferior to the 2D and 3D models during the testing phase (Chen et al., 2012).

Besides numerous references to BPANN applications, methods for determining model inputs, dividing data sets, determining the best model structure, and the comparison between different training algorithms were also reviewed (Maier and Dandy, 2000, Maier et al., 2010, Piotrowski et al., 2014). The determination of the model input variables is one of the main tasks for a BPANN model. Variables that have a significant effect on the output variables should be selected to build a more compact network. Techniques for selecting input variables can be divided into two categories: model-free techniques and model-based techniques. Model-free techniques include analytical methods (linear correlation analysis and non-linear mutual information) and ad-hoc methods based on available data and a priori knowledge of

causal variables. Model-based techniques include stepwise methods like constructive, ad-hoc methods (for example the development of models with different input variables), sensitivity analysis, and global optimization methods like genetic algorithms (Maier et al., 2010). Model-based techniques have a disadvantage in that they are time-consuming, since several different BPANN models need to be built and tested in order to find the optimized input variables.

2.3 Surface Water Quality Models

Water quality models can be effective tools to simulate and predict pollutant transport in a water environment. Many hydrodynamic models and water quality models have been studied and applied in the field of surface water quality modeling.

The Environmental Fluid Dynamics Code (EFDC) is a general purpose surface water modeling system that can be used to simulate aquatic systems in one, two, and three dimensions, which includes hydrodynamics, sediment transport, toxic contaminant transport, and water quality-eutrophication components (Park et al., 2000). EFDC uses stretched or sigma vertical coordinates and Cartesian or curvilinear, orthogonal horizontal coordinates to represent the physical characteristics of a waterbody. EFDC approximates the horizontal plane domain with a set of discrete quadrilateral and optional triangular cells. The hydrodynamic equations are solved in a horizontal curvilinear and orthogonal coordinate system (Tetra Tech, 2002). The user interface of EFDC is based on standard text input file templates, which allows the user to modify input files using text editing software and modify files on remote computing systems and in heterogeneous network environments (Park et al., 2000). It has been applied to various water bodies including rivers, lakes, reservoirs, wetlands, estuaries, and coastal regions in environmental assessment and management processes (Wool et al., 2003, Park et al., 2005, Çalışkan and Elçi, 2009, Jeong et al., 2010, Wang et al., 2013, Wu et al., 2014).

The Water Quality Analysis Simulation Program (WASP) is a dynamic compartment-modeling system for 1, 2, and 3 dimensional aquatic systems. The time-varying processes of advection, dispersion, point and diffuse mass loading and boundary exchange are represented in the WASP model and it has been widely applied in the development of Total Maximum Daily Loads (TMDL) (United States Environmental Protection Agency, 2013). WASP contains a user-friendly Windows-based interface, a pre-processor for the preparation of data to be used in WASP, eutrophication and organic chemical model processors, and a graphical post-processor for viewing and comparison of results. The horizontal domain is divided into several rectangular segments in which WASP solves the equations. For each segment, the user should define the volume, velocity, depth, environmental parameters, initial concentrations, and fraction dissolved. The fluxes of flow and environmental parameters between each segment are necessary parameters in WASP modeling. Thus, the user needs to have a good understanding of how the water body will be segmented and a good knowledge of flow and pollutant transportation in the study area (Wool et al., 2007). The WASP model has been extensively applied in the water quality modeling of rivers, reservoirs, and lakes with the combination of other hydrodynamic models and software, such as EFDC, MIKE 21, and GIS (Tufford and McKellar, 1999, Zhang et al., 2008, Zhang et al., 2008, Peng et al., 2010, Seo et al., 2012, Kannel and Gan, 2013, Tang et al., 2014, VishnuRadhan et al., 2014).

MIKE 21 is a two-dimensional free-surface flow modeling system developed by DHI water & Environment, which is applicable in the simulation of hydraulic and environmental phenomena in lakes, estuaries, bays, coastal areas and seas where stratification can be neglected (DHI Software, 2014). The MIKE modeling system has been widely used in the hydrodynamic simulation, water quality simulation and planning, and eutrophication problem management. The dissolved oxygen distribution in a shallow that presents a complex network structure was studied using the MIKE 21 modeling system; results indicate that the ocean is the main source of oxygen and

the tidal currents are the main factor controlling the oxygen distribution in the main lagoon areas; the water quality model results indicate that the primary production and phytoplankton respiration are the main factors affecting the oxygen budget in the water column; the model can be used as a tool for the study and management of the lagoon system (Lopes et al., 2008). The differences between data-driven models and deterministic models were discussed, and the functions of NAM, MIKE11 HD and MIKE3 FM models were integrated to provide forecast of biological parameters(Bedri et al., 2014). The NAM, MIKE11 HD and water quality models were used to simulate the rainfall-runoff, hydrodynamic processes and oxygen cycle of River Sosiani at the catchment scale and the model efficiency was satisfactory on the basis of the comparison of simulated data and observed data (Chibole, 2013). The MIKE 11 modeling system was applied to the Strymonas River catchment between Bulgaria and Greece for the simulation of surface water management practices (Doulgeris et al., 2012). The hydrodynamic model (MIKE 21 FM) coupled with the water quality model (Ecolab) is established to simulate the current situation of Erhai Lake; the model parameters are calibrated through comparison with observed data and the simulation results are verified to obtain the level of water quality in the lake (Zhu et al., 2013).

The advantage of MIKE models over the EFDC and WASP models is that the MIKE flow model uses the flexible mesh approach in generating the grids. The element size of the mesh varied throughout the model domain and along the boundary of the domain, thus model resolution can be varied across the model area. This unstructured mesh technique allows for the control of node distribution for optimal usage of nodes, the adoption of mesh resolution to the relevant physical scales, and the generation of boundary-fitted mesh (MIKE by DHI, 2013). Furthermore, the powerful visualization functions integrated within the MIKE model are capable of displaying color-coded spatial distribution maps of model results without requiring any extension modules.

2.4 Eutrophication Risk and Heavy Metal Risk Assessment

Eutrophication is a crucial phenomenon to be included in surface water quality assessment. Eutrophication in the surface water environment is a complex problem with adverse effects on the health of the ecosystem and presents one of the major stresses to the surface water environment (Meyer-Reil and Köster, 2000). Great efforts have been made to establish criteria to provide a single quantitative index for the purpose of classifying and ranking lakes, mostly based on physicochemical (e.g., Chl a, nutrients, dissolved oxygen, chemical oxygen demand) and biological indicators (e.g., macro algal abundance, the changes in the distribution of submerged aquatic vegetation, and the increasing occurrence of harmful algal blooms) (Herrera-Silveira and Morales-Ojeda, 2009; Pu et al., 2012; USEPA, 2008).

Various sets of criteria have been reported in the literature for eutrophication risk assessment. The specification for offshore environmental monitoring of China defined the eutrophication index based on three water column parameters: chemical oxygen demand (COD), inorganic nitrogen and phosphate (HJ-442, 2008). Carlson developed 3 Trophic State Indices (TSI_{SD}, TSI_{TP}, TSI_{CHL}) by establishing the relationship between transparency (Secchi disk), total phosphorus and chlorophyll a (Carlson, 1977). Canadian and OECD trophic status criteria use the same variables as the Carlson TSI (Environment Canada, 2004). Besides these three parameters, total nitrogen is also considered in Swedish criteria and Nurnberg criteria (Galvez-Cloutier and Sanchez, 2007). All these trophic status indices have different threshold values for each parameter corresponding to 4 to 6 trigger ranges representing the trophic status.

Previous studies regarding heavy metal ecological risk have primarily focused on the sediment environment, corresponding to the region that has been found to contain the highest heavy metal concentrations (Bastami et al., 2014; Liu et al., 2014; Wu et al., 2014; Yi et al., 2011). Although water bodies tend to have the lowest heavy metal

concentrations in the aquatic system, the accumulation of heavy metals in fish and other aquatic animals results primarily from surface contact with the water and via the food chain. The uptake amount through these routes depends on the environmental levels of heavy metals in the fish habitat. Thus, heavy metal in the water may cause adverse effects on fisheries. The potential ecological risk index developed by Hakanson (Hakanson, 1980) is a comprehensive method that has been widely used to evaluate the combined heavy metal risk in aquatic systems (Bastami et al., 2014; Pekey et al., 2004; Zhang et al., 2013).

2.5 Integration with Geographical Information System

The GIS makes a significant contribution to water quality risk assessment and environmental management systems in virtue of its intrinsic ability to analyze and display large amounts of spatial data. Its powerful visualization of data and analysis results are useful in decision-making processes (Debaine and Robin, 2012). Integration of GIS and water quality assessment and prediction models combines their functions and advantages and could have several benefits for the environmental assessment and management process. Strategies for integrating models with GIS have been discussed and analyzed by many researchers (Huang and Jiang, 2002, McKinney and Cai, 2002, Halls, 2003, Vairavamoorthy et al., 2007, Jiang et al., 2012, Sharma et al., 2013). Two extreme types of coupling--loose and fully tight--were emphasized. Loose coupling consists of data file transferring between GIS and an external modelling program. It enables the exploitation of the cartographic potential of GIS and provides model developers with the required flexibility in choosing the most appropriate models to integrate and the user interface to build. The main drawback is that the loose coupling requires the development of further tools to format data files properly (Santini et al., 2010). With tight coupling, the models are developed entirely within a GIS environment using a macro language such as AML (ESRI Arc Macro Language) and analytical functionalities. However, this type of programming needs

highly technical computer programing skills and is often not able to include complex applications (Al-Sabhan et al., 2003). Besides the two extreme loose and tight coupling, some intermediate types were also applied and analyzed. Brandmeyer and Karimi (2000) arranged the categories into a five layer pyramid describing the increasing level of integration from one-way data transfer to tool coupling, as shown in Figure 2-1. Tool coupling is at the top of the coupling hierarchy, where models are coupled using an overall modeling framework. The framework presents a single graphical user interface and implements shared data storage, also providing functions and tools common to multiple models, while managing data and computing resources.

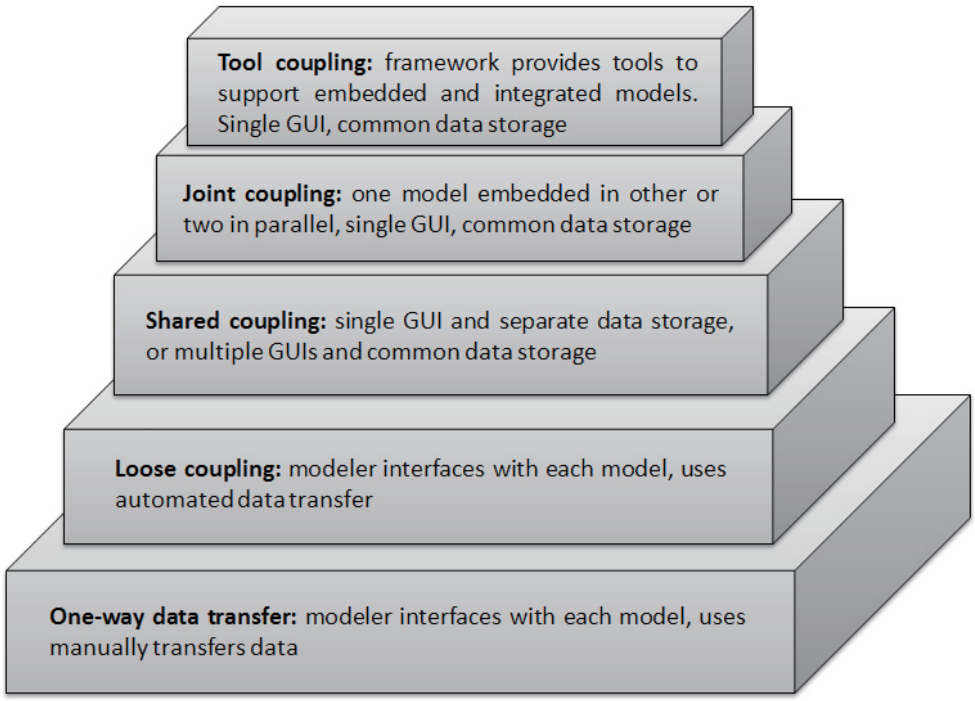


Figure 2-1 Methodologies for coupling environmental models (Brandmeyer and Karimi, 2000)

2.6 Summary

In summary, the results of the fuzzy synthetic assessment model, ecological risk assessment model and the artificial neural network model are just numerical values

without any spatial information, which may not be sufficient for the water quality managers. There has been a lack of research regarding the integration of these models and GIS that is able to facilitate the model calculation and provide spatial reference for the model results, as well as the application of artificial neural network model for the area with complex hydraulic and geographic conditions.

CHAPTER 3 METHODOLOGY

3.1 Overview of the System Developed in This Thesis

The primary components and working process of the developed system named as GIS-SWQAM, MIKE 21 model and other components of the methods included in this thesis are shown in Figure 3-1.

The primary methods applied in this thesis include:

✓ The development of GIS-SWQAM, including:

- Fuzzy-set based water quality assessment, which utilizes the fuzzy synthetic assessment method to determine the water quality status in the study area;
- Eutrophication risk assessment and heavy metal risk assessment to identify the eutrophication conditions and heavy metal risk levels in the study area;
- Back propagation artificial neural network (BPANN) modeling to perform the simulation and prediction of water quality variables;
- Database connection with the above models and ArcGIS Engine. In consequence, the database is capable of providing data storage and management for the above models at the backstage of the system and serves as a data source for ArcGIS Engine;
- ArcGIS Engine integration with the system interface and database, which has the function of generating visually displayed maps of water quality assessment model results and BPANN model results;

✓ Besides the system development that integrates the above methods, the field validation of the commercial MIKE 21 model is carried out in the same study area as BPANN models and the results are compared according to the model performances.

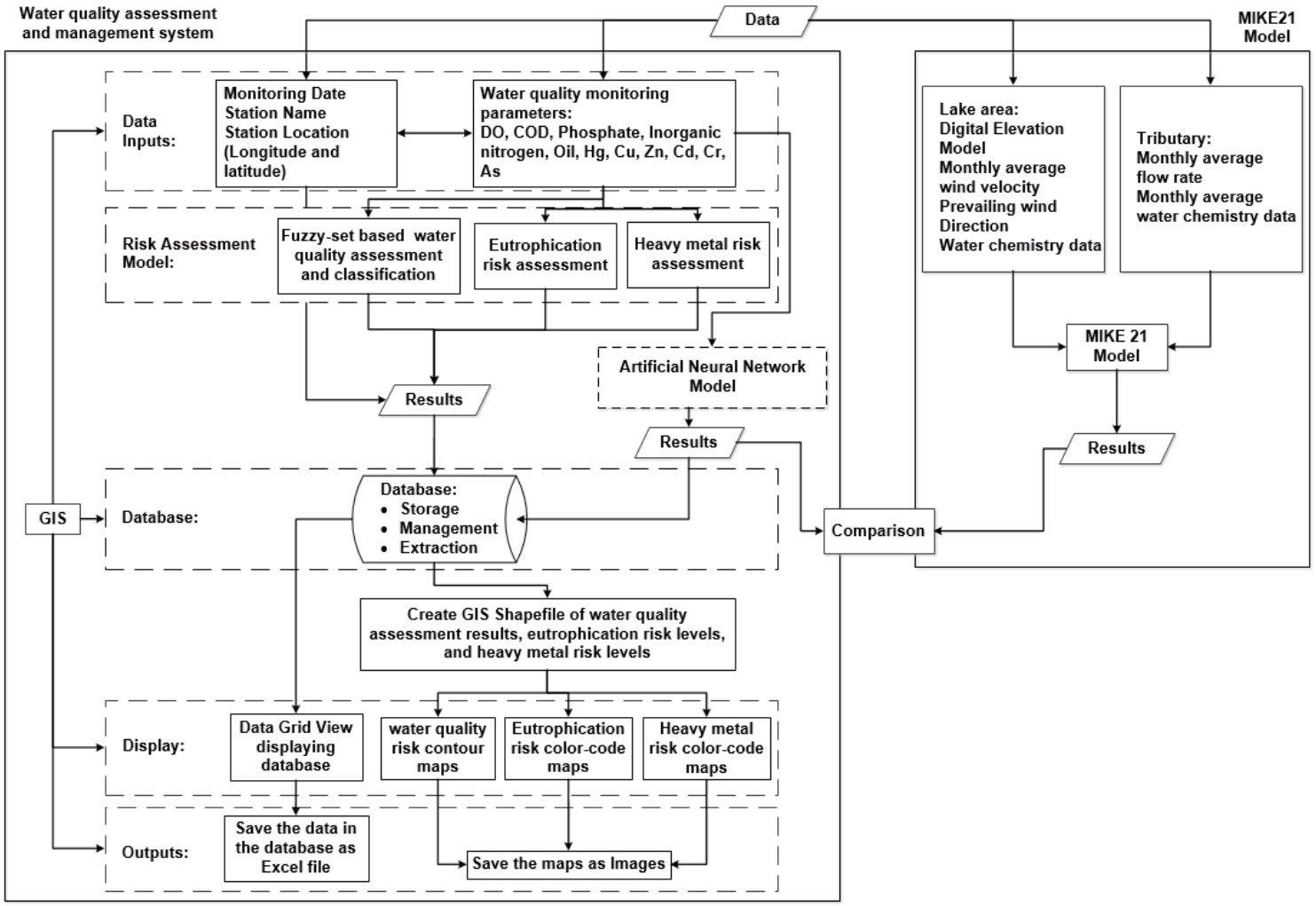


Figure 3-1 Overview of the developed system (GIS-SWQAM) and other methods applied in this thesis

3.2 Fuzzy-Set Based Water Quality Assessment

As one of the fuzzy-set based methods, the fuzzy synthetic assessment method is selected to assess the surface water quality in the system. The fuzzy synthetic assessment consists of four parts: calculation of fuzzy membership matrix; determination of the weight set; final evaluation result determination and water quality classification.

The fuzzy synthetic assessment method for assessing water quality is developed through the following seven steps:

(1) Select water quality parameters used in the assessment.

(2) Establish assessment criteria.

Local water quality standards are used as the assessment criteria of each indicator in this study.

(3) Establish membership functions based on the marine water quality assessment criteria.

In fuzzy logic, the membership functions represent the degree of the specified value belongs to the set. In this study, on the basis of water quality assessment criteria, the membership functions of each assessment factor to assessment criteria at each level can be described by a set of formulae as follows (Eqs.(3.1) - (3.3)) (Icaga, 2007).

$$r_{ij}(x_i) = \begin{cases} 1 & x_i < S_{ij} \\ (S_{i(j+1)} - x_i)/(S_{i(j+1)} - S_{ij}) & S_{ij} \le x_i \le S_{i(j+1)} \\ 0 & otherwise \end{cases} \quad (3.1)$$

$$r_{ij}(x_i)=\begin{cases}(x_i-S_{i(j-1)})/(S_{ij}-S_{i(j-1)}) & S_{i(j-1)}\le x_i\le S_{ij}\\ (S_{i(j+1)}-x_i)/(S_{ij}-S_{ij}) & S_{ij}<x_i\le S_{i(j+1)}\\ 0 & otherwise\end{cases} \tag{3.2}$$

$$r_{ij}(x_i)=\begin{cases}0 & x_i<S_{i(j-1)}\\ (x_i-S_{i(j-1)})/(S_{ij}-S_{i(j-1)}) & S_{i(j-1)}\le x_i\le S_{ij}\\ 1 & x_i>S_{ij}\end{cases} \tag{3.3}$$

where $r(x)$ is the membership function, i is the number of assessment factors, j is the number of assessment criteria levels, x_i is the analyzed value of assessment factor, $S_{i(j+1)}$, S_{ij}, $S_{i(j-1)}$ is the assessment threshold of the ith assessment factor at level $j+1$, j and $j-1$, respectively.

(4) Calculate the fuzzy relationship matrix.

Substituting the data of each assessment factor and the threshold of criteria into the membership functions above, the fuzzy matrix R can be expressed as Eq. (3.4):

$$R=\begin{bmatrix}r_{11} & r_{12} & \cdots & r_{1m}\\ r_{21} & r_{22} & \cdots & r_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ r_{n1} & r_{n2} & \cdots & r_{nm}\end{bmatrix} \tag{3.4}$$

where r_{ij} is the membership degree of the ith assessment factor to the assessment criteria at level j.

(5) Determine the weight matrix.

In the fuzzy synthetic assessment system, every element may play a different role if its weighting coefficient is different. The environmental quality index is used to determine the weight of the factor and then the weights are normalized. This method

not only can highlight the role of the main factors in environmental assessment, also take the difference between the standard values of factors into consideration (Zheng et al., 2007). The weight set is $W = \{w_{01}, w_{02}, \cdots, w_{0n}\}$.

$$w_{oi} = (x_i / S_{oi}) \Big/ \sum_{i=1}^{n} (x_i / S_{oi}) \tag{3.5}$$

where $w_{oi}, x_i, and\ S_{oi}$ are the weight after normalization, the actual value of assessment factor and the average value of all the assessment thresholds of factor *i*, respectively.

(6) Calculate the fuzzy synthetic assessment vector.

$$b_j = \sum_{i=1}^{P} (w_{oi} \times r_{ij}) \tag{3.6}$$

where b_j is the weight of the value of *i*th assessment factor at level *j*, r_{ij} is the membership degree of the *i*th assessment factor to the assessment criteria at level *j*.

The vector then can be expressed as:

$$B = W * R = \begin{bmatrix} w_{01} & w_{02} & \cdots & w_{0n} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \tag{3.7}$$

(7) Determine the water quality category.

The maximum membership principle is an analysis method widely used in environmental assessment. However, this principle chooses the maximum value of the factors in the vector B, and some information may be lost during the process. In view of this issue, the weighted average principle is used in this study to improve over the

maximum membership principle for preserving the information in the assessment coefficients as much as possible (Liu et al., 2010).

$$B^* = \frac{\sum_{j=1}^{m} b_j^k \times j}{\sum_{j=1}^{m} b_j^k} \tag{3.8}$$

where k is a coefficient to control the role of maximum value of, usually k=1 or 2, in this article k=1, B* is the fuzzy marine water quality index. The water quality classification is defined according to local regulations of the study area. The water quality category is the membership class to which the B* value corresponds.

3.3 Ecological Risk Assessment

3.3.1 Eutrophication risk level assessment

Eutrophication is a growing problem in surface water environment in many parts of the world. Meanwhile, the eutrophication status is one of the key factors in the comprehensive risk assessment of environmental sensitive areas like the mariculture zones, the eutrophication state index was included in this study to present the eutrophication risk. Since two case studies from different countries were conducted in this thesis, two assessment indices were considered herein according to local standards.

According to the specification for offshore environmental monitoring of China (HJ-442, 2008), three water column parameters are included in the evaluation of eutrophication for marine water: chemical oxygen demand (COD), inorganic nitrogen and phosphate. The Eutrophication index can be calculated as:

$$E = \frac{C_{COD} \times C_{DIN} \times C_{DIP}}{4500} \tag{3.9}$$

where E is the eutrophication index, $C_{COD}, C_{DIN}, and\ C_{DIP}$ are the concentration of chemical oxygen demand (mg/L), dissolved inorganic nitrogen (μg /L), and dissolved inorganic phosphate (μg /L), respectively.

Tabel 3-1 indicates the eutrophication levels corresponding to eutrophication index intervals in China:

Table 3-1 Eutrophication risk levels in China corresponding to eutrophication index intervals

Eutrophication Index	Eutrophication risk level	Color
E < 1.0	Oligotrophic	
1.0 ≤ E < 2.0	Light eutrophic	
2.0 ≤ E < 5.0	Mesotrophic	
5.0 ≤ E < 15.0	High eutrophic	
E ≥ 15.0	Severe eutrophic	

Carlson TSI is recommended by the United States Environmental Protection Agency (U.S. EPA) to evaluate the trophic status in surface water. In this method, three trophic state indices were calculated as follows (Carlson, 1977):

$$TSI(\mathrm{SD}) = 10(6 - \frac{\ln SD}{\ln 2}) \tag{3.10}$$

$$TSI(\mathrm{Chl}) = 10(6 - \frac{2.04 - 0.68 \ln Chl}{\ln 2}) \tag{3.11}$$

$$TSI(\mathrm{TP}) = 10(6 - \frac{\ln \frac{48}{TP}}{\ln 2}) \tag{3.12}$$

It might be redundant to obtain three indices for the same purpose, thus the above three indices were averaged to compute a TSI_{AVG}. By taking the average, the physical response (SD), the biological response (Chl a) and the nutrient (TP) were combined into one index, which is simpler for comparative and management purposes (Kratzer and Brezonik, 1981, Galvez-Cloutier and Sanchez, 2007).

According to Carlson TSI, the TSI_{AVG} ranges along a scale from 0-100. Table 3-2 shows a summary of trigger ranges of each parameter and the corresponding trophic status for lakes.

Table 3-2 Trophic status classification

TSI_{AVG}	Classification	Color
<40	Oligotrophic	
40-50	Mesotrophic	
50-70	Eutrophic	
>70	Hypereutrophic	

3.3.2 Heavy metal risk level assessment

Eight Parameters were taken into consideration in Hakanson's study, including PCB, Hg, Cd, As, Pb, Cu, Cr and Zn. to the aquatic system (Pu et al., 2012).

$$RI = \sum_{i=1}^{n} T_r^i C_s^i / C_n^i \tag{3.13}$$

where T_r^i is the toxic-response factor for a given substance, C_s^i represents a metal's concentration in the water samples. C_n^i values are the reference concentration of heavy metals to which the risk is considered.

The toxic-response factor for the seven selected heavy metals are listed in Table 3-3 (Hakanson, 1980).

Table 3-3 Toxic-response factor for selected heavy metals

Heavy metal	Hg	Cd	As	Pb	Cu	Cr	Zn
T_r^i	40	30	10	5	5	2	1

3.4 Back-Propagation Artificial Neural Network (BPANN)

An artificial neural network is an artificial representation of the behavior of a human brain emulating the learning process of biological neurons (Dogan et al., 2009). ANNs have the ability to reflect the complex linear or non-linear relationships amongst input and target data through training and calibration, so that the ANN's output becomes as near as possible to target output. A large amount of historical data is needed in the training part of an ANN model. On the basis of satisfactory training, the ANN model should be able to produce output for a new data set that has never been used in the training process (Palani et al., 2008).

The most commonly used ANN is the three-layer feed-forward model, which is comprised of three distinctive layers: the input layer, the hidden layer, and the output layer. ANN architectures with one hidden layer is applied here, since a neural network with one hidden layer is recognized to be capable of approximating any finite non-linear function with high accuracy (He et al., 2011).

3.4.1 Back propagation neural network and learning algorithm

Back-propagation (BP) is a commonly used learning algorithm in ANN applications, which uses the back-propagation algorithm as the gradient descent technique to minimize network error. Each layer in the BPANN has several neurons and each neuron transmits input values and processes to the next layer. Initially, each input is weighted with a small arbitrary value. The sum of the weighted inputs and the bias are used as the input of the transfer function of the hidden layer. The results of the hidden layer will subsequently be used as the input signals of the output layer and process through the transfer function. As the learning progresses, the weights are adjusted in response to the training data and the output values. An output will be predicted after the information has gone through the network in a forward direction. Since all the weights are random, the output might be different from the target. The

error between the target and the predicted output is calculated at the output layer and then redistributed back through the model using the back propagation algorithm. The weights are adjusted accordingly until the error meets the pre-specified goal or the pre-determined maximum epoch is achieved through several iterations (Sahoo and Ray, 2006, Li et al., 2012).

As can be seen in Figure 3-2, the input variables are multiplied by the connection weights w'_{ij} between the input and hidden layer. The weighted signals and bias from the input neurons are summed by the hidden neurons and then projected through a transfer function f_h. The results of the function f_h are weighted by the connection weights w''_j between the hidden and output neurons and sent to the output nodes. The transfer function f_o is then projected by the output neurons. The output of this neuron is the predicted response $\hat{y}$ (Dieterle, 2003).

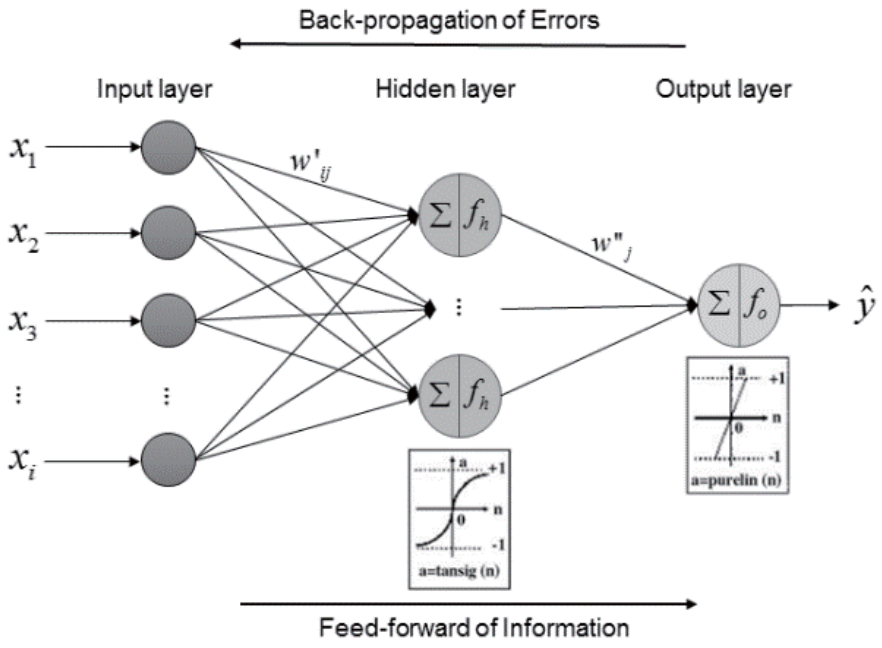


Figure 3-2 Architecture of the developed ANN model in this thesis

In this study, a hyperbolic tangent sigmoid transfer function (Eq. (3.14)) is used to transfer the values from the input neurons to the hidden neurons (f_h), and a linear

transfer function (Eq. (3.15)) is adopted to transfer the values from the hidden neurons to the output neurons (f_o) (Sedki et al., 2009).

$$X_j = \tanh(\sum_{i=1}^{n} x_i w_{ij}' + b_{jh}) \tag{3.14}$$

$$\hat{y} = \sum_{j=1}^{m} X_j w_j'' + b_{ko} \tag{3.15}$$

where i, j and k are the number of neurons in the input, hidden, and output layers, respectively; x_i is the value of the input variable, X_j is the result obtained by the active function from the input neurons, w'_{ij} is the connection weight between the input and hidden neuron, and w''_j is the connection weight between the hidden neuron and output neuron, b_{jh} and b_{ko} are the bias for the jth and kth neuron, respectively.

Specifically for the development of the BPANN model in this thesis, x_i (i=1, 2, 3,…, n) in Figure 3-2 represent the data of water quality parameters that are used as the input for the BPANN model. $\hat{y}$ represents the desired model output. For example, in the case of predicting chlorophyll a concentration, x_1 represents dissolved oxygen concentration, x_2 is pH values, x_3 stands for water temperature, and $\hat{y}$ represents chlorophyll-a concentration; in the case of predicting total phosphorus concentration, x_1 represents dissolved oxygen concentration, x_2 is total nitrogen concentration, x_3 stands for total phosphorus concentration in inflow tributaries, and $\hat{y}$ represents total phosphorus concentration in the lake water.

The constructed model is trained using the Levenberg-Marquardt algorithm (LMA) in this study, since the LMA is faster than other back-propagation algorithms that are used to update weights and bias (MathWorks-trainlm, 2014). The application

of the BPANN model in this study aims at training the most generalized neural network rather than the one that optimally fitted the training set.

3.4.2 Input variables and data processing

The techniques for selecting input variables can be classified into model-free techniques and model-based techniques (Maier et al., 2010). These techniques can be classified into model-free techniques and model-based techniques. Model-free techniques include analytical methods based on linear correlation and non-linear mutual information and ad-hoc methods on the basis of available data and domain knowledge. Model-based techniques include stepwise methods, sensitive analysis, global optimization methods, and ad-hoc methods. Among these methods, the most commonly used technique is linear correlation, which is used in conjunction with the consideration of domain knowledge to select the alternative input for the BPANN model. Models with different combination of alternative input variables are further carried out to determine the optimum input of the BPANN model that has the best performance.

The complete lake water quality data are divided into three sub-sets: the training, validation, and testing data sets. The training data set is used for computing the gradient and adjusting the weights and biases of the network. The validation data set is used to decide when to stop training in order to prevent over-fitting problems during network training. The testing data set is a new data set that has never been used during the training process, and it is used to assess the generalization ability of the trained model and to predict the selected target outputs using the trained network (MathWorks-data division, 2014).

Since different variables have different ranges, the variables should be scaled to uniform ranges that are coincident with the limits of the activation functions used in

the output layer (Maier and Dandy, 2000). The normalization could be done using the following equation (Liu and Chen, 2012):

$$Y_N = (y_{\max} - y_{\min})(\frac{x_i - x_{\min}}{x_{\max} - x_{\min}}) - y_{\min} \quad (3.16)$$

Y_N denotes the data value after normalization, x_{max} and x_{min} are the maximum and the minimum values of data, respectively, and y_{max} and y_{min} denote the boundary values of the specific range determined by the user, which are taken as 1 and -1 respectively in this study.

3.4.3 Optimization of the BPANN structure

The optimal structure of the BPANN model is determined based on the minimum value of the mean square error (MSE) of the training and validation data sets. MSE is used as a target error goal in network training. The training will stop when it reaches the setting MSE or the maximum epoch. The MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (3.17)$$

$\hat{y}_i$ and y_i denote the estimated response and the target output of the BPANN model, respectively.

The neuron numbers in the hidden layer (N_h) play important roles in the optimization of the model structure. This is because a network with few hidden neurons may not be sufficient to perform a satisfactory training, while too many hidden neurons will lead to over-fitting. Palani et al. (2008) supported that N_h can lie between I and 2I+ 1 and it should be greater than the maximum of I/3 and O. Within this range, the optimum hidden neuron number is determined through a trial and error approach. Networks with few hidden nodes are preferable to that with many hidden

nodes, because the networks with fewer hidden neurons usually have better generalization capabilities and fewer over-fitting problems. The one that has the minimum mean square error in training and validation data sets is chosen as the optimum network structure.

3.4.4 Model performance evaluation

To determine the performance of the trained model, the estimated responses and the target outputs are compared. Both graphical techniques and quantitative statistics are used in model evaluation herein.

For graphical techniques, the standard regression line and the response curves are applied to indicate the match degree between estimated responses and target outputs. The slope represents the relative relationship between estimated responses and target outputs. The y-intercept of the regression line indicates the lag between observed data and model predictions (Moriasi et al., 2007).

For quantitative statistical methods, two different criteria are used: coefficient of determination (R^2), and RMSE-observations standard deviation ratio (RSR). In the following equations, $\hat{y}_i$ and y_i denote the estimated response and the target output of the BPANN model, respectively; y_{mean} is the mean value of the observed data; n is the total number of observations in each data set.

The coefficient of determination (R^2) is used to evaluate the model performance, which represents the percentage of variability that can be explained by the model. It ranges from 0 to 1 and higher values indicate better model performance. Typically, R^2 values greater than 0.5 are considered acceptable (Singh et al., 2012). It is calculated as:

$$R^2 = \left[\frac{n\sum_{i=1}^{n} y_i \hat{y}_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} \hat{y}_i)}{\sqrt{\left[n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2 \right] \times \left[n\sum_{i=1}^{n} \hat{y}_i^2 - (\sum_{i=1}^{n} \hat{y}_i)^2 \right]}} \right]^2 \quad (3.18)$$

RSR is calculated as the ratio of the RMSE and standard deviation of observed data. The RMSE is a measure of the goodness-of-fit and describes an average measure of the error in prediction. It can be computed as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (3.19)$$

The RSR is calculated as shown in equation (3.20):

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}}{\sqrt{\sum_{i=1}^{n} \frac{(y_i - y_{mean})^2}{n}}} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - y_{mean})^2}} \quad (3.20)$$

RSR incorporates the benefits of error index statistics and includes a normalization factor. RSR ranges from zero, which indicates perfect model simulation, to any larger positive values. A lower RSR indicates a lower RMSE and better model performance. The model performance can be judged as satisfactory if RSR is less than 0.70 (Moriasi et al., 2007).

3.5 Water Quality Assessment and Management System Development

The GIS-based water quality assessment and management system is developed using the C sharp programing language and is integrated with the fuzzy synthetic evaluation model, the eutrophication risk and heavy metal risk assessment modules, the BPANN model for water quality variable prediction, ArcGIS Engine 9.3, and the Microsoft Access database.

The system contains three submodules: the marine water quality assessment module (MWQ), the lake water quality assessment module (LWQ), and the artificial neural network module (ANN). Each module can be accessed through the corresponding buttons on the welcome interface of the system (Figure 3-3). The primary components of the system developed support the following functions: survey information input, the risk assessment model (including fuzzy synthetic water quality assessment, eutrophication risk and heavy metal risk assessment), the back propagation artificial neural network model, data storage and management, the production of risk level maps and the distribution maps of water quality variables with retrieved data, the display of the database and the produced maps, and the export of the results. The implementation of the developed system will be illustrated later in this chapter.

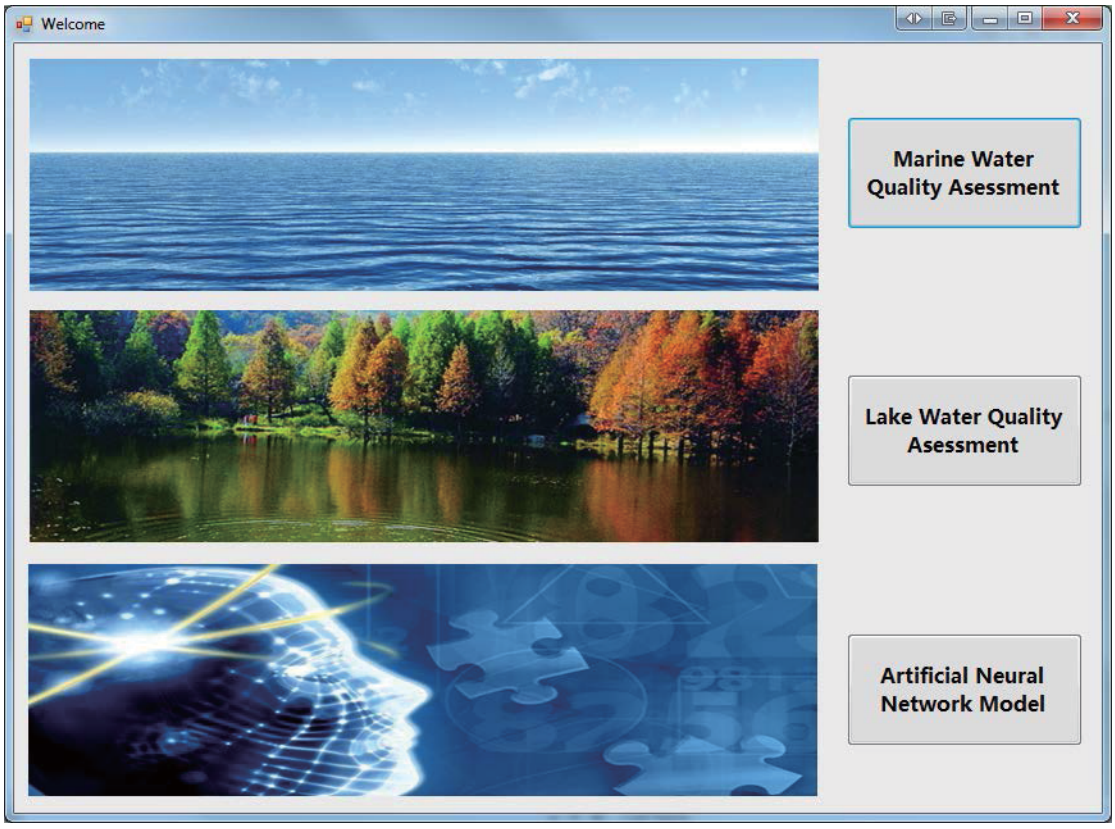


Figure 3-3 Welcome interface of the developed system

3.5.1 Data processing and management

The database is designed to achieve three main functions: data acquisition, data processing, and data management. Data acquisitions are achieved by capturing the input of the user through the interface. The acquired data are processed through the risk assessment model and back propagation artificial neural network model. The input data and model results are subsequently stored in the database and displayed in the data table tab on the user interface.

Database management is conducted using the Microsoft Access software. Some basic management options are linked to the interface by considering the convenience, including the sorting of columns and the deletion of selected rows. To avoid unintentional data alteration and deletion, the modification of data and the direct removal of rows or columns in the data table are unavailable. The data are exported to an Excel file for further reorganization and calculations.

Data in the database are grouped by survey date, which can be extracted from the corresponding column. All of the survey dates and water quality parameters in the input module are added to the drop-down menus in the “Create Maps” module that is described in the following section. Using this menu, users can obtain the resulting contour maps for different water quality parameters in each investigation and produce color-coded maps for eutrophication risk and heavy metal risk.

3.5.2 Development of a GIS-based user interface system

The developed system facilitates calculations of the fuzzy synthetic evaluation module and ecological risk assessment module, and the integration of an artificial neural network (ANN) model into the system enables the function of water quality prediction. The results are intuitively demonstrated trough integration with GIS-based visualization. By integrating the ArcGIS Engine 9.3 into the system, GIS is used as a

post-processor of the risk assessment model and BPANN model. According to the methodologies for coupling environmental models illustrated in Section 2.5, the joint coupling method is applied in this study, since it avoids the need for high-level programming skills and is able to achieve the desired functions. The joint coupling method includes two parallel models or one model embedded in the other. The models have a single graphical user interface and common data storage. The ArcGIS Engine component is linked to the database, which is the data source for creating risk level maps. This component processes the retrieved data, and subsequently creates and displays risk maps and distribution maps. Moreover, this component also contains some common GIS functions that are listed in the toolbox at the top of the map window.

Integration of the ANN model with the system developed in the .NET environment can combine the advantages of both modules. Many methods are able to achieve the integration of an ANN model and the system developed in C# programming language. MATLAB software provides many related functions for developing an ANN model. After the model is trained and validated, a dynamic link library (DLL) for interfacing with other applications can be created from the programing code of the trained model through the MATLAB Compiler (MathWorks). This method needs less effort in coding, however, its application is not flexible, because the model is developed for predetermined purposes and after being processed through the compiler, the model parameters (e.g., layers and neuron numbers) cannot be modified, thus the DLL can only be used for the specific purposes. The other more flexible method is to compile the ANN algorithm using C#, including the data processing, the active functions, finding the weights and bias, and the error calculation, etc. This method requires a good knowledge of each step in the ANN algorithm and a large quantity of code to present the algorithm. Many frameworks for scientific computing in .NET have been developed in recent years, such as the AForge.NET framework (Kirillov, 2013), the Accord.NET framework (Open HUB)

and the Encog machine learning framework (Heaton Research, 2010, Heaton Research, 2014). These frameworks provide support for neural networks, image processing, machine learning and many other advanced algorithms. The ANN algorithms can be called by using less code in development, thus an ANN model with dynamic structures and parameters can be built in .NET environment.

Encog is an Artificial Intelligence (AI) Framework for Java and .Net that provides the tools to create many different neural network types, including feed-forward, recurrent, self-organizing maps, radial basis function and Hopfield neural networks (Heaton Research, 2010, Heaton Research, 2014). Encog 2.4 framework is used in this thesis for the development of back propagation artificial neural network algorithm in the system.

The interfaces of the three submodules of the developed system are shown in Figure 3-4 to Figure 3-11, respectively.

3.5.2.1 Implementation of MWQ module

The implementation of the marine water quality assessment module of the system can be achieved as follows (Figure 3-4):

[I] Enter investigation data, including the information of the monitoring station, the investigation time and observed values of parameters involved in the assessment, in the “Parameter Input” module.

[II] Click the “Run FSE Model”, “Calculate Eutrophication Index” and “Calculate Heavy Metal Risk Index” buttons to obtain the fuzzy synthetic evaluation result and the water quality classification, the eutrophication index and eutrophication risk level, and the heavy metal risk index and heavy metal risk level, respectively.

[III] All input data and results are saved in a database, which can be displayed using the data table tab in the interface (Figure 3-5).

[IV] Refresh the data source, select the date of investigation and parameter and set the interval of the contours in the map. By clicking the “Create Contour Map” button, the contour map is automatically produced and added to the base map. The user can choose to display a contour map or a filled contour map by clicking the checkbox in front of the layer. The contour map includes contour lines with the corresponding fuzzy water quality index labels adjacent to each line. The filled contour map uses black, white and shades of grey to depict the corresponding fuzzy water quality index, with a legend shows the corresponding colors. Then, click the “Create Eutrophication Risk Level Map” and “Create Heavy Metal Risk Level Map” buttons to obtain the respective color-coded maps of risk level (Figure 3-4). The legend for these color-coded maps is located in the legend tab in the interface (Figure 3-5).

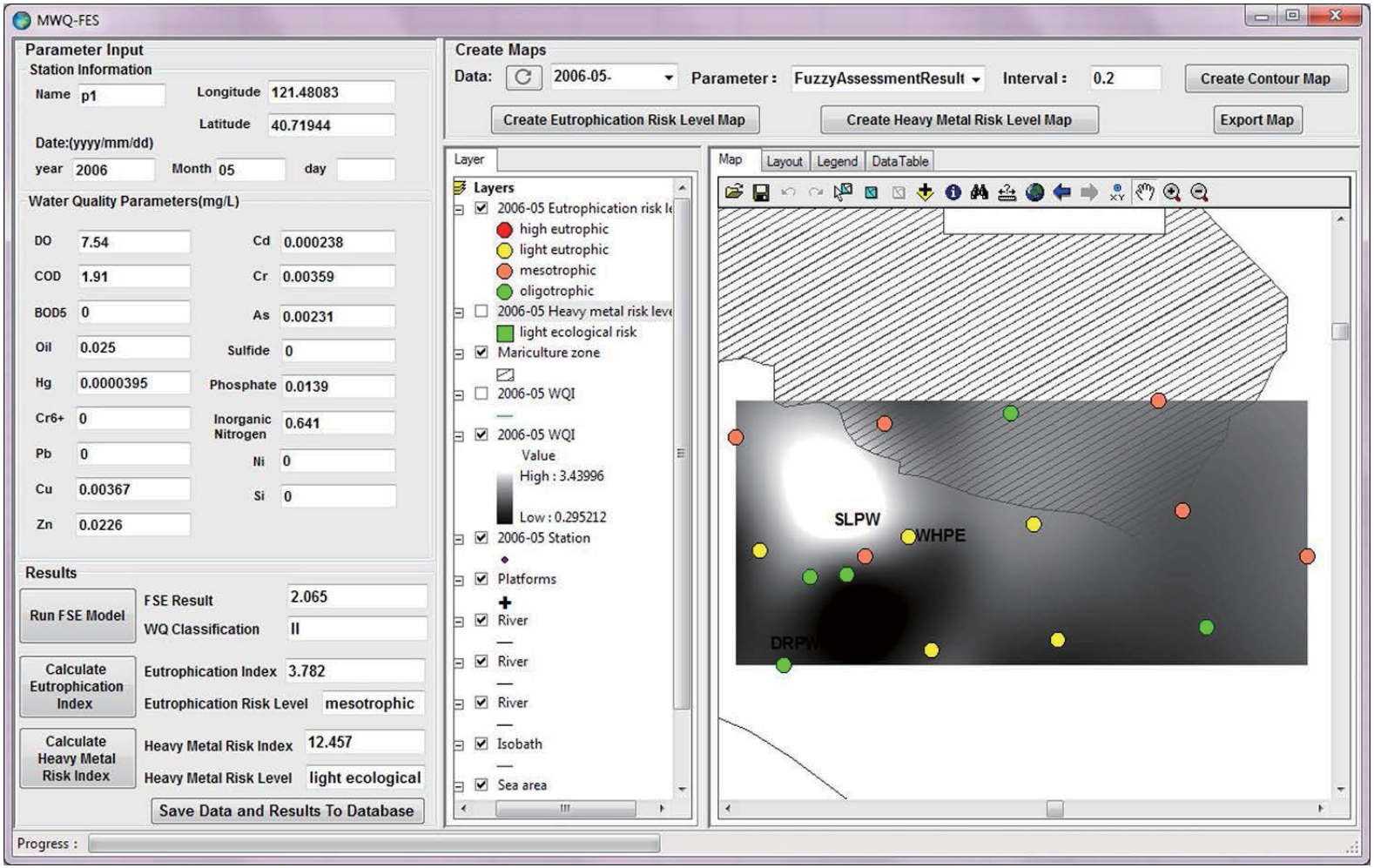


Figure 3-4 Main user interface of MWQ module

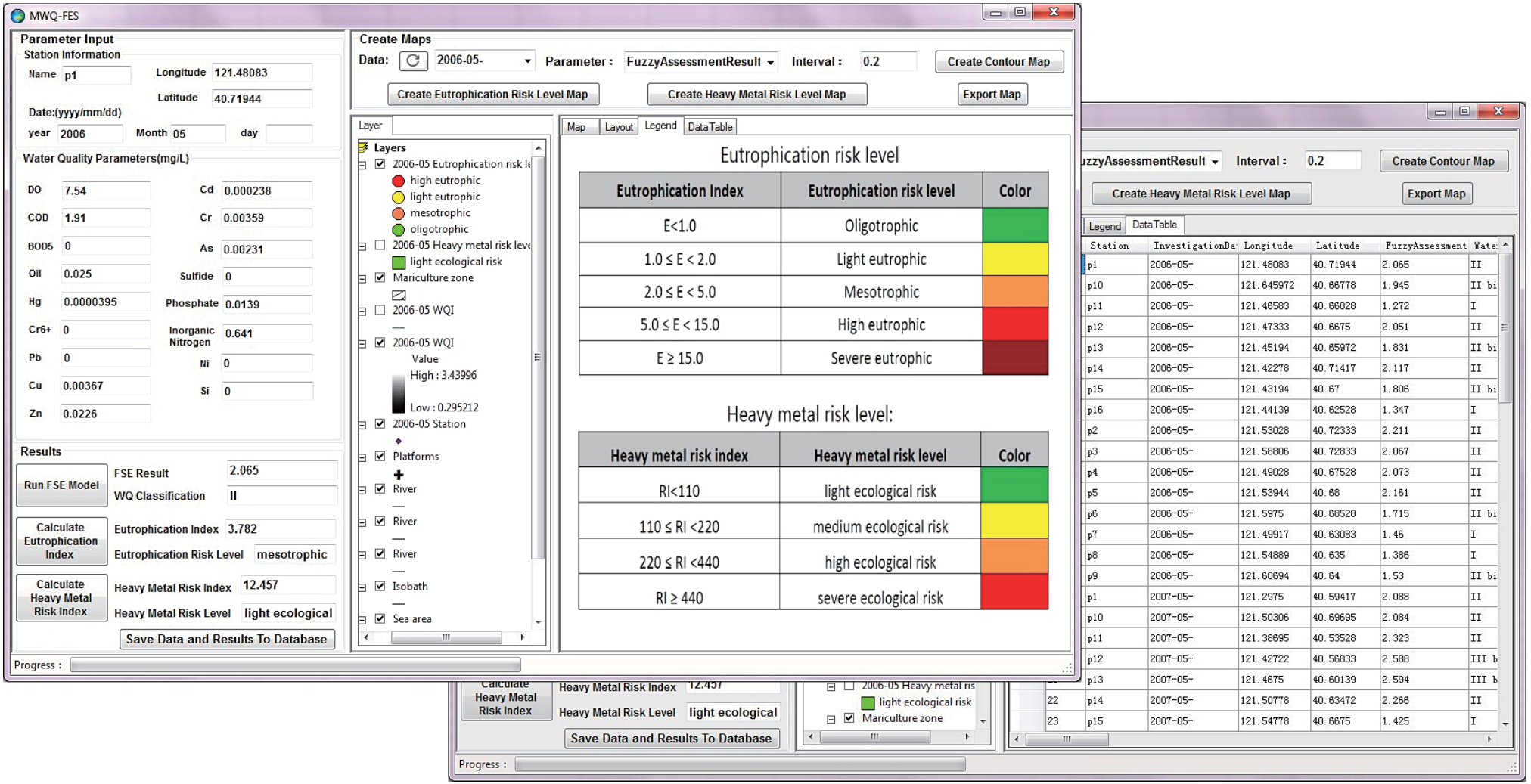


Figure 3-5 Legend and datatable in the interface of the MWQ module

[V] The layout tab in the interface presents the produced maps from the system with longitude and latitude axes, a north arrow, and a plotting scale. To export the produced maps, users can use the “Export Map” button to export the selected maps as images. By right clicking on the selected line in the data table and selecting the export item, the data can be saved as Microsoft Excel files.

For example, input the information for station P1 and the 11 water quality parameters in the input module (Figure 3-4). Then, click the “Run FSE Model” button to obtain the fuzzy synthetic evaluation result and the water quality classification, which is 2.065 and class II, respectively. The eutrophication and heavy metal indices and risk levels can be acquired by clicking the corresponding buttons. By clicking the “Save Data and Results To Database” button, the information and assessment results for station P1 are saved to the database and displayed in the data table. After completing the calculations for all of the stations in one survey, the user can proceed to step IV, choosing the survey date and parameter, setting the contour interval, and subsequently creating contour maps and risk level maps, then exporting the maps and data.

3.5.2.2 Implementation of LWQ module

Similar to the above procedure, the implementation of the lake water quality assessment module of the system can be achieved as follows (Figure 3-6 and Figure 3-7):

[I] Enter the information of the monitoring station, the investigation time and observed values of parameters involved in the assessment, in the “Parameter Input” module.

[II] Click the “Run FSE Model” and “Calculate Trophic State Index” buttons to obtain the fuzzy synthetic evaluation result and the water quality classification, the

trophic state index and trophic status classification, respectively. Results of all of the four trophic state indices calculated previously are displayed in the interface.

[III] All input data and results are saved in a database, which can be displayed using the data table tab in the interface (Figure 3-8).

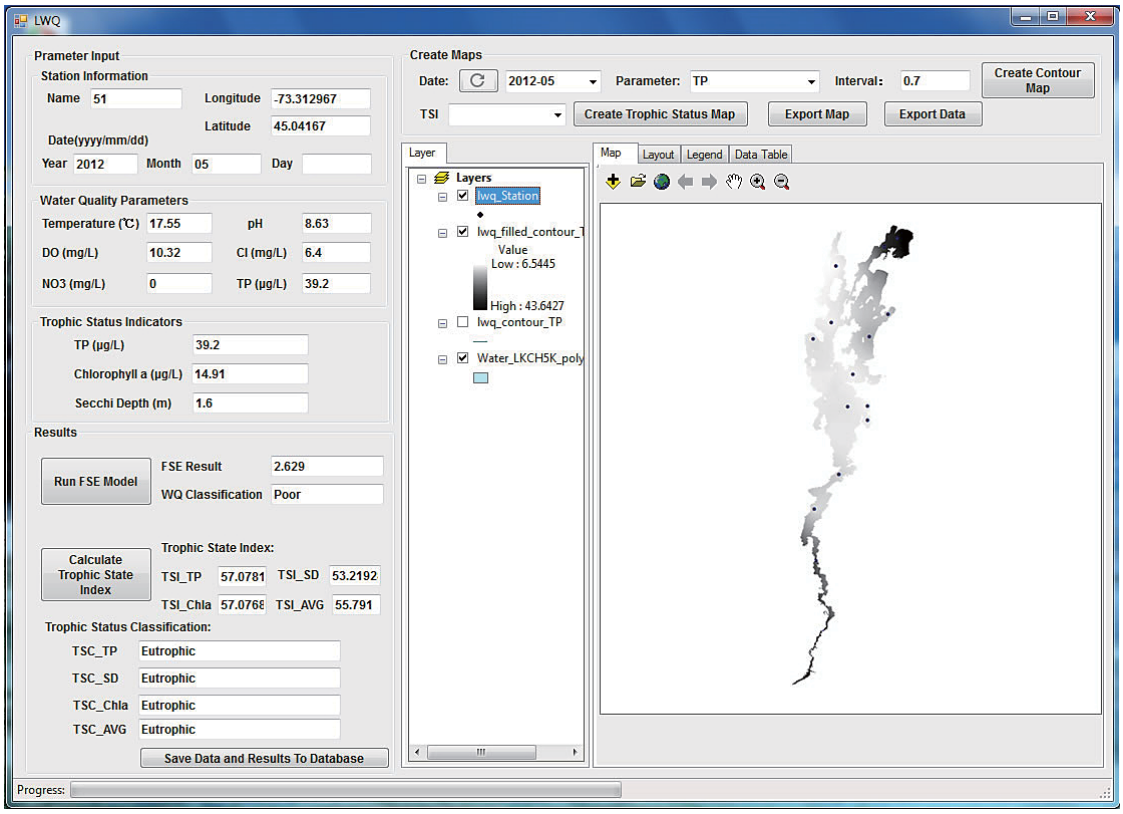


Figure 3-6 Main user interface of LWQ module

[IV] Refresh the data source, select the date of the investigation and the parameter and set the interval of the contours in the map. By clicking the "Create Contour Map" button, the contour map is automatically produced and added to the base map. The user can choose to display a contour map or a filled contour map by clicking the checkbox in front of the layer. The contour map includes contour lines with the corresponding fuzzy water quality index labels adjacent to each line. The filled contour map uses black, white and shades of grey to depict the corresponding fuzzy water quality index, with a legend that shows the corresponding colors.

After selecting the trophic state index in the combobox beside the “TSI” label, click the “Create Trophic Status Map” button to obtain the respective color-coded maps of risk level (Figure 3-7). The legend for these color-coded maps is located in the legend tab in the interface (Figure 3-8).

[V] The layout tab in the interface presents the produced maps from the system with longitude and latitude axes, a north arrow, and a plotting scale. To export the produced maps, users can use the “Export Map” button to export the selected maps as images (Figure 3-9). By clicking the “Export Data” button, the data can be saved as Microsoft Excel files.

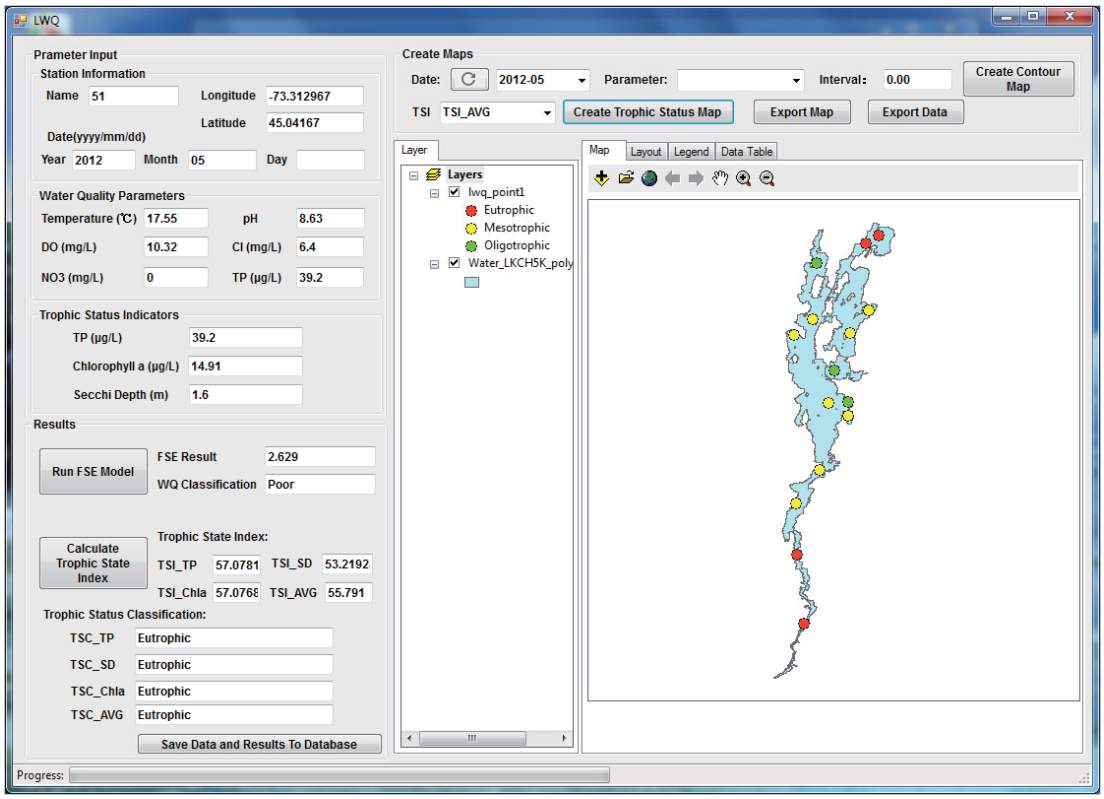


Figure 3-7 Interface of the trophic status assessment

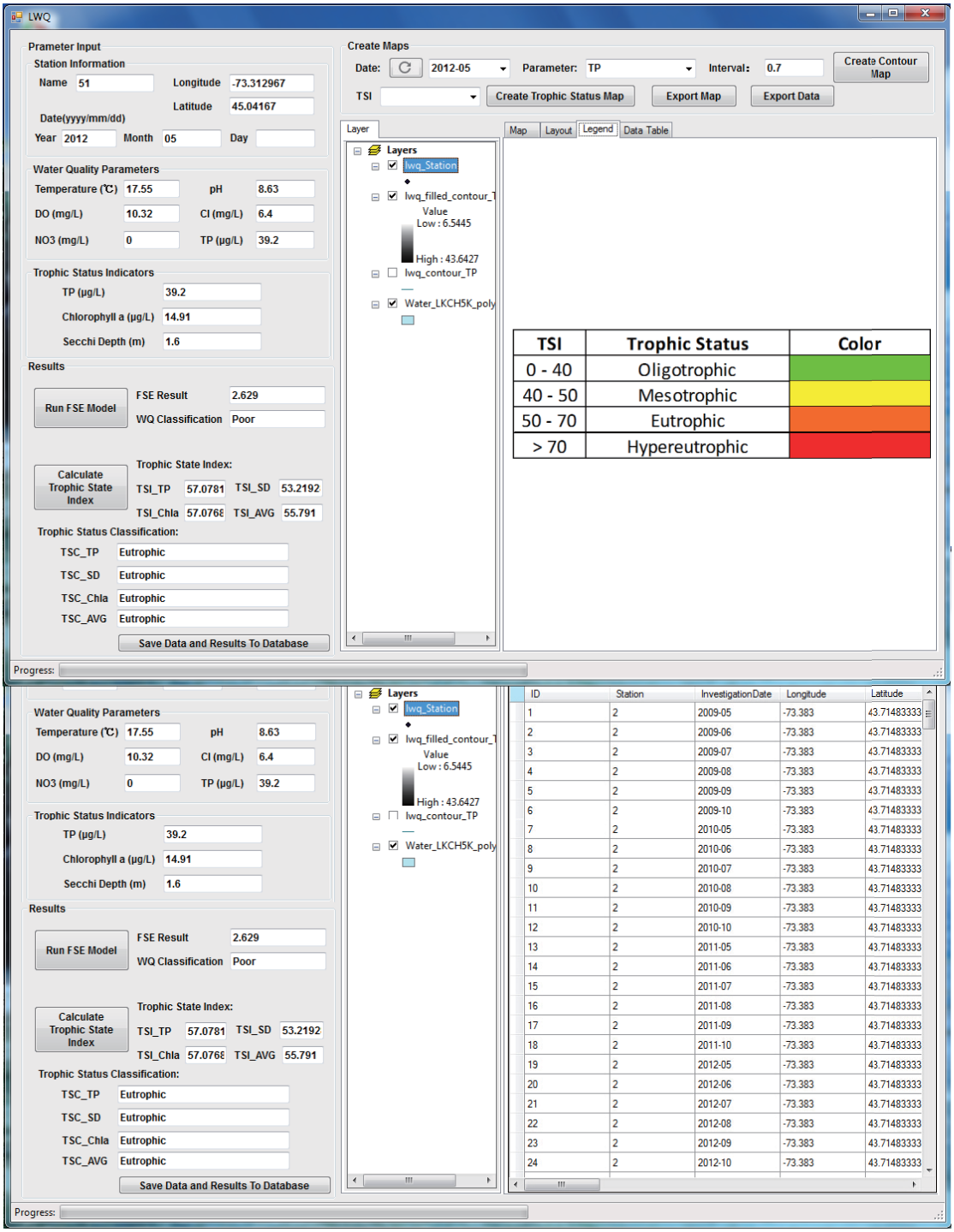


Figure 3-8 Legend and datatable in the interface of the LWQ module

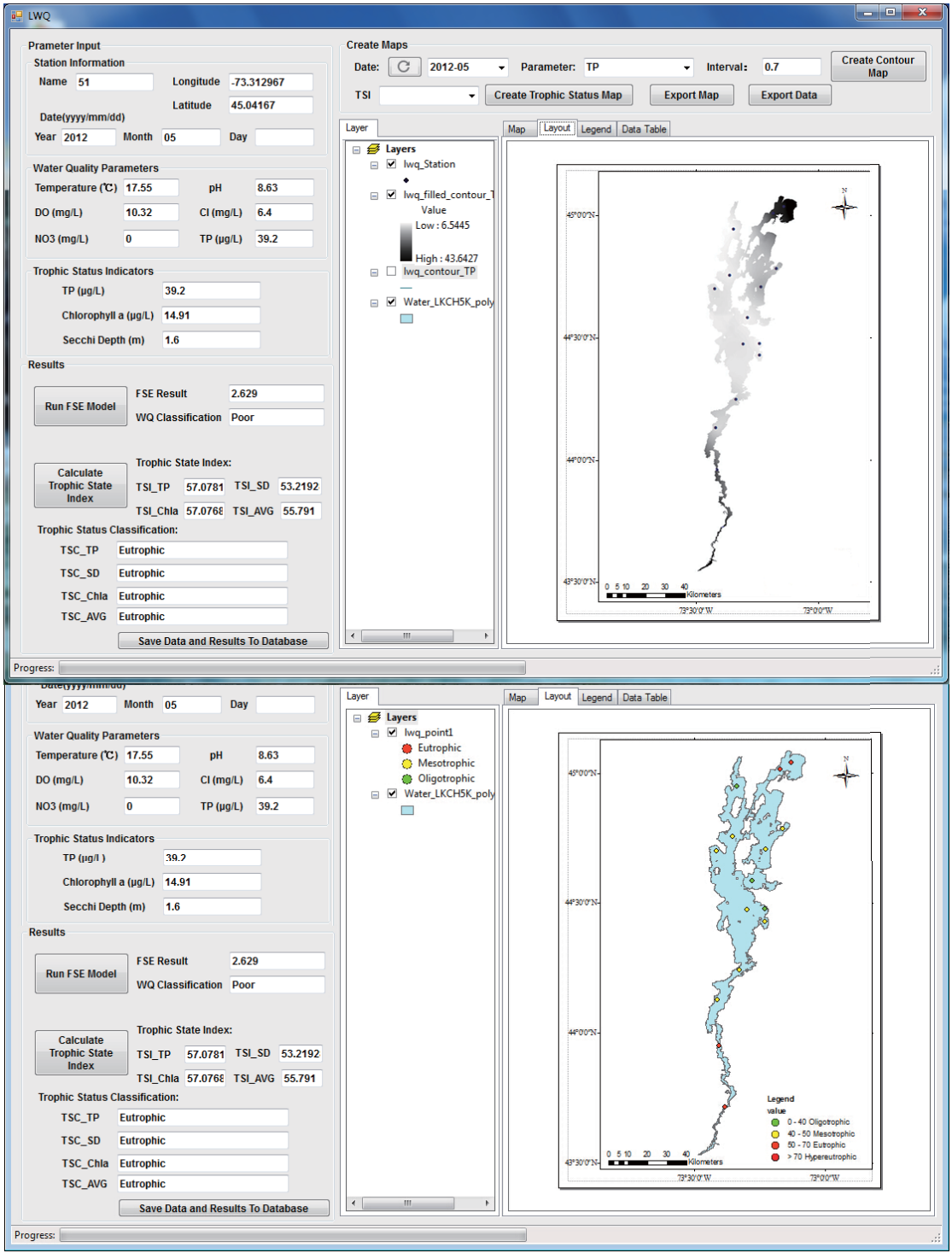


Figure 3-9 Layout page of LWQ and the trophic status assessment results

For example, input the information for station 51 and the seven water quality parameters in the input module (Figure 3-6). Then, click the “Run FSE Model” button

to obtain the fuzzy synthetic evaluation result and the water quality classification, which is 2.629 and within the water quality class of “Poor”, respectively. The trophic state index and trophic status classification can be acquired by clicking the corresponding button. By clicking the “Save Data and Results To Database” button, the information and assessment results for station 51 are saved to the database and displayed in the data table. After completing the calculations for all of the stations in one survey, the user can proceed to step IV, choose the survey date and a parameter, set the contour interval, and subsequently create contour maps and risk level maps, then export the maps and data.

3.5.2.3 Implementation of ANN module

The ANN module of the developed system includes three interconnected parts: training, prediction, and map view. These submodules are able to perform the training of the network, prediction using the trained network, and display the prediction results in the form of spatial distribution maps.

[1] The first step for applying the ANN module is to import training data and validation or testing data from CSV files by clicking the “Load” buttons. The format of these data files should be consistent with the sample data files. The input data files of training, validation and testing data sets contain the following information: the names of the stations, sampling dates, and data of other parameters in the input data set. Similarly, the target data file for training the network and the actual data files for validating and testing the network are comprised of the name of the stations, sampling dates, and data of parameters in the target data set and actual values of the predicted parameter. The names of the stations and sampling dates are key parameters that will be used to pair the input data and the target data or the actual data, as well as the actual values and predicted values.

[2] Determine the structure of the network by selecting the input neuron number, output neuron number, hidden neuron number, and number of hidden layers.

[3] Click the “Start” button on the training tab page to start the training of the network. The network has a tan-sigmoid transfer function for the hidden layer and a linear transfer function for the output layer. The number of epochs and the mean square error of the training data set will be shown on the interface for every 10 epochs. The training can be stopped at any time when the mean square error is satisfied. Sequentially, by clicking the “Save” button the trained network can be saved to a local destination selected by users with the filename extension of “.ntwrk”. The structure, weights, and bias of the trained network is packaged in this file (Figure 3-10).

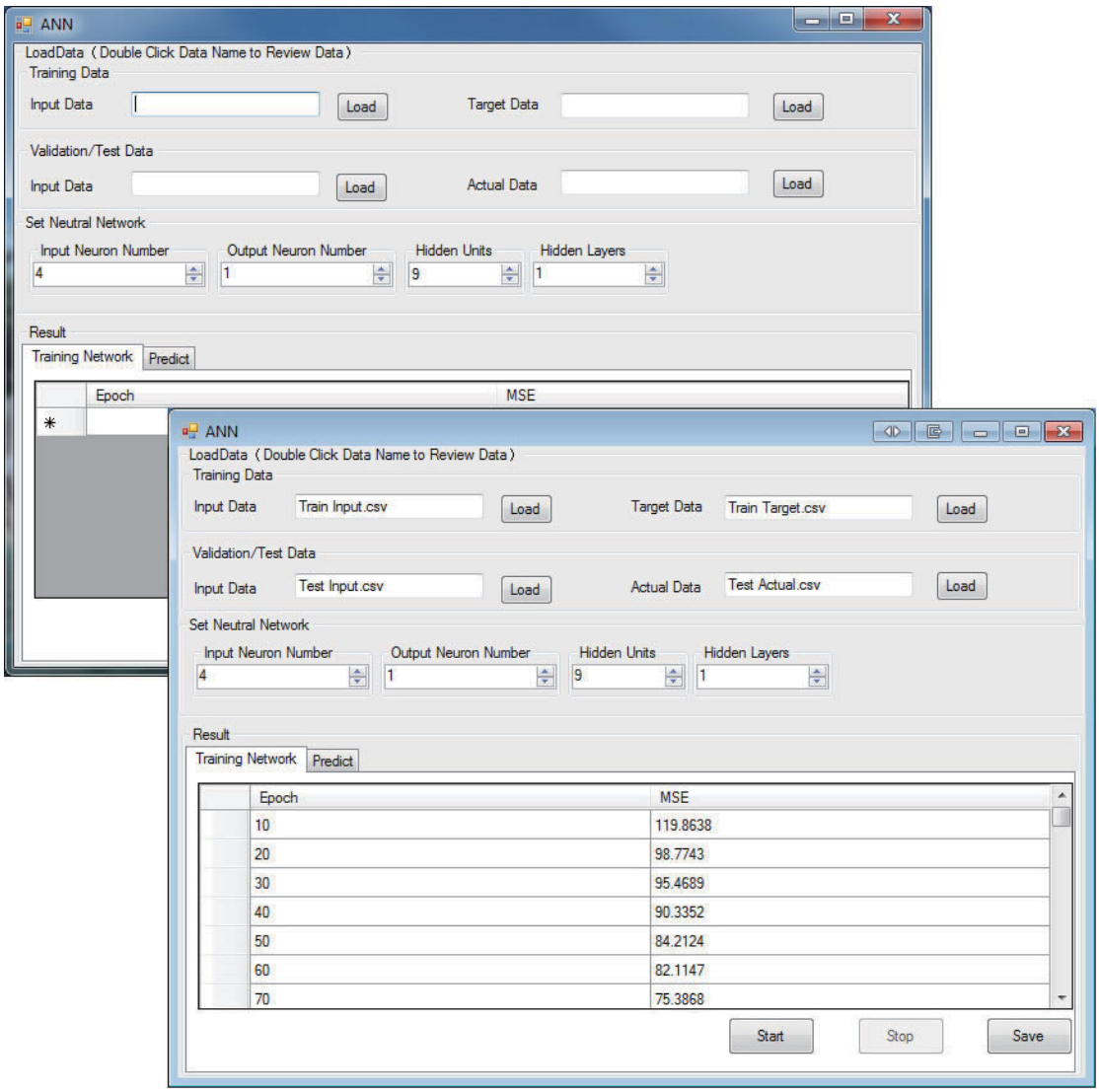


Figure 3-10 Interface of the training submodule of the developed ANN module

[4] Load the saved network file in the prediction tab page through the "Load Network" button, then click the "Predict" button to obtain the predicted values via the trained network. Prediction results will be shown in the data table on the interface, including the name of the stations, dates, actual monitoring values, predicted values, and the absolute errors between the actual and predicted values (Figure 3-11). This operation needs to be done twice to achieve the validation and testing processes by loading the respective validation and testing data.

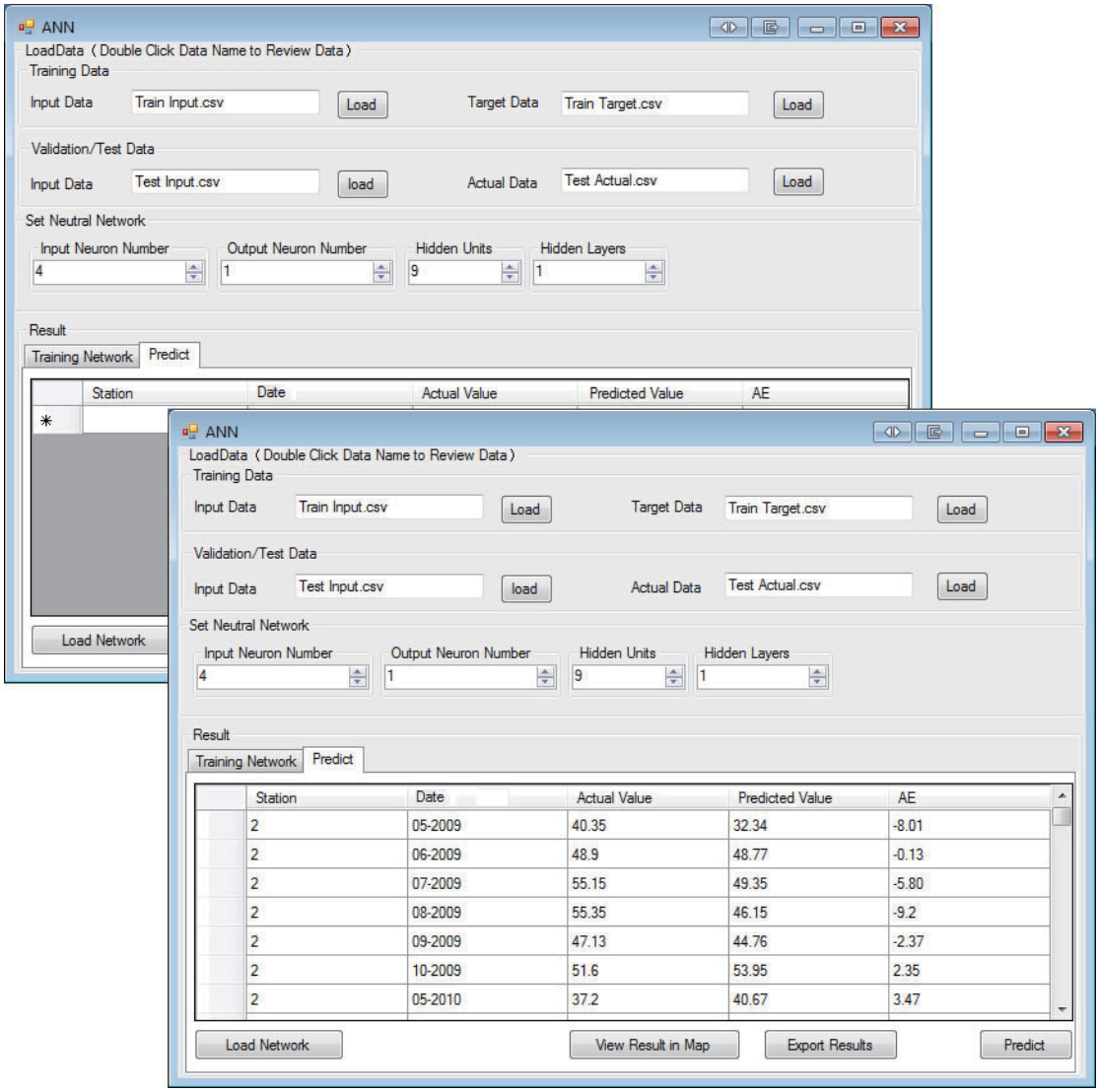


Figure 3-11 Interface of the prediction submodule of the developed ANN module

[5] After each training and prediction, the results could be exported to CSV files by clicking the "Export Results" button at the bottom of the interface. The data can be further organized for the calculation of coefficient of determination (R^2) and RMSE-observations standard deviation ratio (RSR). Click the "View Result in Map" button to create the spatial distribution map of predicted results in the map view window (Figure 3-12). After selecting the date and interval of contours, the map can be generated and displayed in the interface. Two options are available for the display of contour maps: filled contour maps and contour lines with number labels. The maps

with a north arrow, a scale bar, and the legend can be exported to image files after adjusting the position and size of the map in the legend tab page (Figure 3-13).

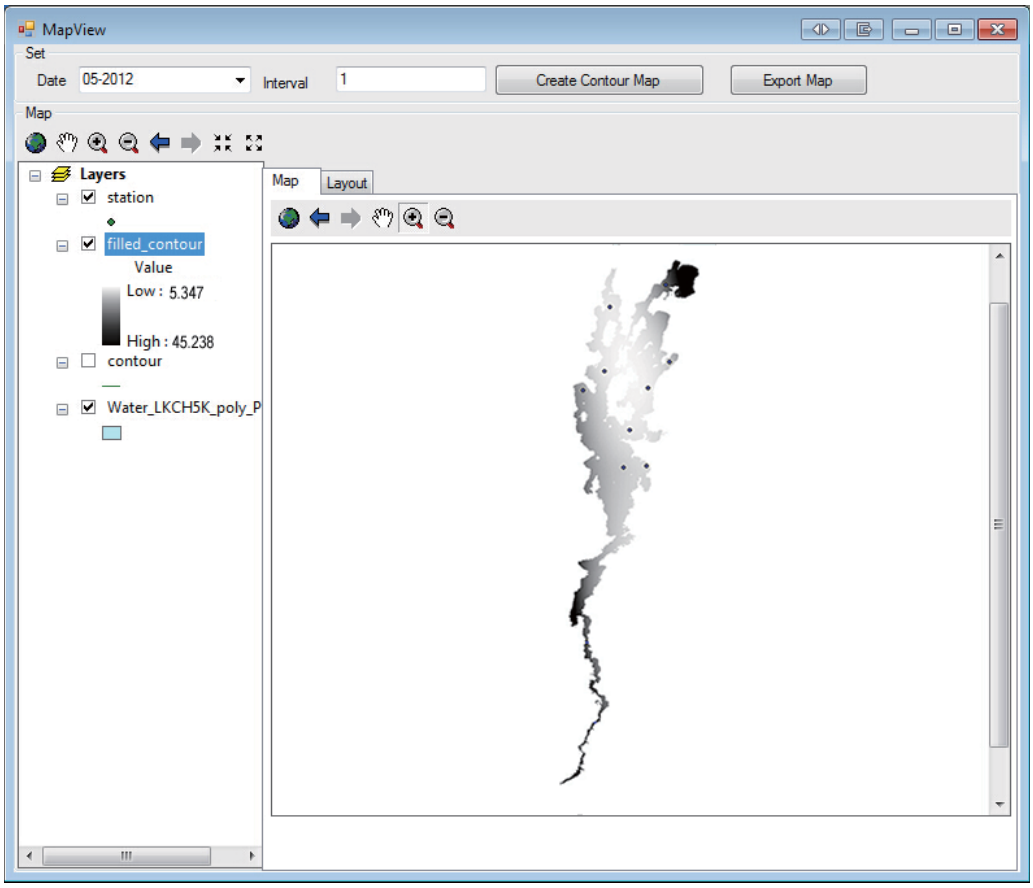


Figure 3-12 Interface of the map view window for ANN predicted results

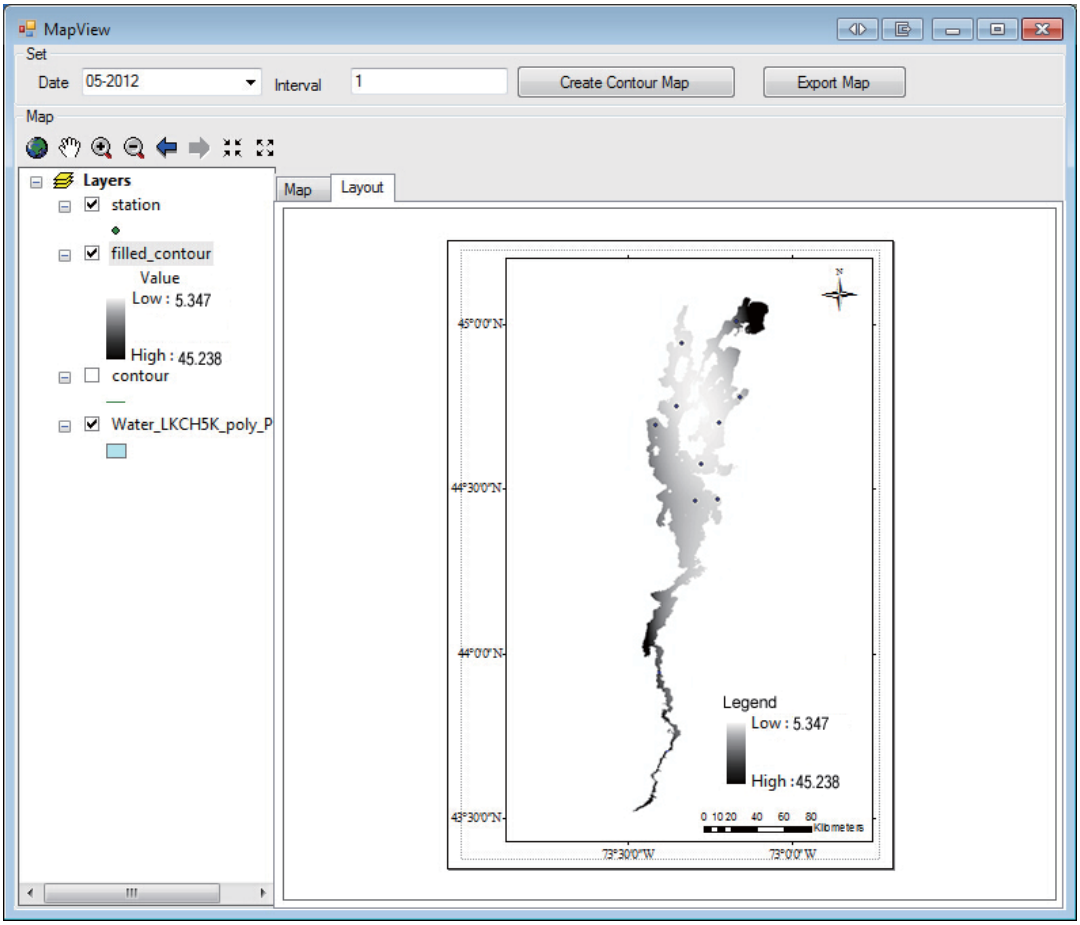


Figure 3-13 Interface of the layout tab page in the map view window

3.6 MIKE 21 Flow Model FM

Considering that the model results of the MIKE Flow model will be compared with the BPANN model that utilized the monitoring data of TP only in the epilimnion layer and unstratified water, the MIKE 21 model was chosen herein with the assumption that the lake water is unstratified.

3.6.1 Description of MIKE 21 Flow Model FM

The MIKE 21 Flow Model is operated through a fully windows integrated graphical user interface (Figure 3-14).

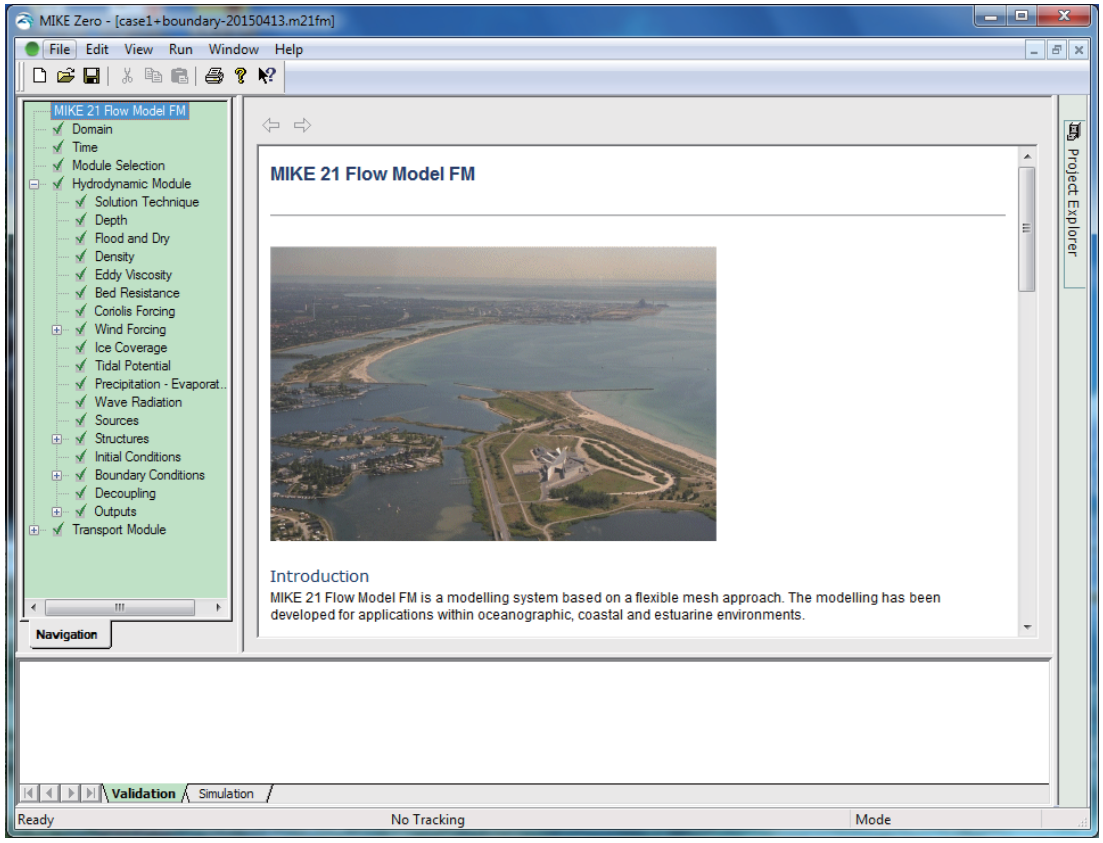


Figure 3-14 The user interface of the MIKE 21 Model FM

All modules in the modelling system are supported by the user interfaces including efficient tools for mesh generation, data retrieve and analysis, 2D/3D visualization, etc. (DHI Water & Environment, 2006). The flexible mesh approach applied in the MIKE 21-HD model has the advantage that the element size in the mesh is varied throughout the model domain, thus model resolution can be varied across the model area.

Two modules were included in this thesis, the hydrodynamic module and the water quality module. Since the ECOLab module includes a large number of parameters to simulate the biological processes in the model domain, among those the Biological oxygen demand (BOD) data is a compulsory parameter. However, BOD is not included in the tributary water chemistry data and the Lake Champlain monitoring program, as organic pollutant is not the main concern of water quality in Lake

Champlain. Consequently, the transport module was utilized in this thesis to simulate the transport of the pollutants.

3.6.2 Hydrodynamic module

The hydrodynamic module is the basic module in the MIKE 21 Flow model system. It is able to simulate the hydrodynamic variation under the effect of many forces, which performs as the basic module for water quality module, sand transport module, oil spill module, etc.

3.6.2.1 Governing equations in Cartesian Co-ordinates (MIKE by DHI, 2012)

The two-dimensional shallow water equations are as follows:

$$\frac{\partial h}{\partial t}+\frac{\partial}{\partial x}(hu)+\frac{\partial}{\partial y}(hv)=hS \tag{3.21}$$

$$\frac{\partial h\bar{u}}{\partial t}+\frac{\partial h\bar{u}^{2}}{\partial x}+\frac{\partial h\overline{uv}}{\partial y}=f\bar{v}h-gh\frac{\partial \eta}{\partial x}-\frac{h}{\rho_0}\frac{\partial P_a}{\partial x}-\frac{gh^2}{2\rho_0}\frac{\partial \rho}{\partial x}+\frac{\partial}{\partial x}\left(hT_{xx}\right)+\frac{\partial}{\partial \mathrm{y}}\left(hT_{xy}\right)+ hu_sS+\frac{\tau_{sx}}{\rho_0}-\frac{\tau_{bx}}{\rho_0}-\frac{1}{\rho_0}\left(\frac{\partial s_{xx}}{\partial x}+\frac{\partial s_{xy}}{\partial y}\right) \tag{3.22}$$

$$\frac{\partial h\bar{v}}{\partial t}+\frac{\partial h\overline{uv}}{\partial x}+\frac{\partial h\bar{v}^{2}}{\partial y}=-f\bar{u}h-gh\frac{\partial \eta}{\partial y}-\frac{h}{\rho_0}\frac{\partial P_a}{\partial y}-\frac{gh^2}{2\rho_0}\frac{\partial \rho}{\partial y}+\frac{\partial}{\partial x}\left(hT_{xy}\right)+\frac{\partial}{\partial \mathrm{y}}\left(hT_{yy}\right)+ hv_sS+\frac{\tau_{sy}}{\rho_0}-\frac{\tau_{by}}{\rho_0}-\frac{1}{\rho_0}\left(\frac{\partial s_{yx}}{\partial x}+\frac{\partial s_{yy}}{\partial y}\right) \tag{3.23}$$

where t is the time; x and y are the Cartesian co-ordinates; η is the surface elevation; d is the equilibrium water depth; $h=\eta+d$ is the total water depth; u and v are the velocity components in the x and y direction; $f=2\omega\sin\phi$ is the Coriolis parameter (ω is the angular rate of revolution and ϕ is geographic latitude); g is the gravitational acceleration; ρ is the density of water; ρ_0 is the reference density of

water; s_{xx} , s_{xy} , s_{yx} , s_{yy} are components of the radiation stress tensor; P_a is the atmospheric pressure. S is the magnitude of the discharge due to point sources and (u_s, v_s) is the velocity that the water is discharged into the ambient water. The overbar indicates a depth average value. T_{ij} represents the lateral stresses and are estimated using the eddy viscosity formulation based on the depth average velocity gradients.

3.6.2.2 Bottom stress

The Quadratic friction law is used to determine the bottom stress:

$$\frac{\vec{\tau}_b}{\rho_0} = c_f \vec{u}_b \left|\vec{u}_b\right| \tag{3.24}$$

where c_f is the drag coefficient; $\vec{u}_b = (u_b,\ v_b)$ is the flow velocity above the bottom. In 2D modeling calculations, $\vec{u}_b$ is the depth-average velocity and the drag coefficient can be determined from the Chezy coefficient (C) or the Manning roughness (M):

$$c_f = \frac{g}{C^2} \tag{3.25}$$

$$c_f = \frac{g}{(Mh^{1/6})^2} \tag{3.26}$$

3.6.2.3 Wind stress (MIKE by DHI, 2012)

Since the data used in this thesis were collected during the months that have no ice coverage, the effects of ice coverage are not considered herein.

The surface stress in areas not covered by ice is determined by the winds above the surface. The stress is given by:

$$\bar{\tau}_s = \rho_a c_d |u_w| \bar{u}_w \tag{3.27}$$

where ρ_a is the density of air, c_d is the drag coefficient of air, and $\bar{u}_w = (u_w, v_w)$ is the wind speed 10 m above the water surface.

The drag coefficient can either be a constant value or it depends on the wind speed. The parameterization of the drag coefficient is determined by the following empirical formula:

$$c_d = \begin{cases} c_a & w_{10} < w_a \\ c_a + \dfrac{c_b - c_a}{w_b - w_a}(w_{10} - w_a) & w_a \le w_{10} < w_b \\ c_b & w_{10} \ge w_b \end{cases} \tag{3.28}$$

where c_a, c_b, w_a and w_b are empirical factors, w_{10} is the wind velocity 10 m above the water surface. The default values for these empirical factors in open sea area are $c_a = 1.255 \times 10^{-3}, c_b = 2.425 \times 10^{-3}, w_a = 7m/s$ and $w_b = 25m/s$. The drag coefficient over the lakes is larger than the open sea area.

3.6.2.4 Numerical solution

(1) Spatial discretization

The 2D shallow water equations in Cartesian coordinates can be written as:

$$\frac{\partial U}{\partial t} + \frac{\partial (\mathrm{F}_x^I - \mathrm{F}_x^V)}{\partial x} + \frac{\partial (\mathrm{F}_y^I - \mathrm{F}_y^V)}{\partial y} = S \tag{3.29}$$

where the superscripts *I* and *V* are the inviscid and viscous fluxes, respectively. *U* is the vector of conserved variables, *F* is the flux vector function and *S* is the vector of source terms.

Both a first order and a second order scheme can be applied for the spatial discretization. An approximate Riemann solver is used to calculate the convective fluxes at the interface of the cells in 2D modelling system.

(2) Time integration

For 2D simulations, two methods are available for the time integration for the shallow water equations and the transport equations: A lower order method and a higher order method.

The lower order method is a first order explicit Euler method:

$$U_{n+1} = U_n + \Delta t G(U_n) \tag{3.30}$$

where Δt is the time step interval.

The higher order method uses a second order Runge Kutta method:

$$\begin{aligned} U_{n+\frac{1}{2}} &= U_n + \frac{1}{2}\Delta t G(U_n) \\ U_{n+1} &= U_n + \Delta t G(U_{n+\frac{1}{2}}) \end{aligned} \tag{3.31}$$

(3) Boundary conditions

Normal fluxes along closed boundaries like land boundaries are forced to zero for all variables. For the momentum equations, this leads to full-slip along land boundaries. The open boundary conditions can be specified in the form of a unit discharge or as the surface elevation for the hydrodynamic equations. A specified value or a specified gradient can be given for transport equations.

The depth in each element/cell is monitored and the elements are classified as dry, partially dry or wet. The element faces are also monitored to identify flooded boundaries.

The wetting depth h_{wet}, drying depth h_{dry}, and flooding depth h_{flood} are used to define whether an element is dry, wet or partially dry. These three factors must satisfy the following relationship:

$$h_{dry} < h_{flood} < h_{wet} \tag{3.32}$$

The default values are $h_{dry} = 0.005m$, $h_{flood} = 0.05m$, and $h_{wet} = 0.1m$.

3.6.3 Transport module

The transport module simulates the transport and fate of dissolved or suspended substances in an aquatic environment under the influence of the fluid transport and dispersion processes. The hydrodynamic module provides the hydrodynamic basis for the transport module. Typical applications of transport module include tracer simulations, flushing studies and simple water quality studies, which is the purpose of applying MIKE 21 model in this study (DHI Software for Water Environment).

The conservation equation for calculating the transport of a scalar quantity is given by

$$\frac{\partial C}{\partial t} + \frac{\partial uC}{\partial x} + \frac{\partial vC}{\partial y} + \frac{\partial wC}{\partial z} = F_C + \frac{\partial}{\partial z}(D_v \frac{\partial C}{\partial z}) - k_p C + C_S S \tag{3.33}$$

The horizontal diffusion term is defined by

$$F_C = \left[\frac{\partial}{\partial x}\left(D_h \frac{\partial}{\partial x} \right) + \frac{\partial}{\partial y}\left(D_h \frac{\partial}{\partial y} \right) \right] C \tag{3.34}$$

The conservation equation is integrated over depth for 2D calculations and defined by (MIKE by DHI, 2012)

$$\frac{\partial h\overline{C}}{\partial t}+\frac{\partial h\overline{u}\overline{C}}{\partial x}+\frac{\partial h\overline{v}\overline{C}}{\partial y}=hF_C-hk_p\overline{C}+hC_SS \tag{3.35}$$

where t is time, x, y and z are the Cartesian coordinates, D_v is the vertical turbulent diffusion coefficient, S is the magnitude of discharge due to point sources, F_C is the horizontal diffusion term, D_h is the horizontal diffusion coefficient, h is the water depth, $\overline{u}$ and $\overline{v}$ are the depth-averaged velocity components, C is the concentration of scalar quantity, k_p is the linear decay rate of scalar quantity, and C_S is the concentration of scalar quantity in source.

Information about the components to simulate should be input to the transport model, including component type, dispersion coefficients, decay information, initial conditions, and boundary conditions (DHI Software for Water Environment).

3.7 Summary

This chapter described the methodologies of the development of a surface water quality assessment and prediction system (GIS-SWQAM) and the MIKE 21 flow model for the surface water quality modeling. Specifically, the methodologies used in this thesis are summarized as follows:

(1) Developing the GIS-based marine water quality assessment system (MWQ module) and the lake water quality assessment system (LWQ module), which consists of a fuzzy synthetic assessment model combined with a eutrophic risk assessment model and a heavy metal risk assessment model, in the C# programing language, integrated with the ArcGIS Engine and a database. The

ArcGIS Engine is employed to handle the spatial information. Through integration with ArcGIS Engine, it allows for spatial analysis on both monitoring and assessment results with a georeferenced database, as well as the visualization of results. The application of the MWQ module and the LWQ module will be illustrated in Chapter 4 and Chapter 5, respectively.

(2) Developing the back propagation artificial neural network model (BPANN) using Matlab software. Afterwards, the BPANN algorithm is integrated with the ArcGIS Engine and a database using C# code to build the GIS-based surface water quality prediction system (ANN module). The ArcGIS Engine is used as the post-processor in this part to achieve the spatial visualization of model results. The methods for selecting the optimal input parameter of the model, optimizing the model structure, and two statistical parameters for evaluating the performance of the model are also listed. The application of the BPANN model developed using Matlab will be presented in Chapter 6 (Section 6.2) for the prediction of chlorophyll-a concentration in Lake Champlain. The ANN module developed will be applied to predict total phosphorus concentration in Lake Champlain, which will also be illustrated in Chapter 6 (Section 6.3).

(3) Constructing the MIKE 21 flow model to simulate the concentration of water quality parameters. The comparison of MIKE 21 flow model results and the artificial neural network model results are performed sequentially. The data and parameters included in constructing the MIKE 21 model and the results comparison will be described in Chapter 7.

CHAPTER 4 MARINE WATER QUALITY ASSESSMENT VIA THE GIS-SWQAM (MWQ MODULE) -- A Case Study Near Offshore Oil Platforms in Northern Liaodong Bay of China

4.1 Study Area

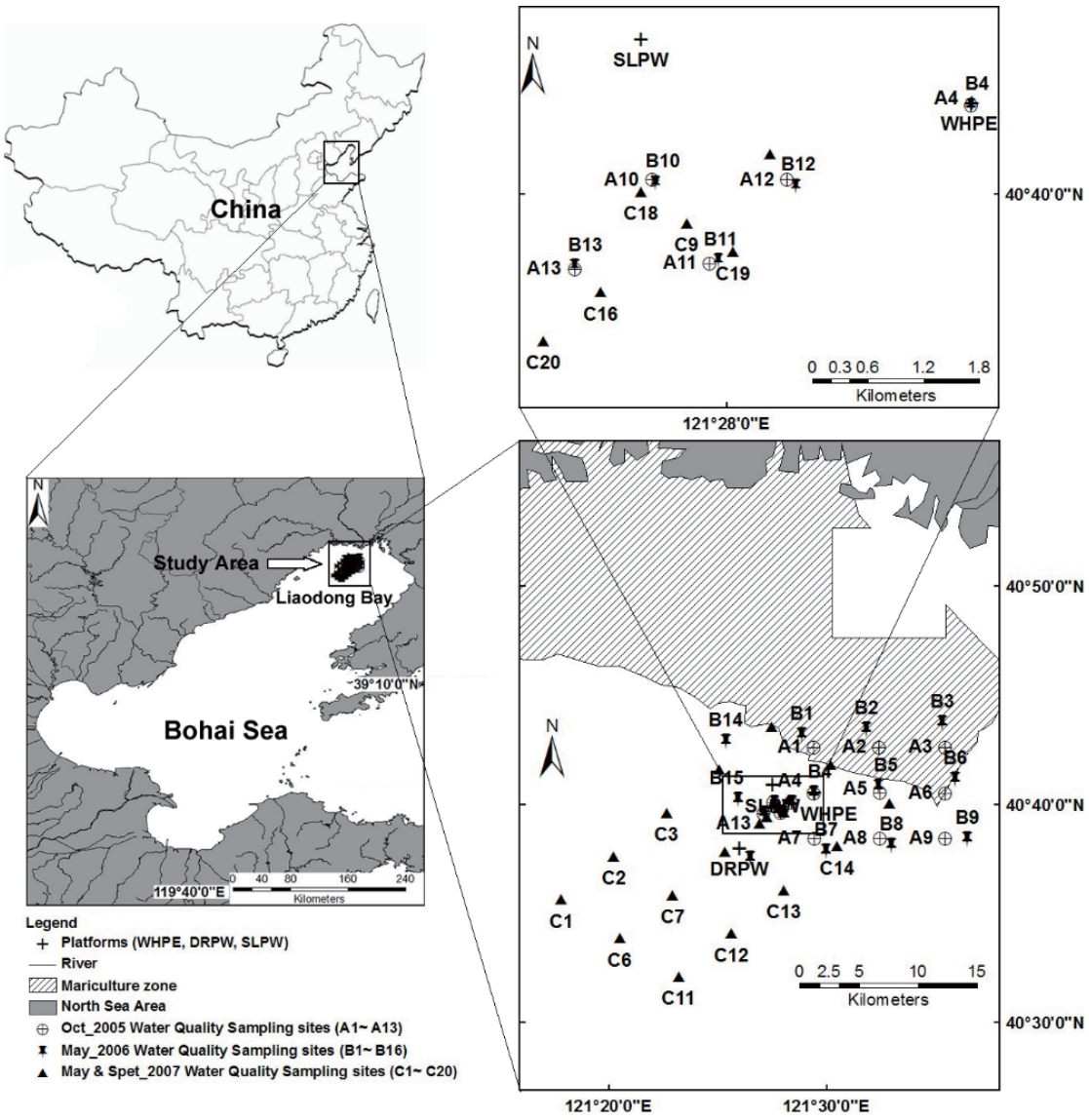


Figure 4-1 Study area and the sampling sites for the 4 field investigations during 2005 to 2007

The location of the study area is presented in Figure 4-1. The minimum distance from the selected platforms to the coastline is 15 km. Water depths in this area range from 6.5 to 10.5 meters. The tide is regular semi-diurnal with a high slack tide of 10 meters and a low tide of 6 meters. The Liao, Shuang-Tai-Xi, Dalinghe and Xiaolinghe Rivers flow into the northern area of Liaodong Bay. Some significant environmentally sensitive areas are located adjacent to the study area, e.g., the mariculture zones. Hence, water quality in this area is of crucial importance to the protection and development of these sensitive areas.

4.2 Data Sources

The data were provided by the North China Sea Monitoring Center. As displayed in Figure 4-1, water samples were collected from 13 sites in October 2005, 16 sites in May 2006, and 20 sites in May and Sept. 2007. The samples were collected at two depths when the water depth exceeded 10 meters, i.e., at approximately 0.5 m from the surface and 2 m above the bottom, corresponding to the surface and bottom levels, respectively. If the water depth was less than 10 m, only the surface water samples were collected. Water samples were collected using CTD and properly stored until laboratory analysis. Water quality parameters, including concentrations of DO, COD, phosphate, inorganic nitrogen, oil, Hg, Cu, Zn, Cd, Cr, and As, were determined by laboratory analysis following standard procedures (GB3097, 1997). Only the analysis of the surface water samples is presented herein.

4.3 System Implementation - MWQ Module

Using the data collected during each survey, the MWQ can be applied according to the five steps introduced in Section 3.5.2.1 to obtain the water quality assessment, eutrophication risk and heavy metal risk evaluation results.

Standard values based on the marine water quality standards in China (GB3097, 1997) were used as the criteria for water quality classification in the developed system as listed in Table 4-1.

According to the values of the fuzzy marine water quality index (B*), seven water quality classes were derived; the corresponding water quality statuses are defined in Table 4-2. The classification "Class II biased toward I" means that the water quality was determined to be close to class II but still better than class II; the other biased classifications have similar meanings.

Table 4-1 Criteria of marine water quality

Factors	Criteria of marine water quality grades (mg/L)			
	Grade S_1	Grade S_2	Grade S_3	Grade S_4
Dissolved oxygen (DO)	6	5	4	3
COD	2	3	4	5
Phosphate (p)	0.015	0.03	0.03	0.045
Inorganic nitrogen (IN)	0.2	0.3	0.4	0.5
Oil	0.05	0.3	0.3	0.5
Mercury (Hg)	0.00005	0.0002	0.0002	0.0005
Copper (Cu)	0.005	0.01	0.05	0.05
Zinc (Zn)	0.02	0.05	0.1	0.5
Cadmium (Cd)	0.001	0.005	0.01	0.01
Chromium (Cr)	0.05	0.1	0.2	0.5
Arsenic (As)	0.02	0.03	0.05	0.05

Table 4-2 Water quality classifications based on B* values

Range of B* Values	Water Quality Classification	Water Quality Status
0~1.5	Class Ⅰ	Very good
1.5~2.0	Class Ⅱ biased toward Ⅰ	Good
2.0~2.5	Class Ⅱ	Fair
2.5~3.0	Class Ⅲ biased toward Ⅱ	Bad
3.0~3.5	Class Ⅲ	Very bad
3.5~4.0	Class Ⅳ biased toward Ⅲ	Poor
＞4.0	Class Ⅳ	Very poor

Seven parameters were considered in the heavy metal risk evaluation in this study, e.g., the concentrations of Hg, Cd, As, Pb, Cu, Cr, and Zn. The selection of reference

concentrations of heavy metals was based on the heavy metal concentrations in the water quality standard for fisheries and grade S_2 of seawater quality standard (GB3097, 1997; GB11607-89, 1989; Pu et al., 2012) (Table 4-3).

Table 4-3 Heavy metal risk level corresponding to RI intervals

Heavy metal risk index	Heavy metal risk level	Color
RI<110	Light ecological risk	
110 ≤ RI <220	Medium ecological risk	
220 ≤ RI <440	High ecological risk	
RI ≥ 440	Severe ecological risk	

4.4 Results

The monitoring data and the fuzzy risk assessment results were exported from the database to an excel file. The average values of each parameter and the average of B* values of each investigation were obtained after performing data reorganization. The monitoring data and assessment results were listed in the following tables respectively.

4.4.1 Monitoring results

Comparing the monitoring data of the 11 water quality parameters listed in Table 4-4 with the criteria of marine water quality in Table 4-1, the phosphate (P) concentration exceeded grade S3 of marine water quality criteria at all the sampling sites in Oct. 2005. The result of May 2006 has significantly lower P concentration, as the P concentration was between grade S1 and grade S2 at 25% of the stations, with the P concentration of the rest of the stations met the grade S1 criteria. In May and September of 2007, the P concentration increased significantly over the previous survey. P concentration in September 2007 met the criteria of grade S2 at all stations with a higher concentration than the measured value in May 2006. The situation was more serious in May 2007 with the P concentration of 20% of the sampling stations failing to meet the criteria of grade S3.

Inorganic nitrogen (IN) concentration in Oct. 2005 was the lowest in these four investigations with values lower than the grade S2 criteria at all sites. But the concentration was particularly high in May 2006 and May 2007, failing to meet the criteria of grade S3 at 69% of the sampling sites in May 2006 and at 95% of all sites in May 2007. In Sept. 2007, the IN concentration was lower than previous survey with the values higher than grade S3 criteria at 65% of all sites. Zinc (Zn) concentration in Oct. 2005 met the criteria of grade S1. It was higher than the grade S1 criteria at 37% of the sites in May 2006, 15% of the sites in May 2007 and 65% of the sites in Sept. 2007. The mercury concentration met the criteria of grade S1 at all stations in May 2006 and May 2007, while it was higher than the criteria of grade S1 at 69% of all sites in Oct. 2005 and 10% of the sites in Sep. 2007. The concentration of other parameters met the criteria of grade S1 in the four investigations.

In summary, the main contaminants in the study area were phosphate, inorganic nitrogen, zinc, and mercury.

Table 4-4 Marine water quality monitoring results

	October 2005		May 2006		May 2007		September 2007	
Parameters	Range (mg/L)	Average (mg/L)	Range (mg/L)	Average (mg/L)	Range (mg/L)	Average (mg/L)	Range (mg/L)	Average (mg/L)
DO	6.74~9.07	8.361	7.54~8.76	7.873	8.26~9.34	8.95	6.90~7.58	7.02
COD	0.32~1.36	0.945	0.472~1.91	1.194	1.00~1.84	1.355	0.688~1.29	0.973
Phosphate	0.0336~0.0455	0.0403	0.00463~0.0171	0.0118	0.0194~0.0338	0.0257	0.00572~0.02	0.0168
Inorganic Nitrogen	0.162~0.231	0.208	0.315~0.687	0.515	0.36~1.069	0.709	0.213~0.594	0.433
Oil	0.0155~0.0278	0.0222	0.0172~0.0688	0.0328	0.0299~0.0747	0.0436	0.0116~0.0263	0.0177
Hg	4.15E-05~8.65E-05	5.67E-05	3.64E-05~4.57E-05	3.97E-05	2.86E-05~5E-05	4.06E-05	3.31E-05~5.26E-05	4.09E-05
Cu	2.74E-03~3.58E-03	3.03E-03	2.50E-03~3.88E-03	3.19E-03	2.41E-03~3.31E-03	2.88E-03	2.53E-03~3.42E-03	3.07E-03
Zn	0.0152~0.0178	0.0165	0.0153~0.0229	0.0196	0.0148~0.0204	0.0175	0.0144~0.0277	0.0225
Cd	1.46E-04~1.92E-04	1.65E-04	1.49E-04~2.38E-04	1.97E-04	1.3E-04~1.8E-04	1.50E-04	1.3E-04~1.9E-04	1.57E-04
Cr	2.26E-03~3.27E-03	2.89E-03	2.73E-03~3.67E-03	3.17E-03	2.49E-03~3.29E-03	2.94E-03	2.8E-03~4.14E-03	3.36E-03
As	2.36E-03~4.06E-03	3.08E-03	1.63E-03~2.31E-03	1.20E-03	1.118E-03~1.26E-03	1.21E-03	6.65E-04~3.38E-03	1.60E-03

Table 4-5 Marine water quality assessment results

	Oct. 2005			May 2006			May 2007			Sept. 2007		
Station	B*[a]	WQC[b]	WQS[c]	B*	WQC	WQS[c]	B*	WQC	WQS[c]	B*	WQC	WQS[c]
p1	1.728	II biased toward I	Good	2.065	II	Fair	2.088	II	Fair	1.289	I	Very Good
p2	1.744	II biased toward I	Good	2.211	II	Fair	1.980	II biased toward I	Good	1.337	I	Very Good
p3	1.748	II biased toward I	Good	2.067	II	Fair	2.097	II	Fair	1.337	I	Very Good
p4	1.656	II biased toward I	Good	2.073	II	Fair	2.123	II	Fair	1.680	II biased toward I	Good
p5	1.941	II biased toward I	Good	2.161	II	Fair	2.101	II	Fair	1.761	II biased toward I	Good
p6	1.875	II biased toward I	Good	1.715	II biased toward I	Good	2.464	II	Bad	1.025	I	Very Good
p7	1.937	II biased toward I	Good	1.460	I	Very Good	2.007	II	Fair	1.502	II biased toward I	Good
p8	1.984	II biased toward I	Good	1.386	I	Very Good	2.014	II	Fair	1.986	II biased toward I	Good
p9	1.889	II biased toward I	Good	1.530	II biased toward I	Good	2.098	II	Fair	1.966	II biased toward I	Good
p10	1.638	II biased toward I	Good	1.945	II biased toward I	Good	2.084	II	Fair	2.012	II	Fair
p11	1.773	II biased toward I	Good	1.272	I	Very Good	2.323	II	Fair	1.429	I	Very Good
p12	1.709	II biased toward I	Good	2.051	II	Fair	2.588	III biased toward II	Bad	2.053	II	Fair
p13	1.624	II biased toward I	Good	1.831	II biased toward I	Good	2.594	III biased toward II	Bad	1.601	II biased toward I	Good
p14	——	——		2.117	II	Fair	2.266	II	Fair	1.940	II biased toward I	Good
p15	——	——		1.806	II biased toward I	Good	1.425	I	Very Good	2.116	II	Fair
p16	——	——		1.347	I	Very Good	1.990	II biased toward I	Good	2.053	II	Fair
p17	——	——		——	——		2.163	II	Fair	1.774	II biased toward I	Good
p18	——	——		——	——		2.097	II	Fair	2.028	II	Fair
p19	——	——		——	——		2.122	II	Fair	1.967	II biased toward I	Good
p20	——	——		——	——		2.403	II	Bad	1.968	II biased toward I	Good
Average	1.785	II biased toward I	Good	1.815	II biased toward I	Good	2.151	II	Fair	1.741	II biased toward I	Good

4.4.2 Fuzzy synthetic water quality assessment results

According to the water quality classification and fuzzy assessment results (B*) calculated using MWQ, the assessment results are listed in Table 4-5. Overall, the water quality in the study area was good (class II biased toward I) in the investigations during Oct. 2005, May 2006, and Sept. 2007. The water quality in May 2007 was fair due to the water quality of the two sites were class III and the average water quality was class II. This finding can also be demonstrated based on the single parameter analysis, since the inorganic nitrogen and oil concentrations in May 2007 were the highest among the four investigations.

The MWQ module is capable of generating distribution contour maps of all parameters included in the assessment. In conjunction with the water quality index, the distribution of phosphate, inorganic nitrogen, and oil concentrations were selected to perform comparisons, considering that the phosphate and inorganic nitrogen concentrations were particularly high and significantly affected the water quality in the four investigations, meanwhile the oil concentration is closely related to the production activities of offshore platforms. The contour map of the water quality index in May 2006 is shown in Figure 3-4 as an example to illustrate the system interface. The water quality index ranged from 1.272 to 2.211; the corresponding overall water quality was good (class II biased toward class I). Contour maps of the four selected parameters in 2007 are presented in Figure 4-2 and Figure 4-3. Among these contour maps, the distribution of water quality index indicates that the range of water quality index in May 2007 was 1.425 to 2.594; the overall water quality was fair (class II). Moreover, the overall water quality was good (class II biased toward I) in Sept. 2007 with water quality index in the range of 1.025 ~ 2.116.

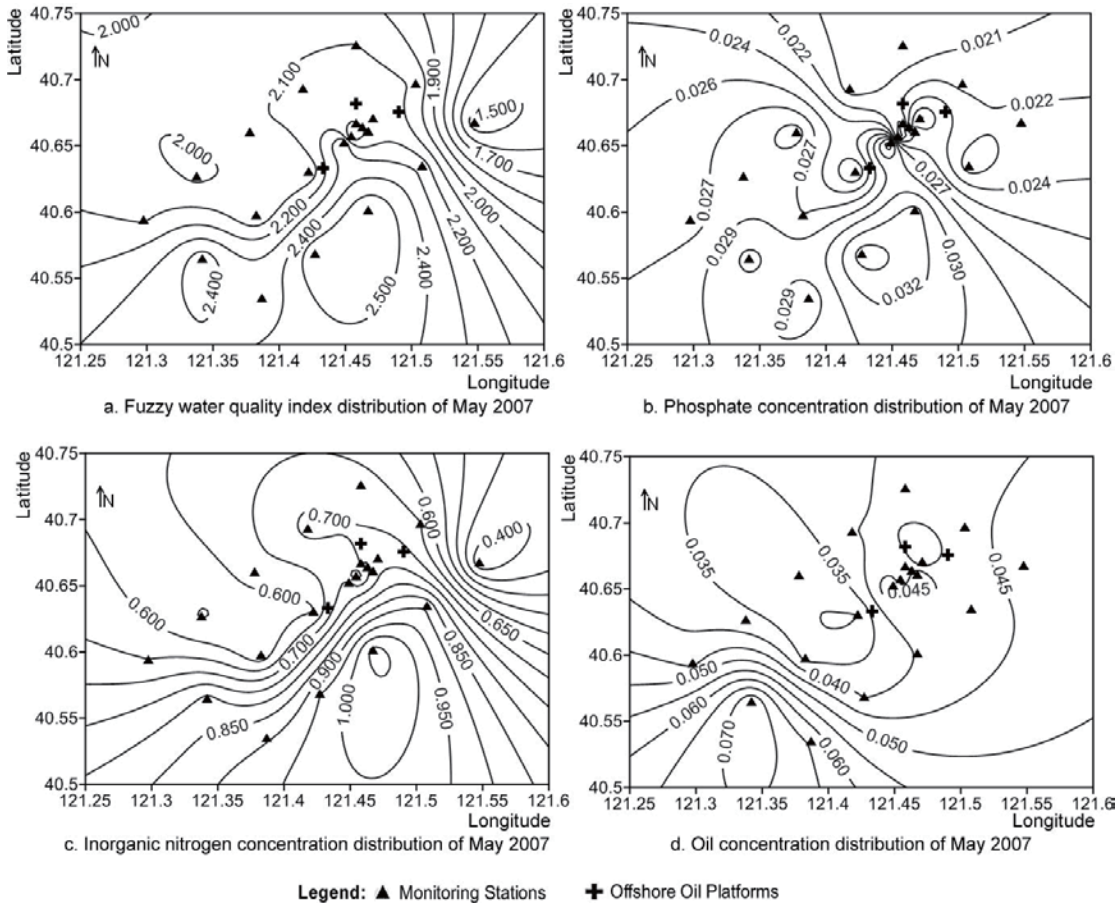


Figure 4-2 Contour map of fuzzy marine water quality index, phosphate, inorganic nitrogen and oil concentration in May 2007

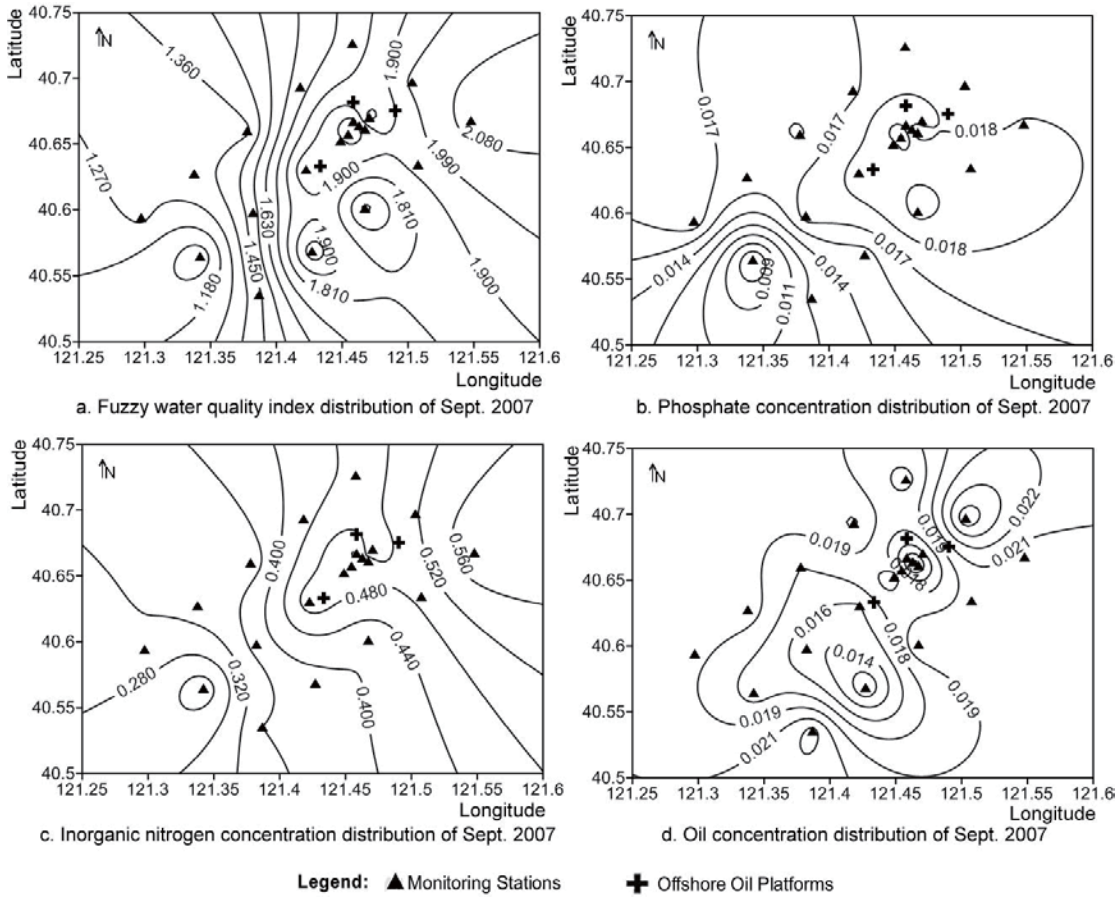


Figure 4-3 Contour map of fuzzy marine water quality index, phosphate, inorganic nitrogen and oil concentration in Sept. 2007

4.4.3 Eutrophication risk and heavy metal risk evaluation

The eutrophication index and heavy metal risk index for the four investigations were calculated using MWQ and the results are provided in Table 4-6. The risk levels of eutrophication and heavy metal are presented as color-coded risk maps (Figure 4-4 and Figure 4-5) , respectively.

According to the risk levels and colors in Table 3-1 and Table 4-3, the ecological risk level of heavy metals in the study area was slight during the four investigations; the eutrophication risk level in Oct. 2005, May 2006 and Sept. 2007 corresponded to

oligotrophic, light eutrophic and mesotrophic conditions, respectively, while the mesotrophic and high eutrophic conditions occurred in May 2007.

Table 4-6 Eutrophication index and heavy metal risk index of the four investigations

	Oct. 2005		May-06		May-07		Sept. 2007	
Station	EI[d]	HMRI[e]	EI	HMRI	EI	HMRI	EI	HMRI
P1	1.976	14.355	3.782	12.457	3.813	12.47	0.993	11.416
P2	2.353	14.666	0.667	13.346	4.534	12.669	1.321	14.14
P3	2.441	12.923	2.498	12.149	4.902	10.901	0.839	11.652
P4	2.318	16.393	1.168	11.278	5.257	8.933	1.397	10.247
P5	0.676	14.584	1.851	13.308	4.448	9.628	1.998	11.012
P6	2.605	13.663	2.171	10.751	5.545	13.127	0.251	10.914
P7	1.616	13.356	1.044	11.216	4.118	10.666	1.471	10.81
P8	0.926	12.171	1.093	10.614	3.598	11.593	2.352	10.828
P9	0.748	13.999	0.942	11.089	6.724	10.683	2.421	9.809
P10	1.464	15.105	2.44	11.305	4.875	11.165	1.627	10.405
P11	2.133	15.711	0.493	12.325	7.148	13.043	0.906	13.695
P12	1.993	20.292	3.013	11.423	7.983	10.249	1.212	12.237
P13	1.113	21.143	0.85	11.64	13.227	11.199	1.806	12.083
P14	——	——	2.433	11.789	7.879	11.925	2.29	10.804
P15	——	——	1.742	11.985	3.387	12.783	1.643	13.088
P16	——	——	0.802	12.689	4.159	11.628	2.024	11.067
P17	——	——	——	——	7.321	10.653	1.656	13.466
P18	——	——	——	——	5.229	10.705	2.348	12.89
P19	——	——	——	——	6.833	9.877	2.149	10.87
P20	——	——	——	——	8.338	11.663	1.979	12.659
Average	1.720	15.259	1.687	11.835	5.966	11.278	1.634	11.705

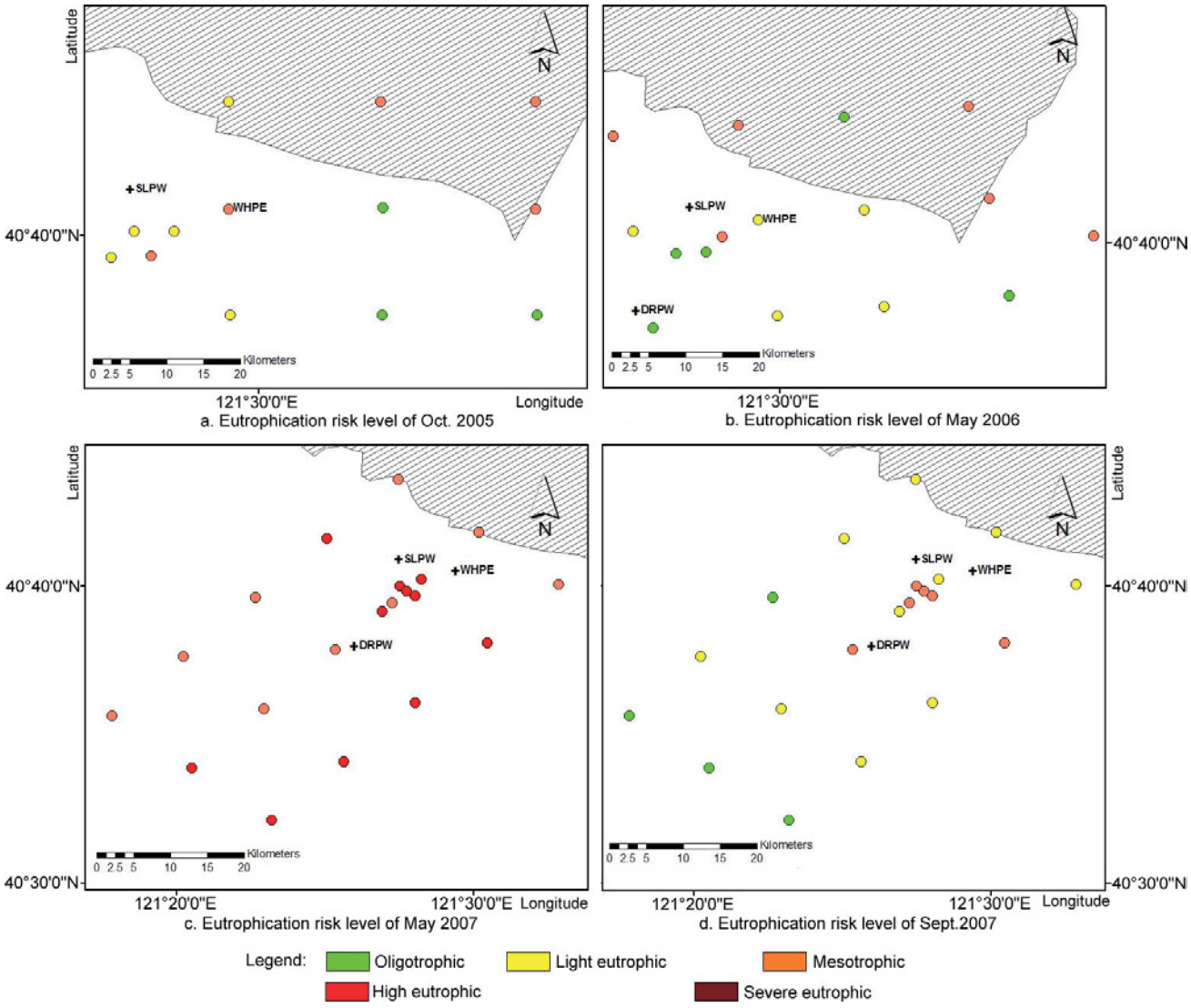


Figure 4-4 Eutrophication risk level of the four investigations

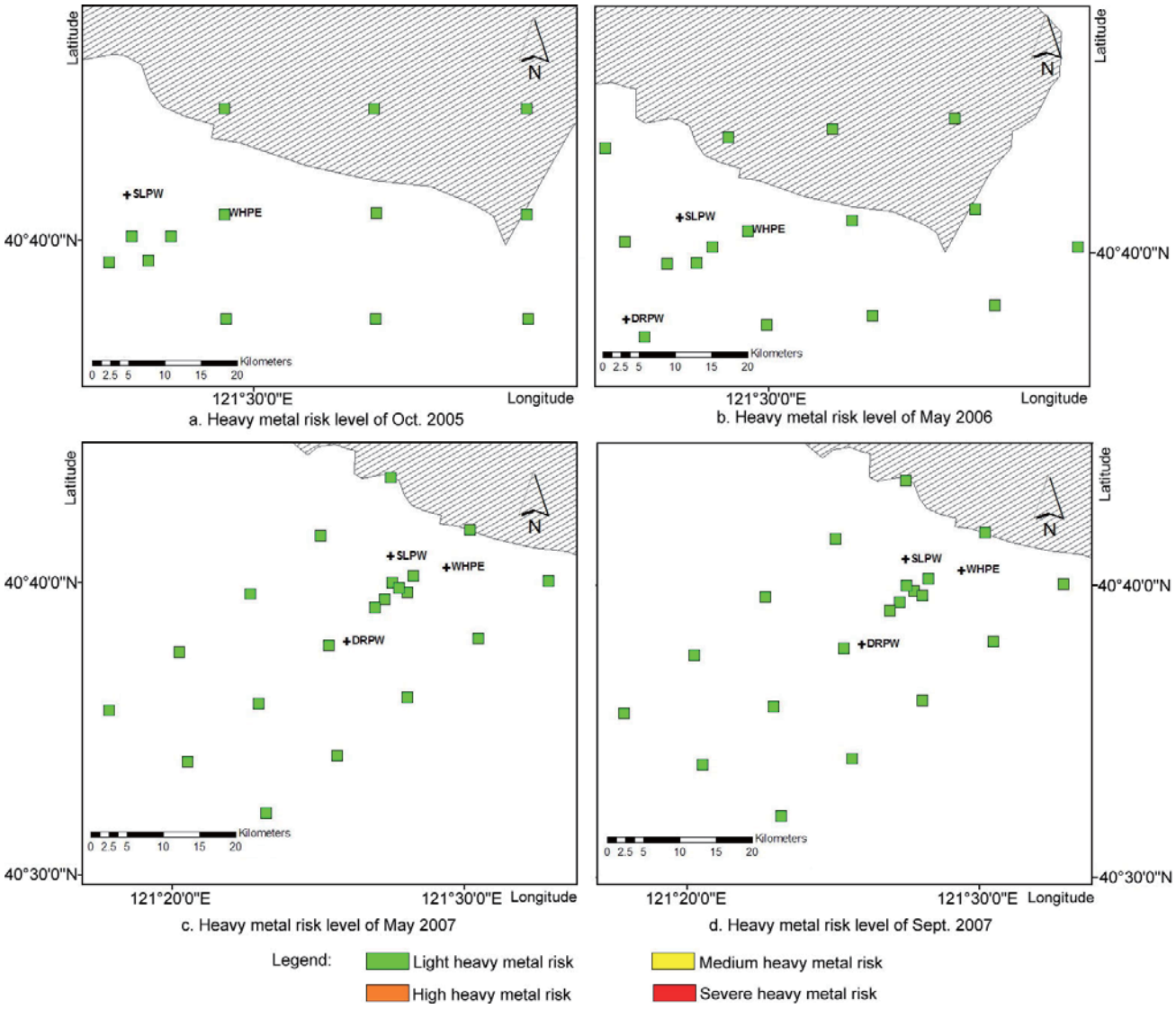


Figure 4-5 Heavy metal risk level of the four investigations

4.5 Discussion

4.5.1 Comparison with results in previous literature

Previous studies have suggested that domestic sewage and industrial wastewater are the main contaminant sources in Liaodong Bay (Fang and Mu 2001; Li, Feng et al. 2001; Wang, Zhang et al. 2001). High concentrations of nutrients were presented in this study, and similar results were noted in previous studies and in coastal environmental reports, indicating that the water quality in the northern portion of Liaodong Bay is primarily affected by inorganic nitrogen and phosphate (Li et al., 2001; Wan et al., 2008). Moreover, the main pollutants in the wastewater discharged into coastal water include large amounts of nutrients (SOA, 2007).

The fuzzy risk assessment results showed that the overall water quality in the study area was class II biased toward class I. Wan (2008) assessed the water quality in the north portion of Liaodong Bay in 2004 using fuzzy mathematics and indicated that the water quality in this region was generally class I with some sites near the estuaries exhibiting worse water quality (class II). The water quality determined in the current study was worse than the values noted in previous studies. This difference is probably due to the following two reasons. First, the weighted average principle method was applied in this research to replace the maximum principle method that was used in Wan's study. Because the nutrient concentrations were particularly high in the study area, the results of the weighted average principle method correspond more closely to the actual situations. Second, the data in Wan's study were collected few years ahead of the current study and the water quality probably declined in this area.

4.5.2 Single parameter analysis

The contour maps generated from the MWQ system in Figure 4-2 and Figure 4-3 indicated close relevance between the fuzzy water quality index and the concentrations of the three selected pollutants in May and Sept. 2007. The contour maps of the fuzzy water quality index reflected the synthetic effects of these parameters on the ambient water quality, i.e., the distribution of the areas with high concentration were similar to the distribution of high water quality index values, which correspond to worse water quality. This finding supports the foregoing analysis that the water quality in the study area is primarily affected by nutrients. Chemical analysis showed that water-based drilling mud also contains phosphate (Zhang 2003). Thus, the leaking of drilling mud related to the operation of the oil platforms may have partially affected the water quality in the surrounding area. The effect of oil exploitation on water quality was reflected by the elevated concentrations of oil distributed near the platforms. The results of other investigations are analogous to the situation discussed above for the 2007 surveys.

According to studies regarding the current in the northern part of Liaodong Bay (Hu, 2007), the mean current in the study area exhibits a SE to NW flow in summer and a NW to SE flow in winter. According to the contour maps, the parameter distributions did not exactly follow the same trend due to the complex effects of shorelines and offshore facilities on the current in the study area. Consequently, the transport of pollutants will also be affected.

4.5.3 Eutrophication risk and heavy metal risk in mariculture zone

The eutrophication risk level in May 2007 was significantly higher than during the other three investigations, agreeing well with the fuzzy water quality index and nutrient concentrations determined in this investigation.

According to Figure 4-1, some sampling sites were located within the mariculture zone near the study area. The eutrophication risk level and heavy metal risk level at these sites are displayed in the color-coded maps generated by MWQ in Figure 4-4 and Figure 4-5. There was only a slight heavy metal risk in the mariculture zone during the four investigations. The eutrophication risk level for most of the sites located within the mariculture zone was mesotrophic in Oct. 2005, May 2006 and May 2007; light eutrophic conditions prevailed in Sept. 2007. It is well accepted that the probability of having toxic algal blooms and red tides is much higher in eutrophic waters. Blooming and the ultimate collapse of algae blooms may lead to hypoxia or anoxia and cause mass mortality of benthos and fish over large areas. There have been several reports of fishery losses arising from red tides and toxic blooms (Feng et al., 2004; Wu, 1999). Thus, management personnel should pay close attention to the nutrient discharges in this area and take effective measures to reduce the eutrophication risk in the mariculture zone.

4.5.4 Comparison between fuzzy synthetic evaluation and standard index evaluation

The standard index evaluation method is commonly used in surface water quality assessments in China. This method can be represented by the following equation:

$$I_{\mathrm{i}} = C_i / S_i \tag{4.1}$$

where I_{i} is the standard index of each parameter, C_i is the measured concentration of each parameter, and S_i is the evaluation criteria of each parameter.

For the dissolved oxygen evaluation,

$$\begin{aligned} I_i(DO) &= |DO_f - DO| / (DO_f - DO_s) && DO \geq DO_s \\ I_i(DO) &= 10 - 9DO / DO_s && DO < DO_s \end{aligned} \tag{4.2}$$

where DO_f is the saturated concentration of dissolved oxygen in water and DO_s is the dissolved oxygen criteria. Moreover,

$$DO_f = 468/(31.6 + t) \tag{4.3}$$

where t is the water temperature in °C.

The standard indices of all parameters included in the analysis are synthesized to obtain the water quality index using the following equation:

$$WQI = \sqrt{\frac{I^2_{\max} + I^2_j}{2}} \tag{4.4}$$

where $I_{\max}$ and I_j are the maximum and the average values of all standard indices, respectively.

The results of the fuzzy synthetic evaluation and the standard index evaluation are provided in Figure 4-6.

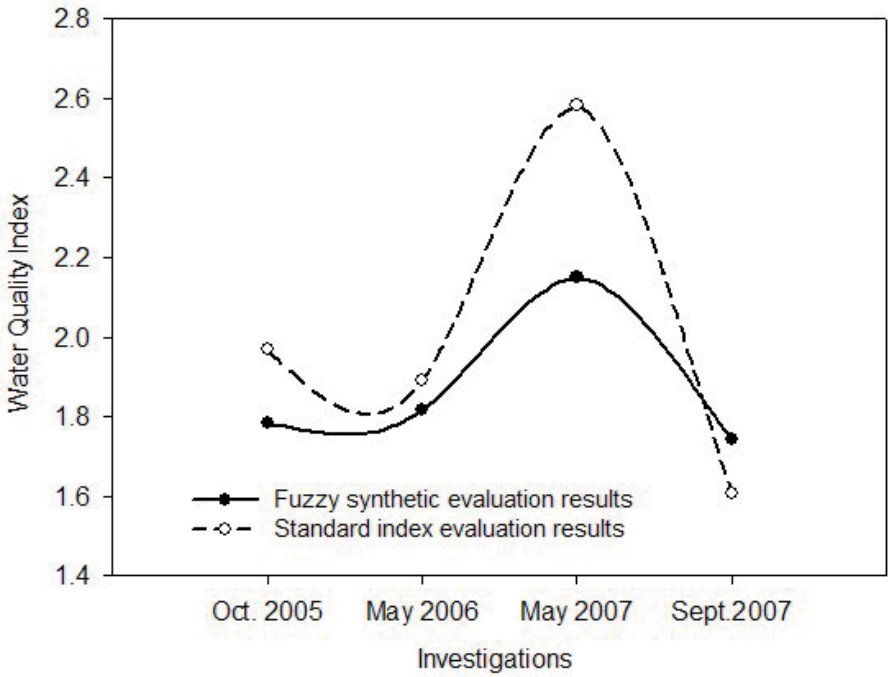


Figure 4-6 Comparisons between the results of the fuzzy synthetic evaluation and standard index evaluation methods

According to Figure 4-6, the results of the two methods were comparable. Moreover, the water quality variation trend over the studied period was the same, indicating that the results obtained from the developed system were accurate. It was also evident that most of the fuzzy water quality indexes were smaller than suggested by the standard index method. This difference was probably due to the fact that the standard index method primarily focused on the parameters with severely high concentrations that exceeded the criteria, which demonstrated the decisive effect of these parameters in the water quality assessment. While the fuzzy synthetic evaluation method sufficiently considered the contribution of each parameter in the assessment process, the contributions were assigned on the basis of the individual weight of each parameter. This can provide a better understanding of the water pollution resulting from multiple factors.

4.5.5 Other uncertainties

Other uncertainties may be associated with the following factors: (i) the uncertainty in the platform and/or land-based discharges, which make it challenging to identify the precise correlation between the water quality variations and offshore oil production, and (ii) the four surveys that were performed in different months and at a distinct interval may not be able to accurately constrain temporal water quality variations. Continuous monitoring is needed in the future to obtain sufficient data to support related studies regarding the water quality in this area. Moreover, the developed MWQ system can be further customized for automatic data processing and for verifying the results.

4.6 Summary

The developed MWQ module of the system was applied to assess the water quality and ecological risk for the area around offshore oil platforms in the northern Liaodong Bay of China. The assessment results were displayed in the forms of contour maps and color-coded maps which intuitively present the spatial distribution of the monitoring and assessment results. Numerical results were also provided in the database and can be exported for further calculation. The developed MWQ module facilitates the tedious calculation of the risk assessments, and makes the assessment process easier.

CHAPTER 5 LAKE WATER QUALITY ASSESSMENT USING THE GIS-SWQAM (LWQ MODULE) -- A Case Study of Lake Champlain

5.1 Study Area

Lake Champlain is one of the numerous large lakes located mainly within the United States (states of Vermont and New York) and partially situated in the province of Quebec, Canada (Figure 5-1). Lake Champlain has a complex bathymetry, with over 70 islands and a shoreline of over 800 km. The acreage of Lake Champlain is approximately 1,269 km^2 in area, the lake is roughly 201 km long, and 0.8 to 23 km wide (The Canadian Encyclopedia). The maximum depth is approximately 120 m. The drainage area of Lake Champlain includes the area between the Adirondack Mountains in northeastern New York State and the Green Mountains in Vermont. The drainage area covers about 21,326 km^2 (The Lake Champlain Basin, 2009). The excessive phosphorus loading in Lake Champlain stimulated noxious algae blooms, which disrupt the lake's ecology and degrade the lake water quality (Ghebremichael et al., 2010).

The lake is divided into five major lake basins created by natural and anthropogenic barriers, including the South Lake, the Main Lake, the Northeast Arm, Malletts Bay, and Missisquoi Bay. The South Lake is a shallow riverine basin, which is narrow and warm. Water flows northward into the Main Lake. The Main Lake includes the deepest portion of the lake and contains over 80% of the total lake volume and almost 60% of the lake surface area. The main lake extends northward to the Canadian border and drains via the Richelieu River into the St. Lawrence River. Malletts Bay is moderately deep. Water from Malletts Bay flows west into Main Lake and north into the Northeast Arm, which is moderately deep and partially isolated from Malletts Bay by a shallow sandbar arising from the sediments deposited by the

Lamoille River. Missisquoi Bay is the northernmost basin of Lake Champlain, which is a shallow and warm basin (Facey et al., 2012).

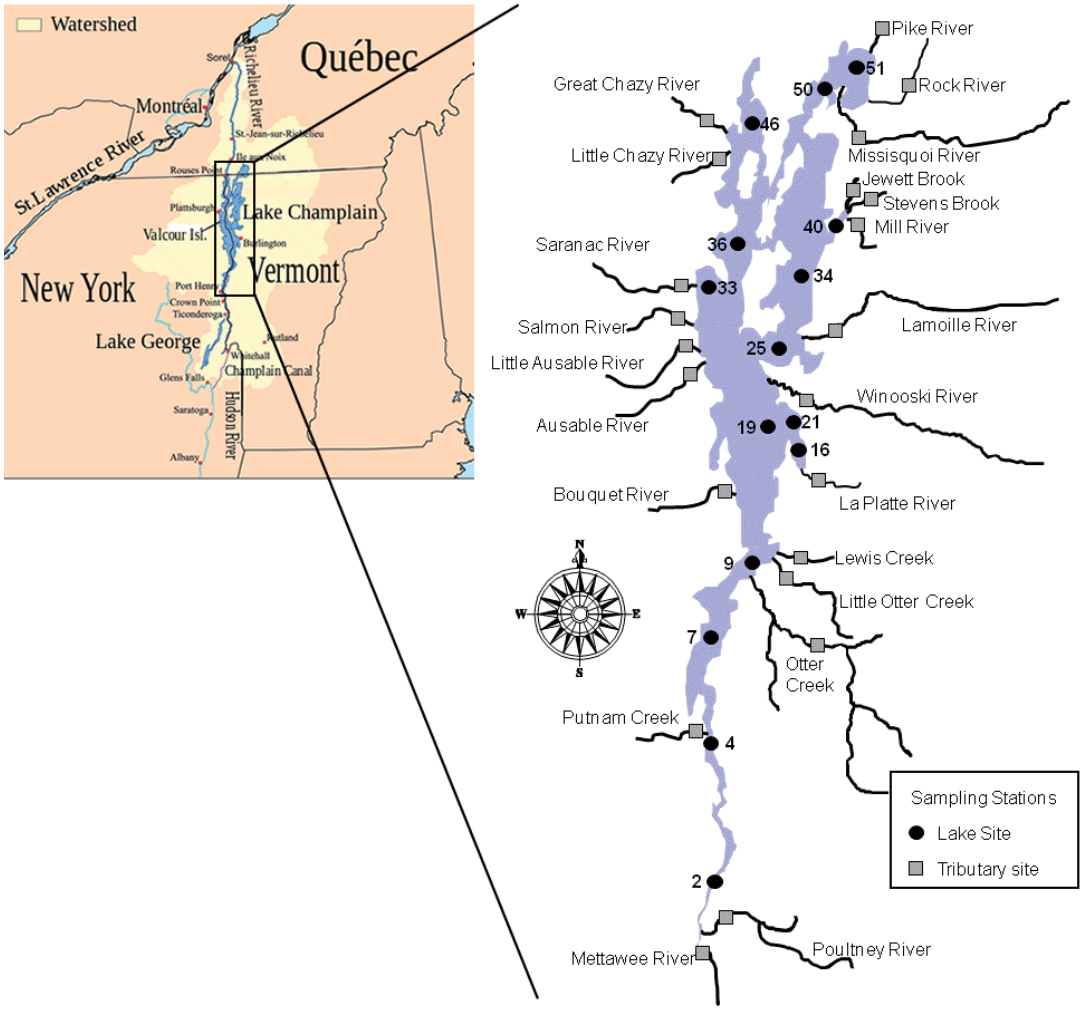


Figure 5-1 Study area and sampling sites (Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation, 2012)

5.2 Lake Water Chemistry Data Collection

The lake water chemistry data used in this thesis were obtained from the Champlain Lake long-term water quality and biological monitoring project database (Champlain Lake long-term water quality and biological monitoring project). Details on sampling and analytical procedures are available in the "Lake Champlain Long-

Term Water Quality and Biological Monitoring Program Description" (Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation, 2012). The water samples were collected at 15 lake stations. Except the stations 9, 16 and 51 that were added after 2001, other stations have been sampled consistently during the entire monitoring period from 1992 to 2012. The location of these sites is shown in Figure 5-1 and listed in Table 5-1.

Table 5-1 List of lake sampling stations and their locations

Lake Station	Station Name	Latitude N	Longitude W
2	South Lake B	43° 42.89'	73° 22.98'
4	South Lake A	43° 57.10'	73° 24.47'
7	Port Henry	44° 07.56'	73° 24.77'
9[1]	Otter Creek	44° 14.53'	73° 19.75'
16[1]	Shelburne Bay	44° 25.55'	73° 13.92'
19	Main Lake	44° 28.26'	73° 17.95'
21	Burlington Bay	44° 28.49'	73° 13.90'
25	Malletts Bay	44° 34.92'	73° 16.87'
33	Cumberland Bay	44° 42.07'	73° 25.09'
34	Northeast Arm	44° 42.49'	73° 13.61'
36	Isle La Motte (off Grand Isle)	44° 45.37'	73° 21.30'
40	Saint Albans Bay	44° 47.12'	73° 09.73'
46	Isle La Motte (off Rouse's Pt.)	44° 56.90'	73° 20.40'
50	Missisquoi Bay	45° 00.80'	73° 10.43'
51[2]	Missisquoi Bay (Central)	45° 02.50'	73° 07.78'
	Notes: [1] Added in 2001. [2] Added in 2006.		

The lake stations were sampled approximately biweekly from May to early November each year. (Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation, 2012). Only data from the epilimnion layer is included in this thesis.

5.3 Water Quality Assessment of Lake Champlain

In this section, the water quality and ecological risk in Lake Champlain were evaluated through the developed LWQ module.

5.3.1 System implementation – LWQ module

Since most parts of Lake Champlain are located in the United States, the criteria for lake water quality of New York and Vermont were applied in this study. The document “Ambient Water Quality Standards and Guidance Values and Groundwater Effluent Limitations” was used as a reference to determine the maximum allowable values of temperature and chloride concentration (John Zambrano and Stoner, 1998). The maximum values of total phosphorus and dissolved oxygen, as well as the allowable range of pH, were determined according to the Vermont water quality standards (Department of Environmental Conservation, 2014). Ranges for fuzzy sets were obtained by dividing the standard values into four intervals (Ocampo-Duque et al., 2006, Icaga, 2007). Temperature and pH were exceptions and the intervals were determined with reference to the fuzzy evaluation of lake water quality classification in Turkey (Icaga, 2007). These ranges are shown in Table 5-2.

Lake Champlain was divided into 13 phosphorus management segments (Figure 5-2) (Lake Champlain Basin Program-nutrients, Smeltzer et al., 2009). According to the Vermont water quality standards, the standard values of total phosphorus varies between lake segments (Lake Champlain Basin Program-nutrients, Department of Environmental Conservation, 2014). The lake water quality monitoring stations were consequently grouped into each segment as shown in Table 5-2.

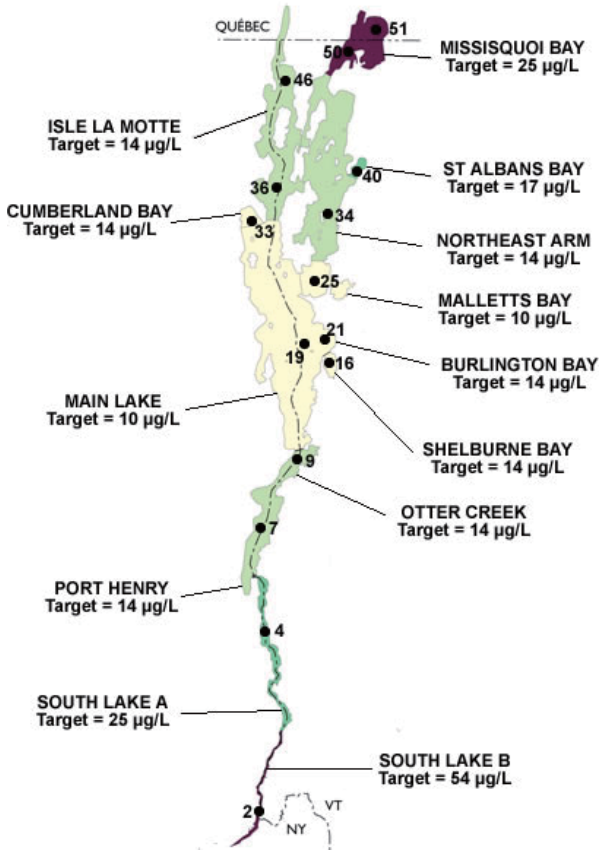


Figure 5-2 Lake Champlain phosphorus management segments and sampling stations (Lake Champlain Basin Program-nutrients, Smeltzer and Quinn, 1996)

Observations used in this part comprise five water quality parameters, including temperature (T, ℃), pH, dissolved oxygen (DO, mg/L), chloride (Cl, mg/L), and total phosphorus (TP, µg/L). Monthly average values of these parameters were calculated based on data retrieved from the database.

Since the standard values of total phosphorus vary between lake segments, the fuzzy water quality index was calculated separately with a different limit concentration of total phosphorus. The range of the fuzzy water quality index is 0 to 4. The corresponding water quality is: good (0-1), fair (1-2), poor (2-3), bad (3-4).

Table 5-2 Limits of membership functions

		Parameters of the membership functions			
Indicator	Units	a	b	c	d
DO	mg/L	7	5	4	3
pH ≥ 7.5	--	7.5	7.75	8.75	9.25
pH < 7.5	--	7.5	6.75	6.25	5.75
Cl	mg/L	62.5	125	187.5	250
Temperature	Degree C	15	20	25	30
TP	µg/L				
Segments	Stations	a	b	c	d
Main Lake	19	2.5	5	7.5	10
Cumberland Bay	33	3.5	7	10.5	14
Malletts Bay	25	2.5	5	7.5	10
Burlington Bay	21	3.5	7	10.5	14
Shelburne Bay	16	3.5	7	10.5	14
Northeast Arm	34	3.5	7	10.5	14
Isle LaMotte	36, 46	3.5	7	10.5	14
Otter Creek	9	3.5	7	10.5	14
Port Henry	7	3.5	7	10.5	14
St. Albans Bay	40	4.25	8.5	12.75	17
Missisquoi Bay	50, 51	6.25	12.5	18.75	25
South Lake A	4	6.25	12.5	18.75	25
South Lake B	2	13.5	27	40.5	54
Fuzzy Water Quality Index		1	2	3	4
		Good	Fair	Poor	Bad

The LWQ module can be applied to the data collected in each survey to obtain the water quality assessment and eutrophication risk results according to the steps introduced in Section 3.5.2.2.

5.3.2 Results and discussion

The results obtained through the LWQ module using monthly average data from May to September in 2012 will be shown and discussed herein.

Comparison between the monitoring values and criteria are shown in Figure 5-3 to Figure 5-7 to present the weights and ranges of each parameter in the water quality assessment. As can be seen from the following figures, the concentrations of chloride

were far lower than the maximum allowable value at all the lake stations, thus the weight of chloride in the assessment was quite small. The dissolved oxygen concentrations were all above the minimum required level for cold fish, which is 5 mg/L in the Vermont water quality standard (Department of Environmental Conservation, 2014). The pH values fell into the allowable range at all the stations, which is 6.5-8.5 as required in in the Vermont water quality standard. All the temperature values were below the maximum value according to the related water quality guidance document in New York State (John Zambrano and Stoner, 1998). As for total phosphorus concentrations, the monitor values were close to or above the maximum allowable concentrations, indicating that the total phosphorus is a major pollutant in Lake Champlain and has been attracting more attention in the lake water quality management. Consequently, the TP concentration has the greatest weight in the assessment.

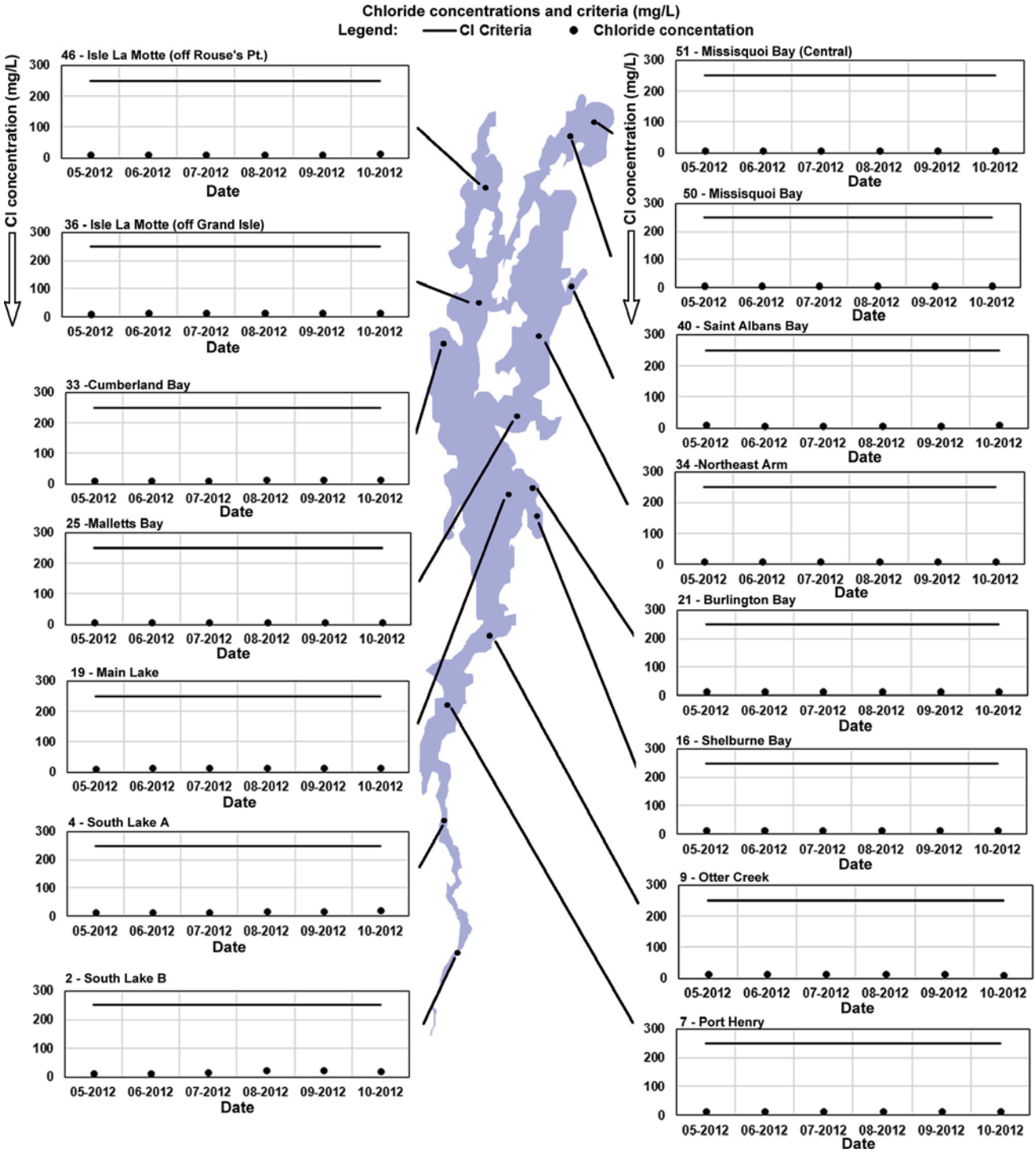


Figure 5-3 Comparison between chloride concentrations in 2012 and criteria

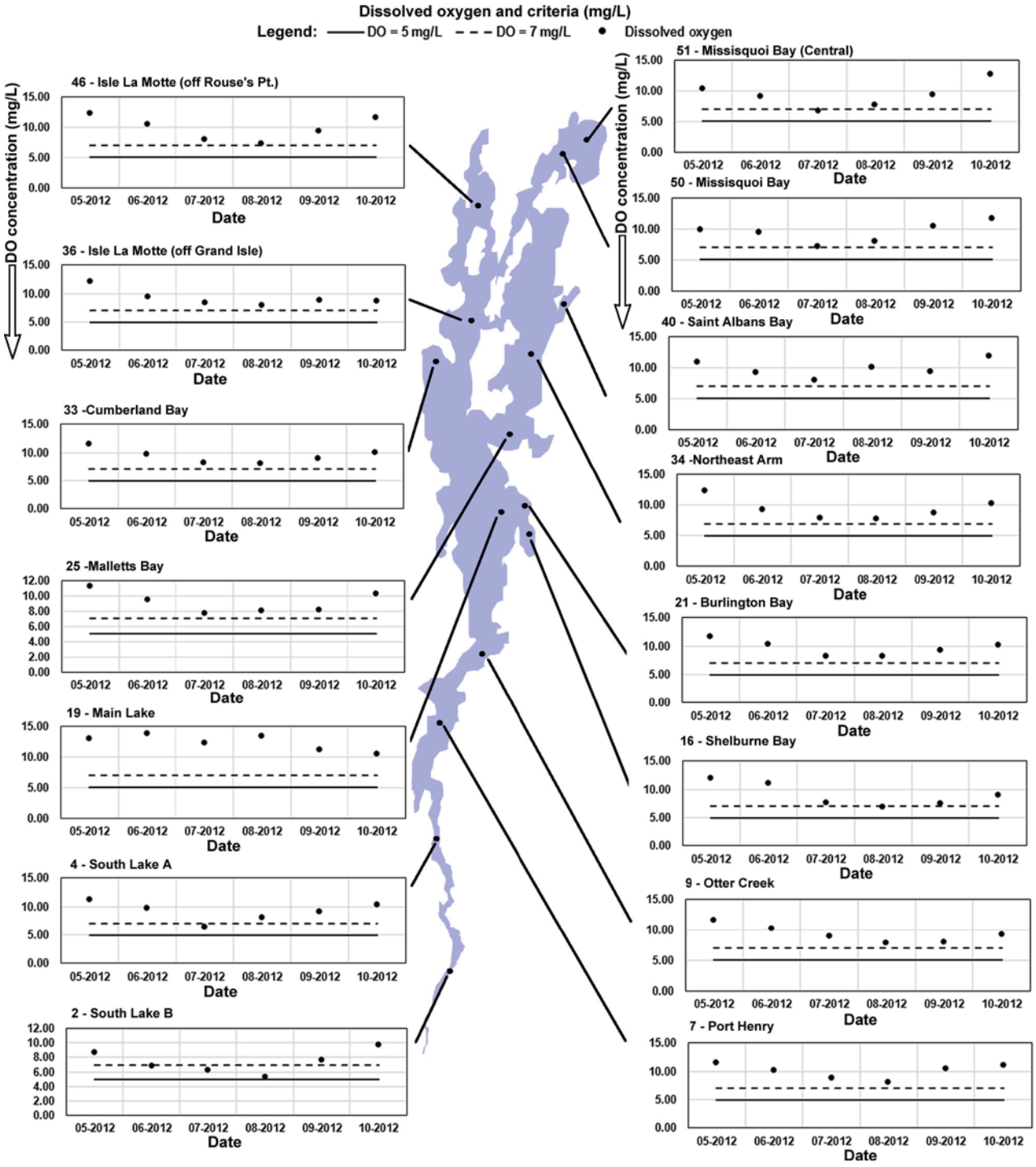


Figure 5-4 Comparison between dissolved oxygen concentrations in 2012 and criteria

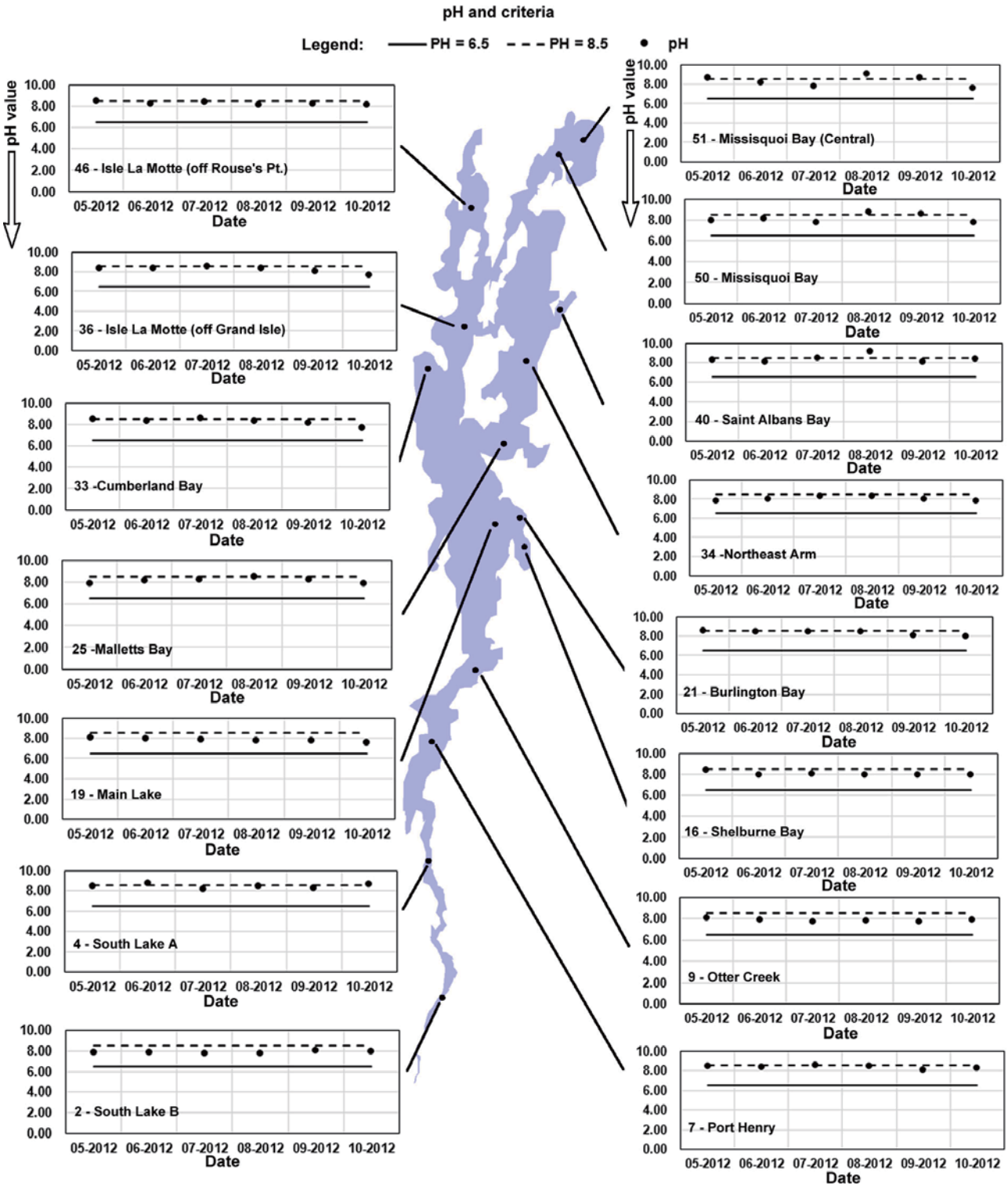


Figure 5-5 Comparison between pH in 2012 and criteria

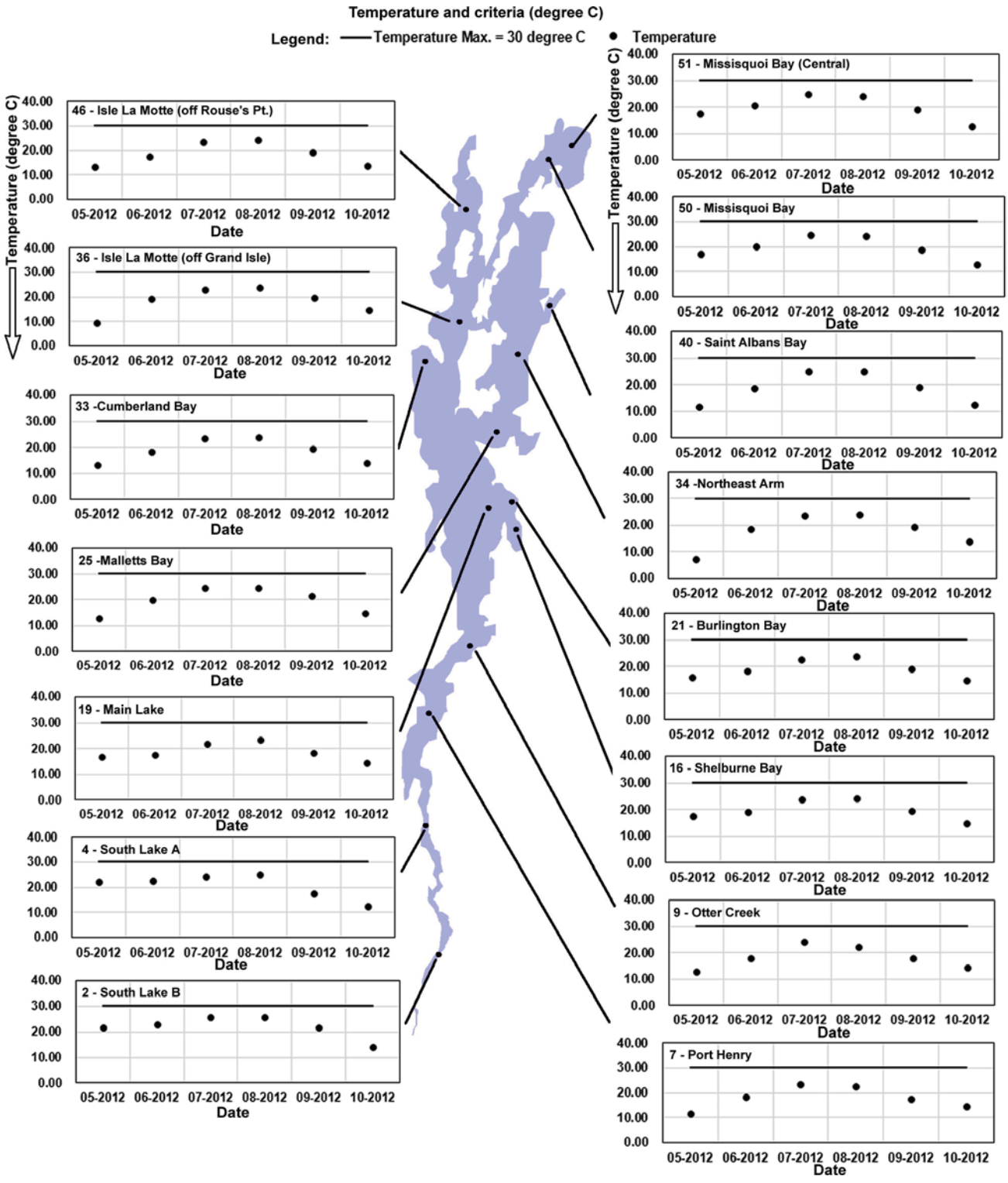


Figure 5-6 Comparison between temperature in 2012 and criteria

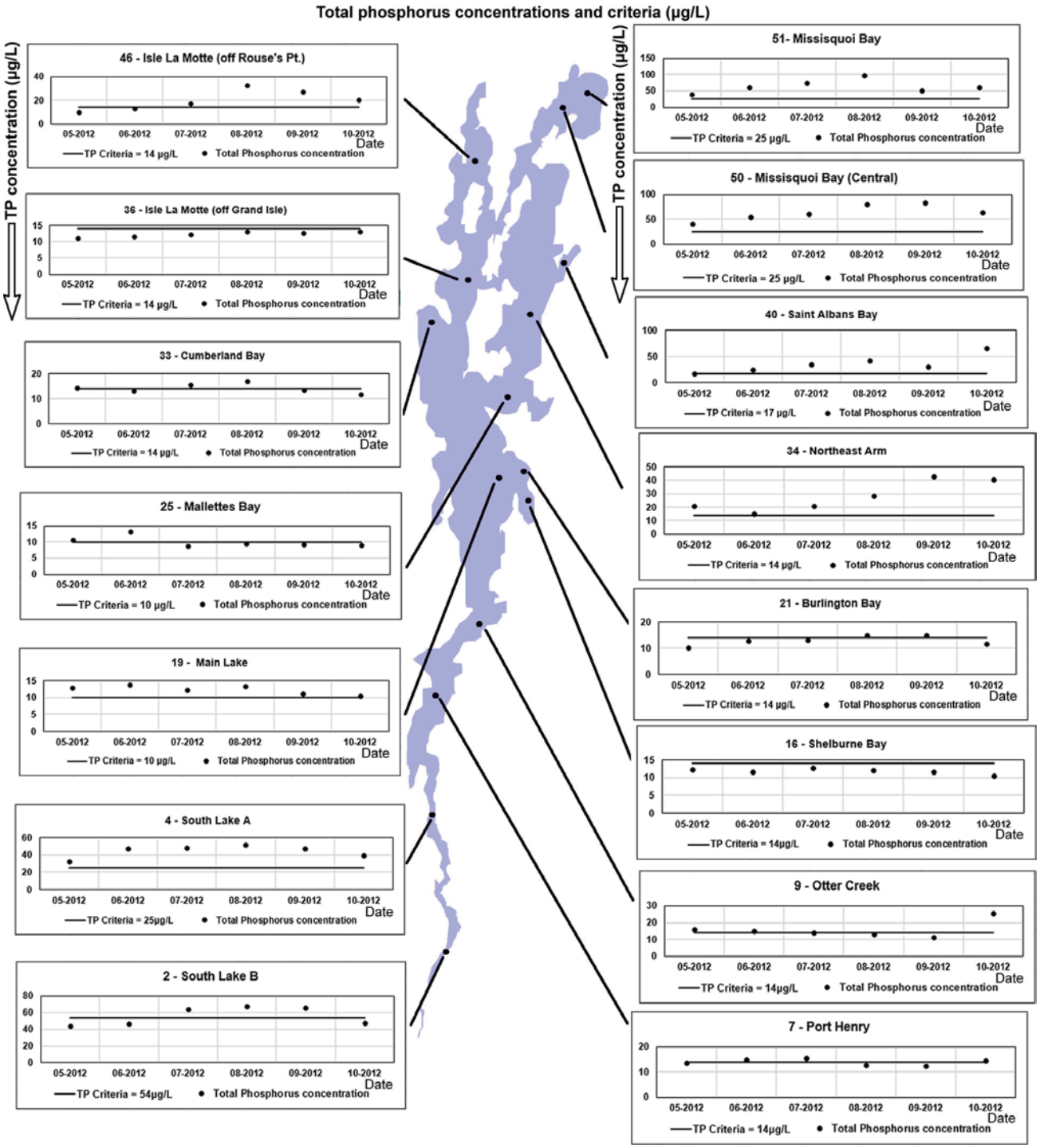


Figure 5-7 Comparison between total phosphorus concentrations and criteria in 2012

5.3.2.1 Fuzzy Synthetic Water Quality Assessment Results

By inputting monitoring data through the graphical user interface of the LWQ module, the fuzzy synthetic water quality index and water quality classification can be obtained according to the water quality classification and fuzzy assessment results (B*). The results were subsequently saved into the database and exported to an excel file for further reorganization.

The temporal variations of assessment results during May to October 2012 were shown in Figure 5-8, in which could be observed that the fuzzy water quality index at the fifteen lake stations in 2012 mostly fell into the interval of "Poor" classification; few of the results were in the range of "Fair" or even "Bad" categories.

Since the concentrations of chloride were quite low compared with the maximum allowable concentration, its weight was small and could be ignored in the synthetic water quality assessment. The following discussion will focus on pH, temperature, DO, total phosphorus, and the fuzzy synthetic water quality index.

The monitoring data and assessment results of May 2012 were used as an example for the spatial distribution maps generated by the LWQ module of the developed system. Select "2012-05" in the date combox and "TP" in the parameter combox; the interval of total phosphorus distribution contour map was set at 0.7. Click the "Create Contour Map" button to generate the spatial distribution map, which has two types: contour lines and filled contour maps (Figure 5-9). Contour maps for pH, temperature, dissolved oxygen and other parameters could be generated in the same manner. To intuitively present the spatial variation of the concentrations, the filled contour maps were utilized in the following analysis.

According to the temporal variation of fuzzy water quality index shown in Figure 5-8, the fuzzy water quality index was relatively higher in July to September than in other months due to higher surface water temperature, which resulted in a lower concentration of dissolved oxygen. Therefore, fuzzy water quality index values were higher under the influence of temperature and dissolved oxygen concentration.

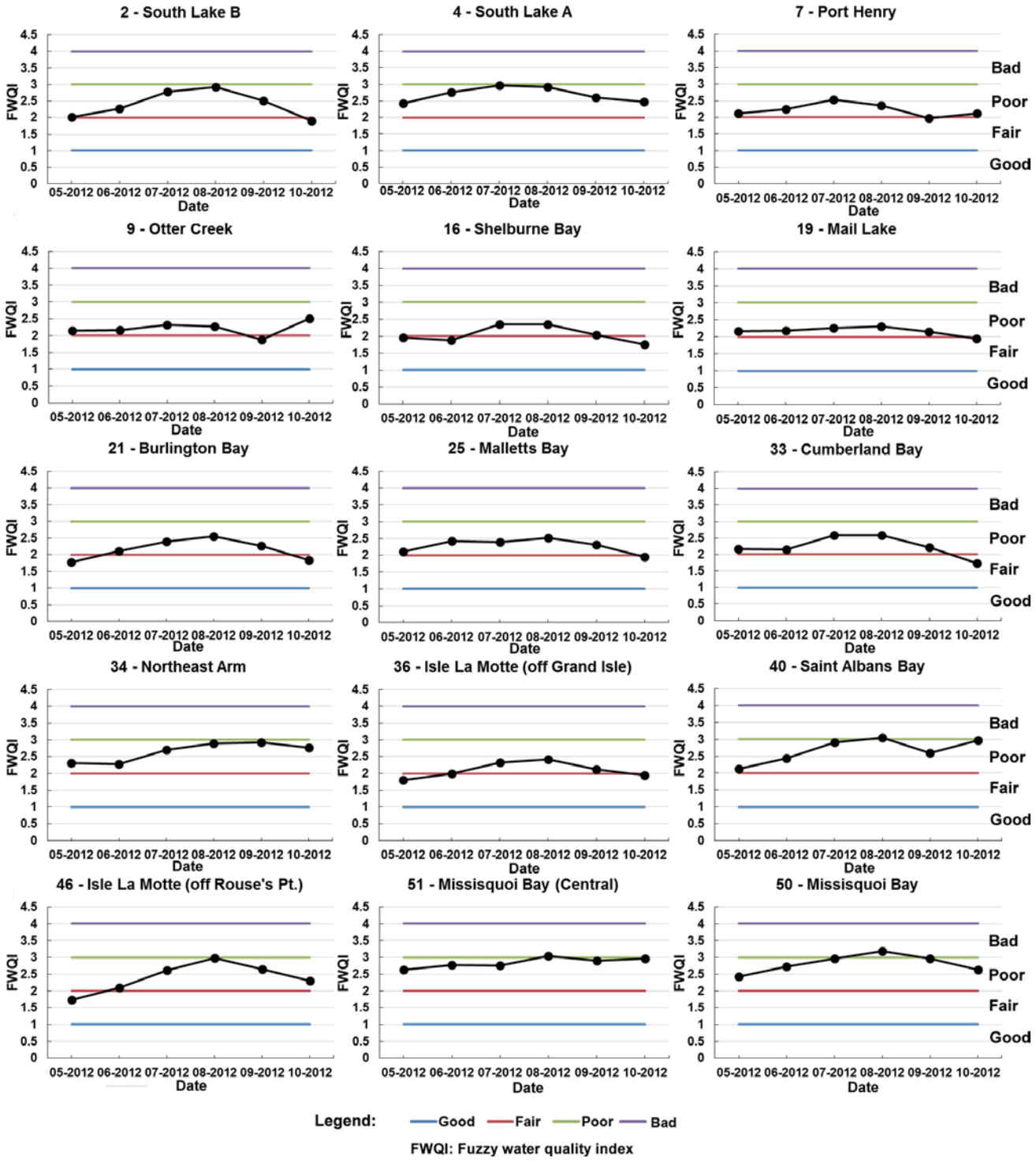


Figure 5-8 Fuzzy synthetic assessment results of May to September in 2012 for all the stations

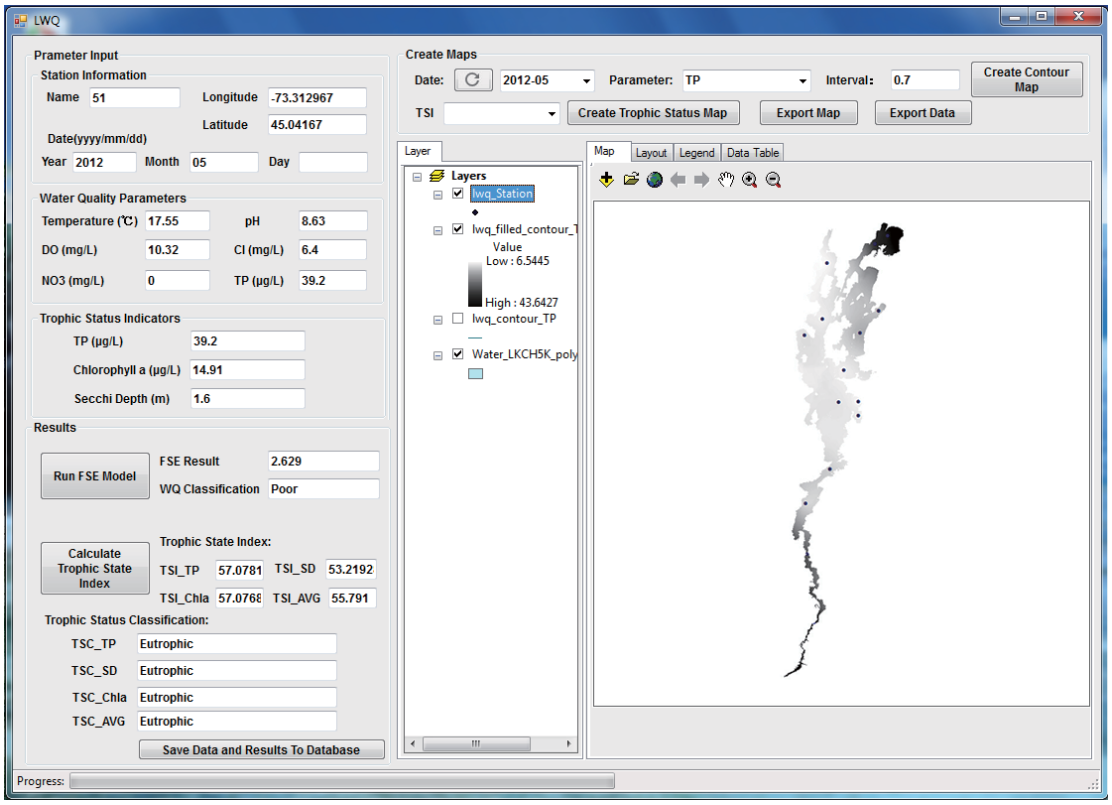


Figure 5-9 Interface of the settings and generated contour map of total phosphorus concentration in May 2012

The generated maps of water quality parameters in May 2012 were exported to images and shown in Figure 5-10. The temperature distribution map shows relative high surface water temperature in the South Lake. The surface water in Main Lake, Burlington Bay, Shelburne Bay, and Missisquoi Bay were also warmer than the water in the Isle La Motte and the Northeast Arm regions. The highest value of dissolved oxygen concentration appeared in the Main Lake region. South Lake and Missisquoi Bay had the lowest DO concentration. The range of pH values over the lake area was 7.66 to 8.82 and higher values showed up in Isle La Motte, Malletts Bay and Missisquoi Bay. Total phosphorus concentration fluctuated greatly over the lake and the high concentration appeared in South Lake, Missisquoi Bay and Northeast Arm.

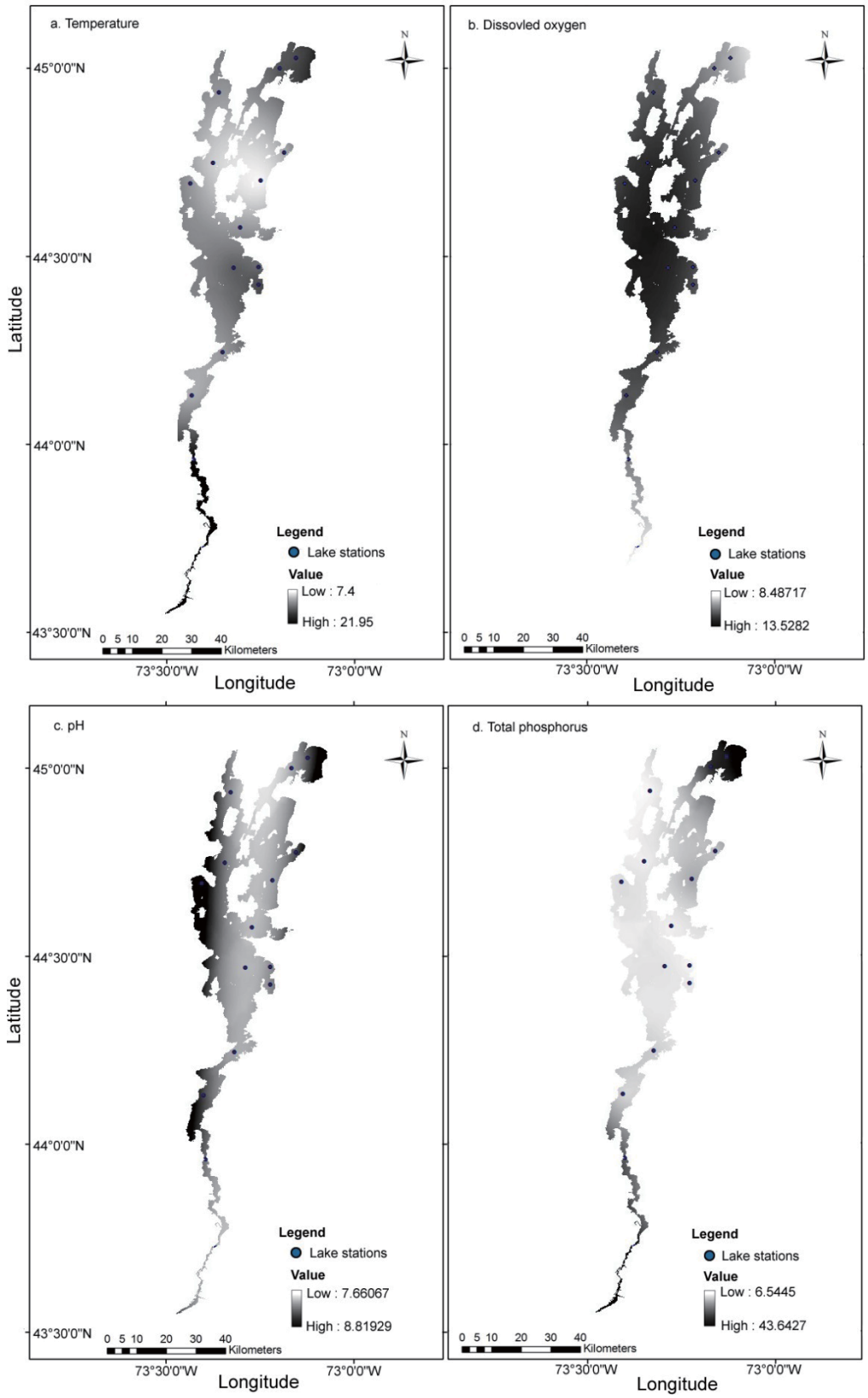


Figure 5-10 Spatial distribution map of (a) temperature, (b) dissolved oxygen, (c) pH, and (d) total phosphorus in May 2012

The fuzzy water quality index of the lake stations was interpolated and displayed in the form of maps (Figure 5-11 and Figure 5-12). The spatial distribution pattern was similar during the period of May to July 2012. Higher values of the fuzzy water quality index that represent worse water quality conditions appeared in the Missiquoi Bay, Port Henry, and South Lake regions. Lower values of the fuzzy water quality index representing better water quality were distributed in other lake regions. In August, the water quality was worse in the northern and southern regions than it was in the middle of the lake. In September and October, the Missiquoi Bay, Northeast Arm, and South Lake regions had obviously worse water quality than other lake regions.

In the results of May 2012, obvious associations can be found between the spatial distributions of total phosphorus concentration and the fuzzy water quality index, as shown in Figure 5-11 and 5-12. Higher concentration of TP was distributed in the Missisquoi Bay and South Lake regions, which also showed higher fuzzy water quality index. Among the five parameters included in the water quality assessment, total phosphorus was the only one that exceeded the limitations greatly. Thus, TP concentration has the highest weight in the assessment process and has the most significant effect on the assessment results. The distributions of temperature and DO were also similar to that of the fuzzy water quality index. The higher temperature and lower DO concentration that results in higher fuzzy water quality index was distributed in the Missisquoi Bay and the south part of the lake.

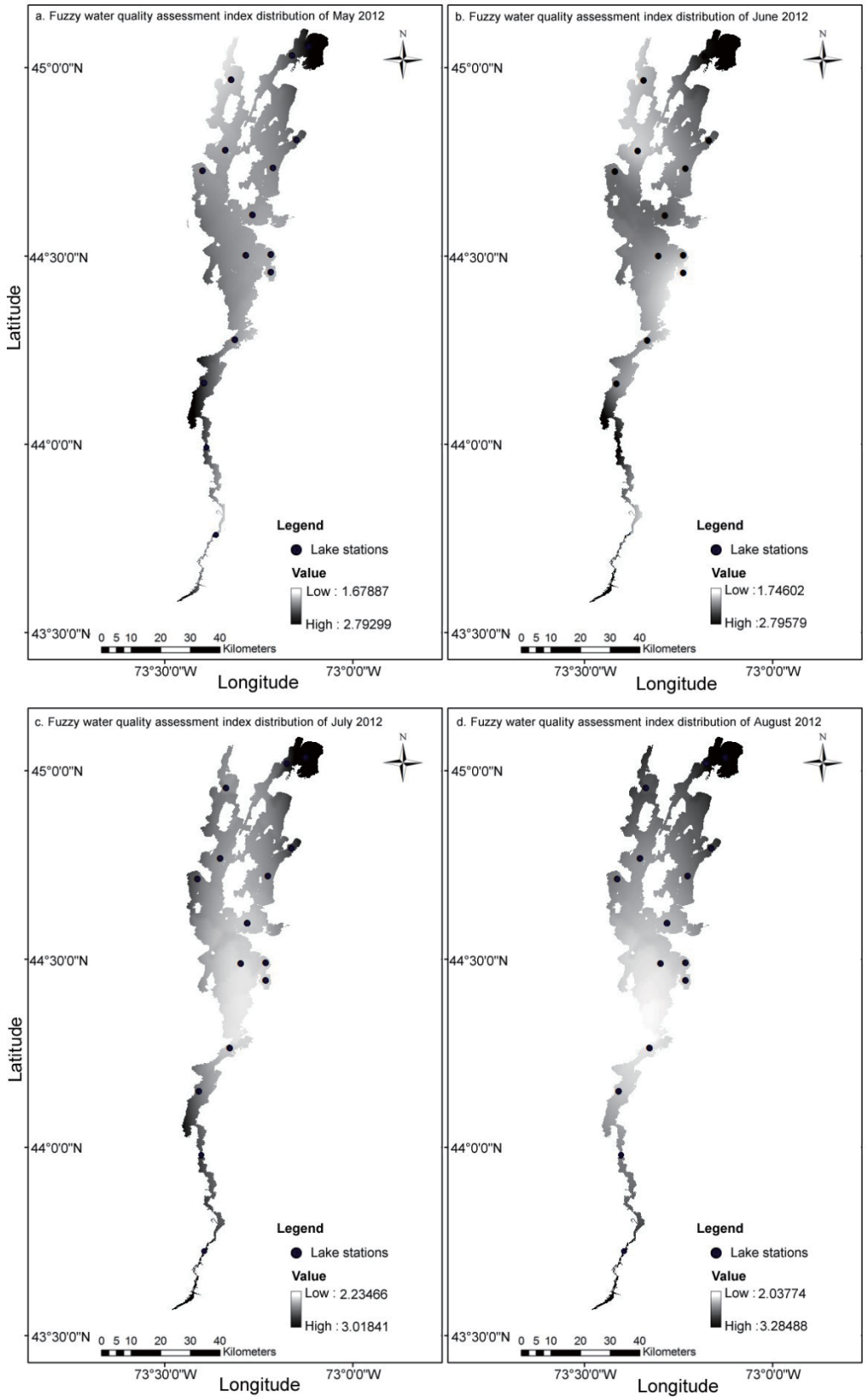


Figure 5-11 Spatial distribution map of fuzzy water quality index in May to August 2012

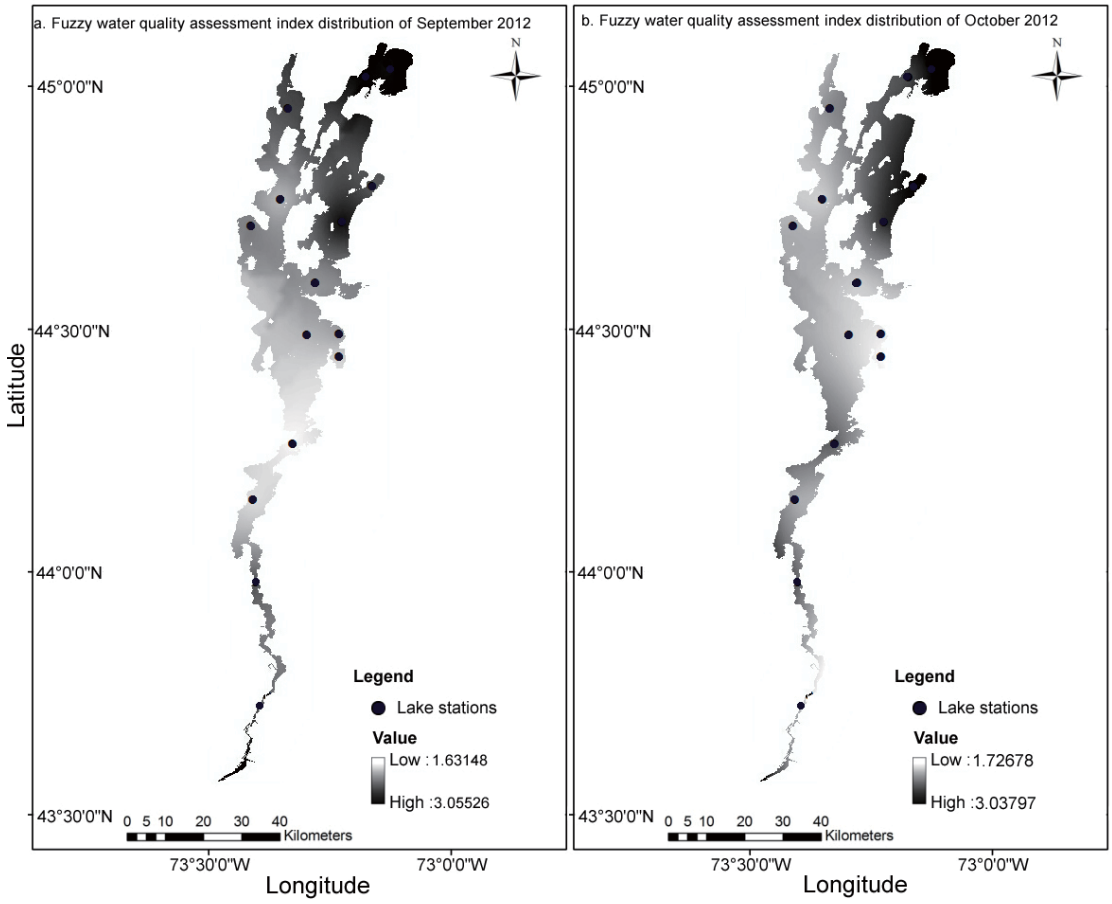


Figure 5-12 Spatial distribution map of fuzzy water quality index in September and October 2012

5.3.2.2 Trophic status evaluation

The temporal variation of trophic state index of the 15 lake stations from May to October in 2012 was calculated using the LWQ module and the results are provided in Figure 5-13 and Figure 5-14. The spatial distribution of the risk levels of eutrophication in 2012 is presented as color-coded risk maps in Figure 5-15 and Figure 5-16. The results indicate that the trophic status determined by the average trophic state index (TSI-AVG) were mostly in the range of [40, 50], which corresponds to mesotrophic, at most of the lake stations during May to October in 2012. The exceptions were the South Lake region (Station 2 and 4) and Missisquoi Bay region (Station 50 and 51), where results for the trophic state index were mostly

in the range of [50, 70] and the trophic statuses were eutrophic, correspondingly. At few stations, the trophic state index fell below 40, which indicated an oligotrophic status.

As seen in Figure 5-15 and Figure 5-16, the eutrophic status distributed in the Missisquoi Bay and South Lake regions over May to October 2012 were all eutrophic. In other lake regions, the trophic status was mostly mesotrophic. Exceptionally, eutrophic condition showed up in St Albans Bay, and five lake regions including the Shelburne Bay, Main Lake, Burlington Bay, Cumberland Bay, and Isle La Motte (off Grand Isle) were under oligotrophic status in August 2012.

The spatial distribution of trophic state index agreed fairly well with the fuzzy water quality index, i.e., the regions with higher trophic state index had higher fuzzy water quality index, and vice versa. This indicated that the eutrophication arose in the lake regions that had worse water quality, which further proved that reducing the loads of TP into Lake Champlain and controlling the TP concentration in lake water are the key factors for improving water quality of Lake Champlain.

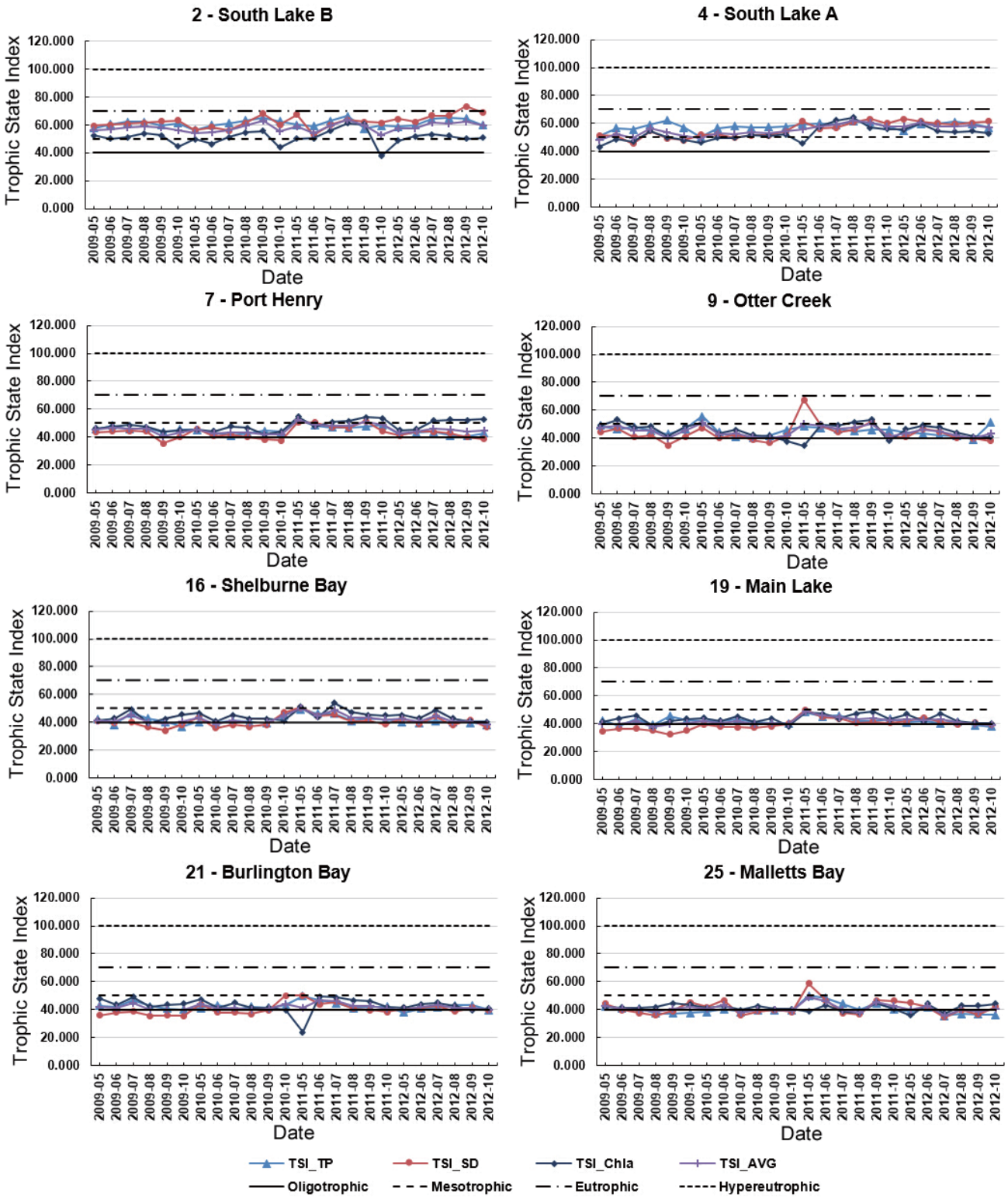


Figure 5-13 Trophic state index of the 15 lake stations from May to October 2012 (I)

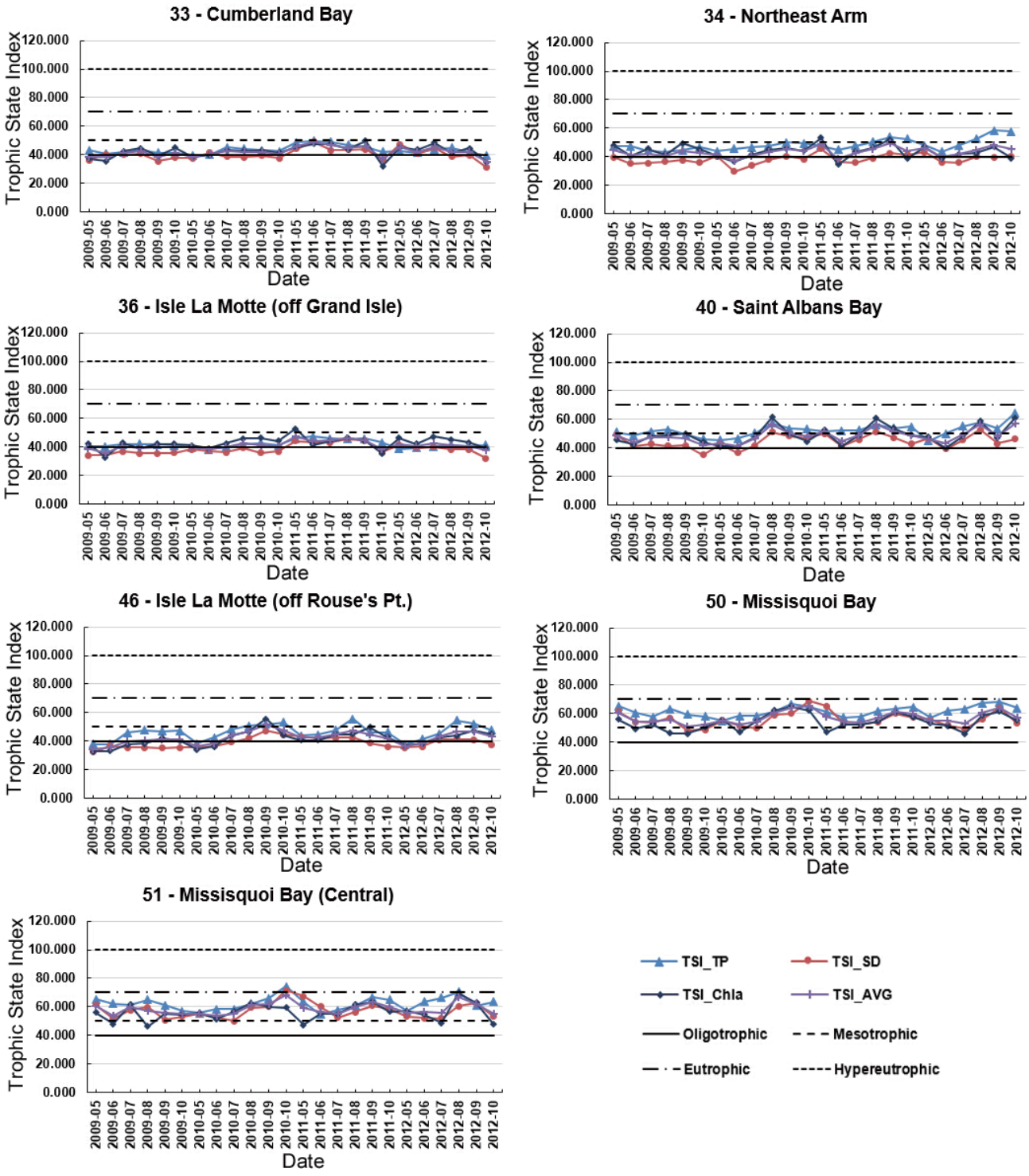


Figure 5-14 Trophic state index of the 15 lake stations from May to October 2012 (II)

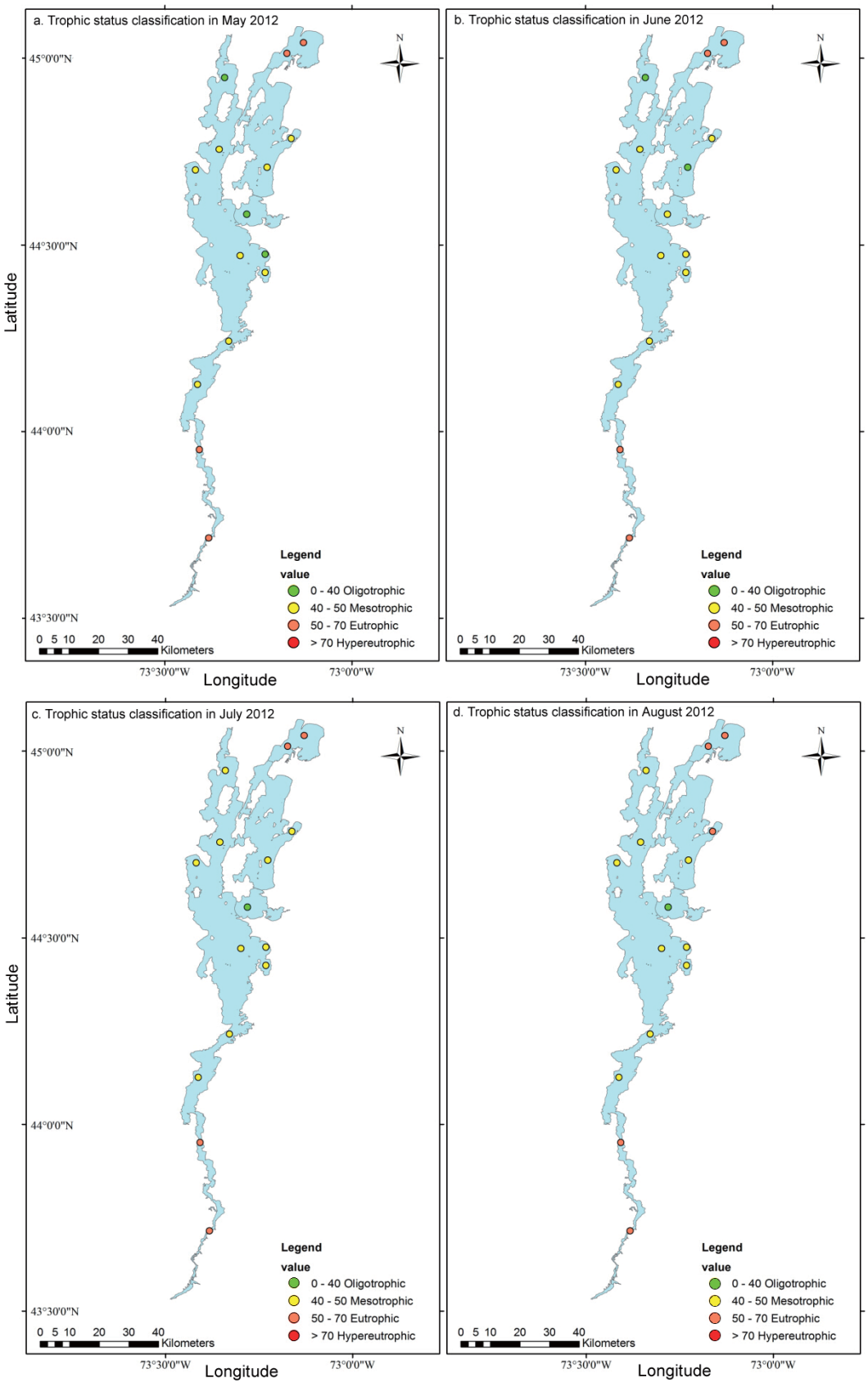


Figure 5-15 Trophic status classification map of May to August 2012

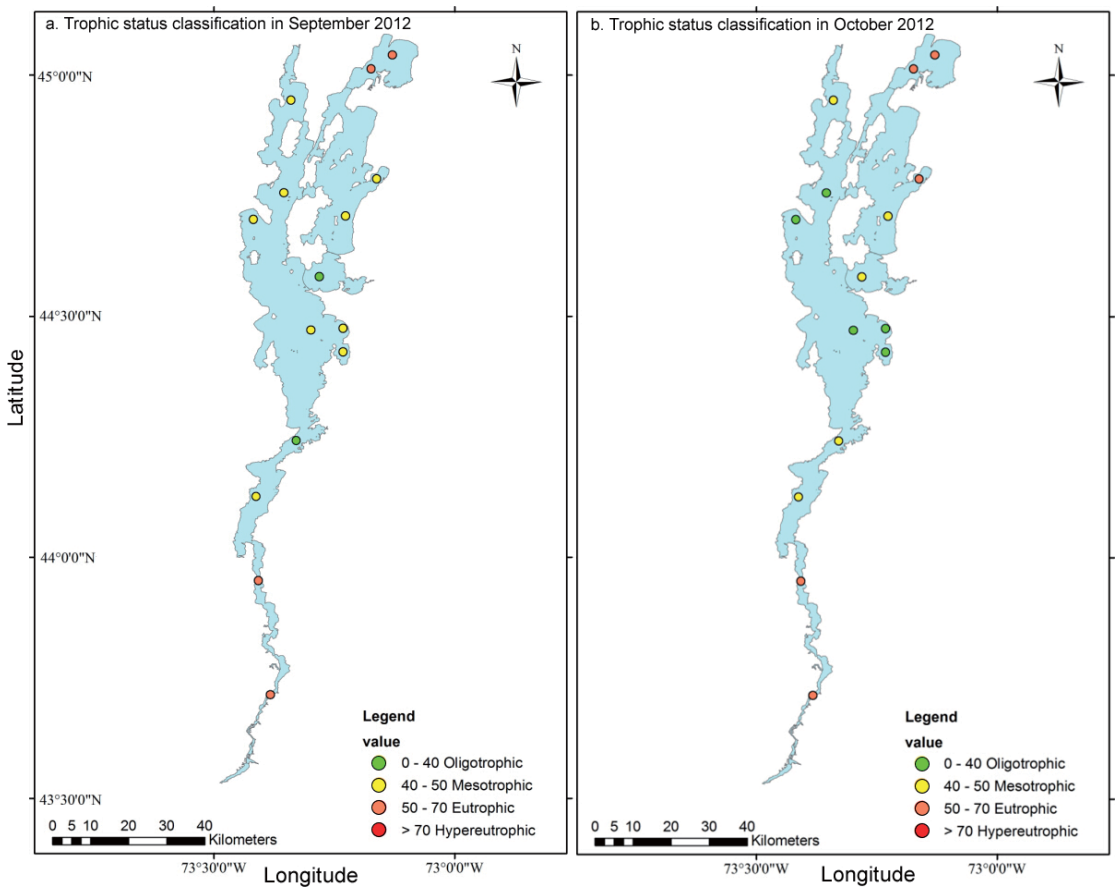


Figure 5-16 Trophic status classification map of September and October 2012

5.4 Summary

A case study was conducted using the developed LWQ module in the assessment of the lake water quality and ecological risk of Lake Champlain. By comparing the spatial distribution maps of the water quality risk assessment results generated via the LWQ module, the parameters that have primary influence of the lake water quality were analyzed. The developed LWQ module is useful for providing support for the decision makers and the lake managers.

CHAPTER 6 WATER QUALITY PREDICTION THROUGH THE GIS-SWQAM (ANN MODULE) -- A Case Study of Lake Champlain

6.1 Study Area and Data Preparation

The information of the study area was illustrated in Section 5.1.

Besides the lake water quality data mentioned in Section 5.2, the water chemistry data and flow rate data of tributaries around Lake Champlain were also included in this part of the study.

6.1.1 Tributary water chemistry data collection

Tributary water chemistry data used in this study were obtained from the Lake Champlain long-term water quality and biological monitoring project database (Lake Champlain Basin Program, 2014) and The Ministry of Sustainable Development, Environment and the Fight against Climate Change (The Ministry of Sustainable Development). These monitoring stations are shown in Figure 5-1 and listed in Table 6-1. The monitoring station of the Richelieu River, the outlet of Lake Champlain, is not included in Figure 5-1. The location and code for this station are listed in Table 6-1.

Table 6-1 List of tributary water chemistry sampling stations and their locations

Tributary Station	Latitude N	Longitude W	Station Code	State
Winooski	44° 31.52'	73° 15.41'	WINO01	VT
Otter Creek	44° 09.94'	73° 15.40'	OTTE01	VT
Missisquoi	44° 55.23'	73° 07.63'	MISS01	VT
Lamoille	44° 37.96'	73° 10.39'	LAMO01	VT
Poultney	43° 34.24'	73° 23.53'	POUL01	VT
Pike	45° 07.40'	73° 04.13'	03040015	QC
Lewis	44° 14.80'	73° 14.77'	LEWI01	VT
Little Otter	44° 12.24'	73° 15.11'	LOTT01	VT
LaPlatte	44° 22.21'	73° 13.01'	LAPL01	VT
Saranac	44° 41.52'	73° 27.19'	SARA01	NY
Ausable	44° 33.63'	73° 26.95'	AUSA01	NY
Mettawee	43° 33.33'	73° 24.10'	METT01	NY
Great Chazy	44° 58.81'	73° 25.96'	GCHA01	NY
Bouquet	44° 21.84'	73° 23.41'	BOUQ01	NY
Little Ausable	44° 35.65'	73° 29.79'	LAUS01	NY
Salmon	44° 38.40'	73° 29.70'	SALM01	NY
Putnam	43° 57.35'	73° 25.99'	PUTN01	NY
Little Chazy	44° 54.12'	73° 24.88'	LCHA01	NY
Rock River	45° 00.95'	73° 03.79'	03040013	QC
Stevens Brook [1]	44° 50.95’	73° 07.15’	STEV01	VT
Jewett Brook [1]	44° 51.37’	73° 09.06’	JEWE02	VT
Mill River [2]	44° 46.46’	73° 08.39’	MILL01	VT
Richelieu River	45° 03.68’	73° 19.95’	03040012	QC
Notes: 1 Added in 2008		2 Added in 2010		

6.1.2 Tributary flow rate data collection

Tributary flow rate data was obtained from U.S. Geological Survey (USGS) (United States Geological Survey) and the Centre d'expertise hydrique du Québec (CEHQ) (Centre d'expertise hydrique du Québec). The location of these gages are shown Table 6-2. The flow rate of Pike River at Notre-Dame-de-Stanbridge was estimated by transferring data from the gage station at Bedford through the drainage

area ratio method. Similarly, the flow rate of Richelieu River at the outlet of Lake Champlain near Rouse's Point was obtained from Rapid Fryers hydrometric station (station number 02OJ007 of Environment Canada, www.wateroffice.ec.gc.ca) on Richelieu River through the drainage-area ratio method.

Table 6-2 List of stream flow gages on monitored tributaries

Tributary	Gage Location	State	Reference	Agency
Ausable River	Au Sable Forks	NY	04275500	USGS
Bouquet River	Willsboro	NY	04276500	USGS
Great Chazy	Perry Mills	NY	04271500	USGS
Little Ausable	Valcour	NY	04273800	USGS
Little Chazy	Chazy	NY	04271815	USGS
Mettawee River	Middle Granville	NY	04280450	USGS
Putnam River	Crown Point Center	NY	04276842	USGS
Salmon River	S. Plattsburgh	NY	04273700	USGS
Saranac River	Plattsburgh	NY	04273500	USGS
Pike River	Bedford	QC	030420	MDDEP
Pike River[1]	Notre-Dame-de-Stanbridge	QC	030424	MDDEP
Lamoille River	E. Georgia	VT	04292500	USGS
LaPlatte River	Shelburne Falls	VT	04282795	USGS
Lewis River	N. Ferrisburg	VT	04282780	USGS
Little Otter River	Ferrisburg	VT	04282650	USGS
Missisquoi River	Swanton	VT	04294000	USGS
Otter Creek	Middlebury	VT	04282500	USGS
Poultney River	Fair Haven	VT	04280000	USGS
Winooski River	Essex Jct.	VT	04290500	USGS
Rock River [1]	St. Armand	QC	030425	MDDEP
Stevens Brook[2]	St. Albans	VT	04292770	USGS
Jewett Brook[3]	St, Albans	VT	04292810	USGS
Mill River [4]	St, Albans	VT	04292750	USGS
Richelieu River	Rapides Fryers	QC	02OJ007	Environment Canada
Richelieu River	Outlet of Lake Champlain	QC		Estimated

1 New gages on the Pike and Rock were installed by Quebec MDDEP in 2002.
2 New gage on Stevens Brook was installed by USGS in 2005.
3 New gage on the Jewett Brook was installed by USGS in 2008.
4 New gage on Mill River was installed by USGS in 2010.

Rapid Fryers station is located at some distance downstream on the Richelieu River. The drainage area ratio is 0.974 (Drainage area of Rouse's Point / Drainage

area of Rapid Fryers = 21437 / 22000) was used to estimate the outflow rate at the outlet of Lake Champlain.

The drainage-area ratio method assumes that the stream flow for the ungaged site can be estimated by multiplying the streamflow for the nearby streamflow-gaging station by the ratio of the drainage area for the ungaged site and the drainage area for the nearby streamflow-gaging station. The drainage-area ratio method is given by (Emerson et al., 2005, OhioEPA, 2010):

$$Q_{ungaged} = \frac{A_{ungaged}}{A_{gaged}} \times Q_{gaged} \tag{6.1}$$

where $Q_{ungaged}$ is the flow at the ungaged site, Q_{gaged} is the flow at the gaged site, $A_{ungaged}$ is the drainage area of the ungaged station, and A_{gaged} is the drainage area of the gaged station.

6.2 Prediction of Chlorophyll-a Concentrations in Lake Champlain

In this part, the BPANN model for predicting the chlorophyll a concentration with respect to time and space was conducted by using the Matlab artificial neural network toolbox.

6.2.1 Data for the prediction of chlorophyll-a

Considering the availability of long-term monitoring data, 12 of the 15 lake sampling stations (except stations 9, 16, and 51) were included in this part.

Monthly average values of the 9 parameters that have long-term monitoring data were calculated on the basis of the monitoring data in the lake water chemistry database from 1992 to 2012, including pH, chloride, water temperature (T), Secchi

depth (SD), dissolved oxygen (DO), dissolved phosphorus (DP), total phosphorus (TP), total nitrogen (TN), and chlorophyll-a. Chlorophyll-a will be used as the output of the BPANN model in this part.

The complete lake water quality data set comprises 12 stations × 21 years × 6 months/year = 1512 samples. The data of first thirteen years of each station were used as the training data, the next four years as the validation data and the remaining four years as the testing data. Thus, the training, validation and testing data sub-sets were comprised of 936 (61.9%), 288 (19.05%) and 288 (19.05%) samples, respectively.

6.2.2 Results and discussion

6.2.2.1 Alternative input variable selection

The alternative input data were obtained using the results of the Pearson's correlation analysis between the concentration of chlorophyll-a and other water quality variables. The absolute value of the correlation coefficient represented the degree of correlation, i.e. more than 0.8 represented very strong correlation, from 0.5 to 0.8 indicated a strong correlation, from 0.3 to 0.5 showed a moderate correlation, from 0.1 to 0.3 represented a modest correlation, and less than 0.1 was weak correlation (HaSS, 2014). Therefore, the variables with the absolute value of correlation coefficients greater than 0.1 were selected as alternative input variables. Based on the correlative analysis of the concentration of chlorophyll-a and eight other parameters (Table 6-3), the variables that have the absolute value of correlation coefficients greater than 0.1 were SD, DP, TP, TN, pH, and T. Moreover, algae produce oxygen in daylight and use up oxygen during the night. Oxygen consumption also occurs during the process of death and decay of algae. Therefore, dissolved oxygen is also correlated with chlorophyll-a concentration. Consequently, seven water

quality variables, including SD, DP, TP, TN, pH, DO and T, were selected as alternative input variables of the BPANN model for chlorophyll-a prediction.

Table 6-3 Correlation analysis of the concentration of chlorophyll-a and other water quality variables

Variables		Chlorophyll-a	Variables		Chlorophyll-a
Dissolved Phosphorus	Pearson Correlation	.340**	Dissolved Oxygen	Pearson Correlation	-.090**
	Sig. (2-tailed)	.000		Sig. (2-tailed)	.000
Chloride	Pearson Correlation	-.047	pH	Pearson Correlation	.263**
	Sig. (2-tailed)	.068		Sig. (2-tailed)	.000
Secchi Depth	Pearson Correlation	-.444**	Water Temperature	Pearson Correlation	.177**
	Sig. (2-tailed)	.000		Sig. (2-tailed)	.000
Total Nitrogen	Pearson Correlation	.453**	Total Phosphorus	Pearson Correlation	.567**
	Sig. (2-tailed)	.000		Sig. (2-tailed)	.000

**. Correlation is significant at the 0.01 level (2-tailed).

6.2.2.2 Spatial variation of the alternative input variables

Seven water quality variables, including SD, DP, TP, TN, pH, DO and T, were prepared as alternative input variables of the BPANN model for chlorophyll-a computation. The basic statistical parameters, i.e. the minimum, median, maximum, mean, the standard deviation, and the coefficient of variation of the selected variables were presented in Table 6-4.

Table 6-4 Basic statistics of the selected water quality variables for Chl-a prediction measured between 1992 and 2012 in Lake Champlain (n=1512)

Variable	Unit	Minimum	Maximum	Mean	Median	Std. Deviation	CV
Secchi Depth	m	0.30	9.05	3.76	4.00	1.80	0.48
pH	--	6.58	9.55	8.01	8.00	0.36	0.04
Temperature	℃	3.63	26.60	17.51	18.43	4.76	0.27
Dissolved Oxygen	mg/L	3.28	14.45	9.51	9.37	1.41	0.15
Dissolved Phosphorus	µg/L	3.00	54.00	10.79	8.00	7.21	0.67
Total Phosphorus	µg/L	5.33	140.50	22.93	16.00	16.15	0.70
Total Nitrogen	mg/L	0.20	1.35	0.42	0.40	0.12	0.28
Chlorophyll-a	µg/L	0.50	40.85	5.80	4.50	4.68	0.81

Dissolved phosphorus, total phosphorus, and chlorophyll-a showed large variations between lake sampling sites. For the twelve sampling sites in Lake Champlain, measured DP, TP, and chlorophyll-a averaged 10.79 ± 7.21 µg/L (mean ± S.D.), 22.93 ± 16.15 µg/L, and 5.8 ± 4.68 µg/L, respectively. Correspondingly, the coefficient of variation of these three variables were 0.67, 0.70, and 0.81, respectively. The large variation in DP and TP in Lake Champlain is likely attributed to the large geographical area of the lake and the different loads from tributaries distributed around the lake, and subsequently resulted in the variation in chlorophyll-a.

6.2.2.3 Performance of BPANN models for chlorophyll-a prediction

Seven scenarios established by the factors obtained through the Pearson correlation analysis are presented in Table 6-5. The input variables were sequentially excluded one by one in descending order of its correlation coefficient with chlorophyll-a. Then, the BPANN model was performed for each scenario. As stated previously, within the range of [I, $2I+1$], different BPANN models were constructed and tested to determine the optimum number of hidden neurons (N_h) for each scenario. Table 6-5 shows the results of performance of each scenario for the training, validation, and testing data sets. Among the seven scenarios, the one that uses all the seven variables as model input showed the highest prediction performance in the testing data set. The topology of this model has seven neurons in the input layer,

twelve neurons in the hidden layer and one neuron in the output layer. The R^2 values of the best performed model in the training set, validation set, and testing set were 0.82, 0.93, and 0.81, respectively. The corresponding RSR values of the three data sets were 0.62, 0.38, and 0.53, respectively.

Table 6-5 Scenarios and results of the chlorophyll-a prediction

Input Variables	Topology	Training Set		Validation Set		Testing Set	
		R^2	RSR	R^2	RSR	R^2	RSR
TP ,TN, SD, DP, pH, T, DO	7-12-1	0.82	0.62	0.93	0.38	0.81	0.53
TP, TN, SD, DP, pH, T	6-11-1	0.82	0.83	0.87	0.93	0.76	1.08
TP, TN, SD, DP, pH	5-11-1	0.79	1.10	0.88	0.67	0.74	1.25
TP, TN, SD, DP	4-7-1	0.76	1.49	0.89	0.78	0.73	1.45
TP, TN, SD	3-5-1	0.81	0.85	0.90	0.89	0.79	0.83
TP, TN	2-4-1	0.75	1.39	0.84	1.33	0.73	1.32
TP	1-3-1	0.76	1.26	0.84	1.10	0.72	1.53

For the optimal model selected above, Figure 6-1 shows the comparison of modeled and measured values of chlorophyll-a in training, validation and testing data sets. To further illustrate the model performance, Figure 6-2 shows the comparison of all the measured and modeled values of monthly chlorophyll-a, the coefficient of determination (R^2) for all the data was 0.87, RSR value was 0.48. The intercepts of the regression line of training data set, validation data set, testing data set and all data are 0.36, 2.4, 2.4, and 1.1, respectively. The results indicated that although some peaks and lows were not well predicted, the BPANN model can satisfactorily reproduce the variations of chlorophyll-a and was able to obtain satisfying prediction results using selected input parameters at sampling stations in Lake Champlain.

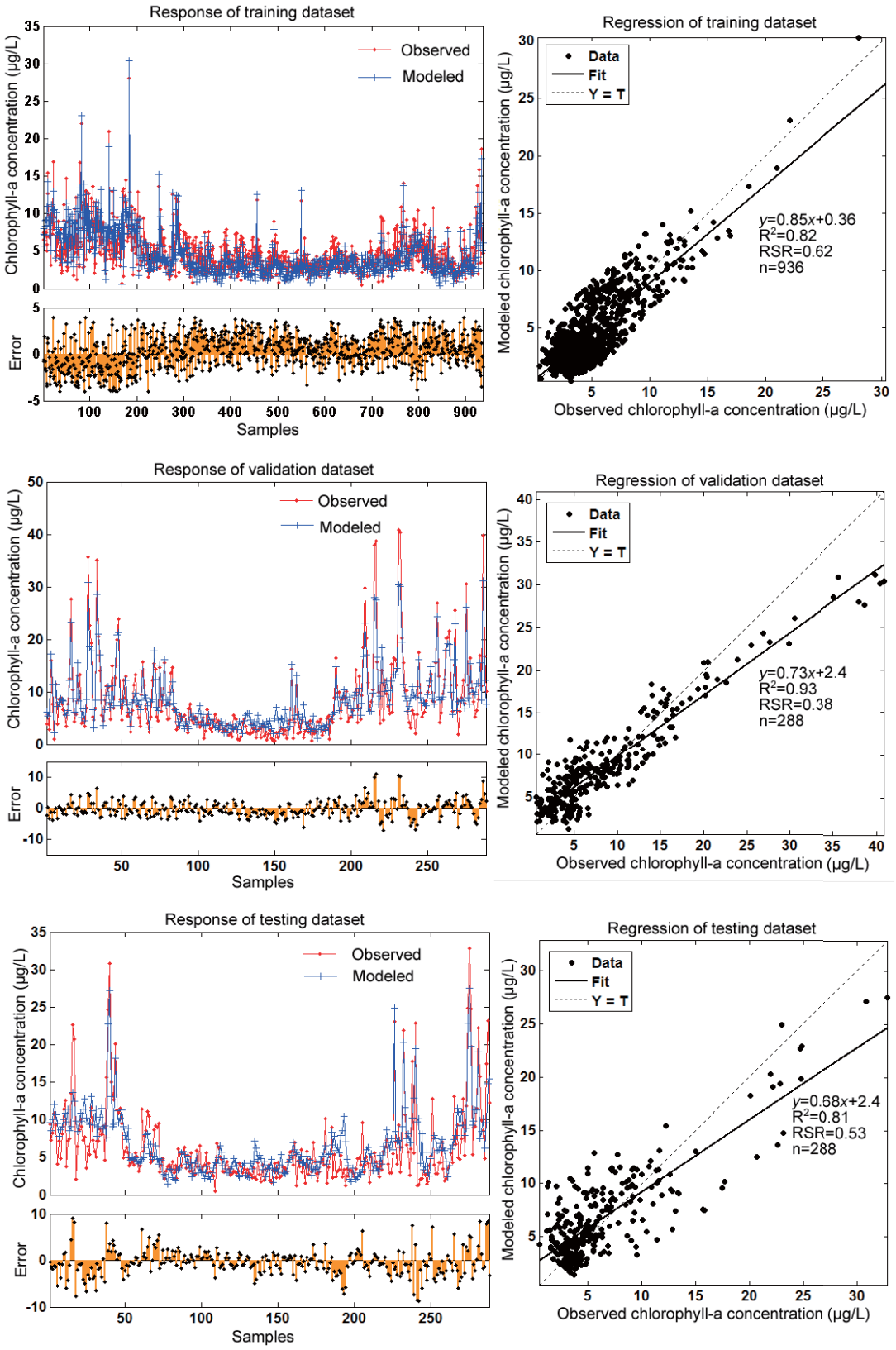


Figure 6-1 Comparison of the modeled and observed values of chlorophyll-a in training data set, validation data set and testing data set

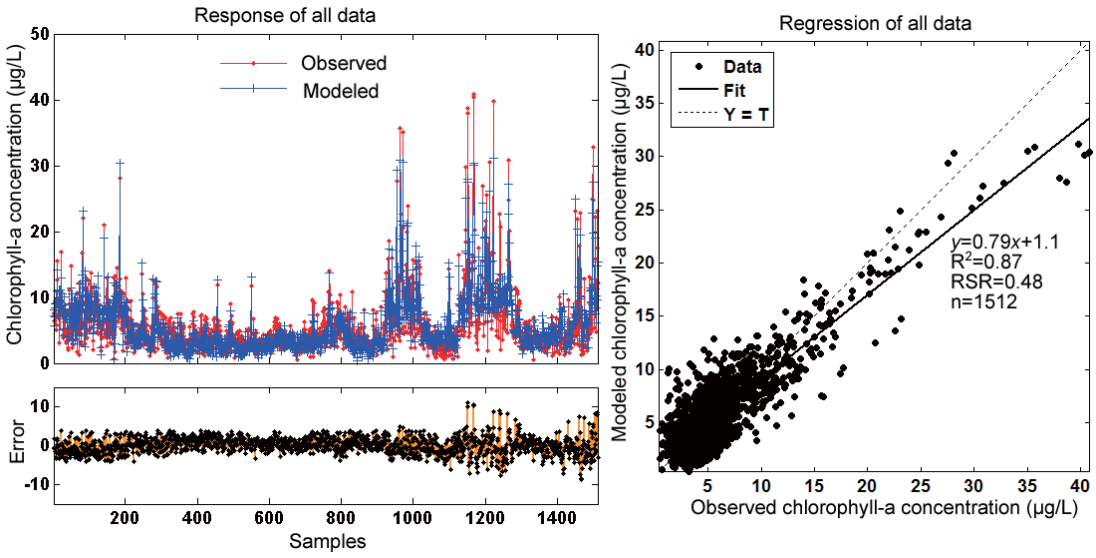


Figure 6-2 Comparison of the modeled and observed values of chlorophyll-a for all data

Though the BPANN predictions reproduce the measured values satisfactorily, as can be seen in Figure 6-1, the regression lines intersected with the dot lines that represent perfect fitting, and there are problems of underestimation for high values of chlorophyll-a concentration and overestimation of low values in the training, validation and testing data sets. For example, the observed chlorophyll-a is 40.85 µg/L while the modeled value is 30.36 µg/L. The reason for these problems in the training set is that the object of this study is to develop a BPANN model that has the better generalization ability, rather than a model that perfectly fits the training data. The underestimation and overestimation in the validation and testing data sets is probably due to the non-homogeneous nature of the input and output water quality variables, as these data were measured over a long period of 21 years and the sampling sites distributed over a large geographical area (approximately 1,269 km^2). Moreover, Maier et al. (Maier et al., 2010) stated that ANN models have the best performance when they are used within the range of the training data, and the statistical properties of the training, validation and testing data sets should be similar in order to develop the best performed model. For the time series prediction in this study, the values of

water quality variables varied during a long time period and over a large geographical area, so it is difficult to ensure the similarity of the statistical properties of the three data sets. As listed in Table 6-6, the range of chlorophyll-a concentration in the training, validation and testing data set was 0.5-28.07 μg/L, 0.64-40.85 μg/L, and 0.5-32.80 μg/L, respectively. The median and standard deviation of the validation and testing data sets were also higher than that of the training data set. The developed model was trained to fit the data in the training data set, which has a smaller range as compared to the other two data sets. The R^2 values of the testing and validation data sets were similar to or even higher than that of the training set. However, the y-intercepts of the validation and testing data sets are larger than that of the training data set, and the slopes of the regression line of the validation and testing data sets are smaller than that of the training data set. Larger y-intercepts and smaller slopes of the regression lines indicate larger lags between the observed and modeled values. These also represent the underestimation problem in validation and testing data sets, and indicate that the model performance in these two data sets are not as good as the model performance in the training data set. In addition, abrupt changes of data, probably arose from the inevitable errors during the sampling and analysis of Chl a, were found during the study period, and created a considerable challenge for the prediction. These factors would probably result in the underestimation or overestimation of the values in the data sets.

Table 6-6 Statistical properties of chlorophyll-a concentrations in training, validation, and testing data sets

	Training set	Validation set	Testing set
N	936	288	288
Mean (μg/L)	4.8960	8.3497	6.1693
Median (μg/L)	4.1950	5.7500	4.3900
Std. Deviation (μg/L)	2.85814	7.35180	5.05556
Minimum (μg/L)	0.50	0.64	0.50
Maximum (μg/L)	28.07	40.85	32.80

Although there are some underestimation problems, Figure 6-2 shows a good agreement for the total response between all the observed and modeled chlorophyll-a values. The coefficient of determination is 0.87. Moreover, the model simulation can be considered satisfactory if RSR≤ 0.7 (Moriasi et al., 2007). In this study, the RSR value of the testing data set is 0.53, which indicates that the developed BPANN model is applicable to the chlorophyll-a prediction.

For the determination of model input, Cho et al. (Cho et al., 2014) stated that the accuracy of the model could be improved by using a relatively small number of input variables instead of using various environmental factors. This is also supported by the research results of Lee et al. (Lee et al., 2003). On the contrary, the results of this study indicate that all the seven variables were selected as the optimal model input rather than using a relatively small number of input variables. Similar results were found by He et al. (He et al., 2011). These contradictions are probably due to the different properties of the water body and the study areas, e.g. the hydraulic conditions in the water body and the meteorological conditions in the study area.

The coefficient of determination (R^2) of the testing data set was 0.81 in this study, which was lower than the result of Cho et al.(2014) (R = 0.85). The possible reasons include: (1) The training data used in Cho's study was continuous monthly data from January to October, while the training data used in this study was obtained from May to October; (2) The prediction period of Cho's study was one year, but a longer prediction period (four years) was applied in this study. It would be reasonable that the prediction accuracy will decrease while increasing the number of data to be predicted.

6.3 Prediction of Total Phosphorus Concentrations in Lake Champlain

The ANN module in the developed system was utilized in this part for the prediction of total phosphorus concentration in Lake Champlain.

6.3.1 Data for the prediction of total phosphorus

Considering the results of the prediction of TP concentration using the BPANN model will be compared with the results of MIKE 21 model, which is conducted with data from all over the lake, data from all the 15 lake stations were included in this part.

Lake Champlain is plagued by an excessive influx of nutrients. Phosphorus loading from tributaries is an important element in modeling of total phosphorus in Lake Champlain. The concentration of chlorophyll-a was not included in the modeling of total phosphorus because the increase of chlorophyll-a concentration is always considered as a consequence of the increase of total phosphorus concentration. So there were in total 12 parameters included in the ANN modelling of total phosphorus: temperature (T), chloride (Cl), total nitrogen (TN), Secchi depth (SD), dissolved phosphorus (DP), dissolved oxygen (DO), total phosphorus (TP), inflow and outflow rate of all tributaries around the lake, inflow concentration of total phosphorus from tributaries (TP_Inflow), and outflow concentration of total phosphorus from the Richelieu River (TP_Outflow). Monthly average values of these parameters were obtained from the data retrieved from the database. In this part, total phosphorus concentration was used as the output of the ANN models.

In this part, the complete lake water quality data set is comprised of 1698 samples, including data from May to October of 21 years at the same 12 stations as in chlorophyll a model (from 1992 to 2012), data from May to October of 12 years at

station 9 and 16 (from 2001 to 2012), and data from May to October of 7 years at station 51 (from 2006 to 2012). The data from 1992 to 2008 of each station was used as the training data, the next four years as the validation data and the remaining four years as the testing data. Thus, the training, validation and testing data sub-sets were comprised of 984 (57.95%), 354 (20.85%) and 360 (21.2%) samples, respectively.

6.3.2 System implementation – ANN module

6.3.2.1 Alternative input variable selection

Based on the correlative analysis of the concentration of total phosphorus and nine other parameters (Table 6-7), the variables that have the absolute value of correlation coefficients greater than 0.1 were chosen as alternative input variables of the BPANN model for total phosphorus prediction, including SD, DP, TN, DO, T, and TP_Inflow.

Table 6-7 Correlation analysis of the concentration of total phosphorus and other water quality variables

Variables		Total Phosphorus	Variables		Total Phosphorus
Dissolved Phosphorus	Pearson Correlation	.862**	Dissolved Oxygen	Pearson Correlation	-.343**
	Sig. (2-tailed)	0.000		Sig. (2-tailed)	0.000
Secchi Depth	Pearson Correlation	-.728**	Water Temperature	Pearson Correlation	.201**
	Sig. (2-tailed)	0		Sig. (2-tailed)	0.000
Chloride	Pearson Correlation	0.063**	TP_Outflow	Pearson Correlation	0.020
	Sig. (2-tailed)	0.010		Sig. (2-tailed)	0.411
Total Nitrogen	Pearson Correlation	.587**	TP_Inflow	Pearson Correlation	.119**
	Sig. (2-tailed)	0		Sig. (2-tailed)	0.000
pH	Pearson Correlation	-.036	Inflow Rate	Pearson Correlation	-.029
	Sig. (2-tailed)	.133		Sig. (2-tailed)	0.236
Outflow Rate	Pearson Correlation	-.033			
	Sig. (2-tailed)	0.169			

**. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed).

6.3.2.2 Spatial variation of the alternative input variables

The basic statistical parameters, i.e. the minimum, median, maximum, mean, the standard deviation, and the coefficient of variation of the selected six alternative variables along with total phosphorus were presented in Table 6-8.

Table 6-8 Basic statistics of the selected water quality variables for TP prediction measured between 1992 and 2012 in Lake Champlain (n=1698)

Variable	Unit	Minimum	Maximum	Mean	Median	Std. Deviation	CV
Secchi Depth	m	0.30	9.05	3.76	4.00	1.79	0.48
Temperature	℃	3.63	26.60	17.59	18.44	4.72	0.26
Dissolved Oxygen	mg/L	3.28	14.45	9.53	9.40	1.41	0.15
Dissolved Phosphorus	μg/L	3.00	58.65	10.86	8.00	7.37	0.68
Total Phosphorus	μg/L	5.33	140.50	23.04	15.70	16.58	0.72
Total Nitrogen	mg/L	0.20	1.35	0.43	0.40	0.13	0.30
TP_Inflow	μg/L	201.27	4565.29	1333.3	1187.65	627.12	0.48

Similar to the statistical results in the chlorophyll a concentration modeling, dissolved phosphorus and total phosphorus showed relatively large variations. Measured DP and TP concentrations in Lake Champlain averaged 10.86 ± 7.37 μg/L (mean ± S.D.) and 23.04 ± 16.58 μg/L, respectively. The coefficient of variation of these two variables were correspondingly 0.72 and 0.68.

6.3.2.3 BPANN model training and prediction of total phosphorus using the developed ANN module

The alternative input variables were sequentially excluded one by one in descending order of its correlation coefficient with total phosphorus to establish six scenarios as presented in Table 6-9. The corresponding input data of each scenario was prepared separately. The input data along with the total phosphorus data were divided into three parts according to the sampling date, i.e., the training data set, validation data set, and testing data set. The input and output data were organized into separate CSV files. For the training data set, the input data was organized into the

CSV file named “Train Input”, which contains groups of input data corresponding to each scenario over the period of 1992 to 2008. The target data were organized into the file named “Train Target”, which contains the monitoring concentration of TP of all the stations from 1992 to 2008 and was used as the training target for the BPANN model. Analogously, the input and output data were organized as the CSV files named “Test Input” and “Test Actual” respectively for the validation and testing data sets. The validation and testing of the model should be performed separately, although the file name of the data was the same. Data in the “Test Actual” file was used for comparison with the predicted concentration of the BPANN model.

The scenario with four input variables was used as an example to illustrate the application of the ANN module. The ANN module was implemented following the steps illustrated in Section 3.5.2.3. Examples of interfaces of loading data, model structure selection and results display were shown in Figure 3-8 to Figure 3-11. The input neuron number was four and the output neuron number was one. In this case, the ANN model with a single hidden layer was applied. The optimal number of hidden neurons (N_h) was determined within the range of [I, $2I+1$] through the trial and error method, where I is the number of input variables. Each model was retrained ten times and all the results were saved to CSV files for further comparison. Model performances were sequentially evaluated according to the calculated R^2 and RSR values. The generated contour maps were able to provide the spatial distribution of the predicted total phosphorus concentrations corresponding to the selected date.

6.3.3 Results and discussion

For each scenario, the optimal performance was selected among the results of each data set obtained by applying different model structures. Best model performances of the training, validation, and testing data sets of the six scenarios were

listed in Table 6-9. As observed in Table 6-9, the model with five input variables has the best performance. The topology of this model has five neurons in the input layer, ten neurons in the hidden layer and one neuron in the output layer. The values of R^2 in the training, validation, and testing data set were 0.90, 0.92, and 0.92, respectively. Correspondingly, the RSR values of the three data sets were 0.44, 0.39, and 0.40, respectively.

Table 6-9 Scenarios and results of the total phosphorus prediction

Input Variables	Topology	Training Set		Validation Set		Testing Set	
		R^2	RSR	R^2	RSR	R^2	RSR
DP, SD, TN, T, DO, TP_Inflow	6-8-1	0.90	0.43	0.92	0.40	0.73	0.42
DP, SD, TN, T, DO	5-10-1	0.90	0.44	0.92	0.39	0.92	0.40
DP, SD, TN, T	4-7-1	0.89	0.46	0.86	0.55	0.91	0.41
DP, SD, TN	3-5-1	0.89	0.46	0.89	0.46	0.92	0.40
DP, SD	2-4-1	0.86	0.51	0.92	0.42	0.91	0.43
DP	1-3-1	0.85	0.53	0.90	0.44	0.90	0.46

Comparison of predicted and measured concentrations of total phosphorus in training, validation and testing data sets is presented in Figure 6-3. Figure 6-4 further presents the model performance by comparing all the data of predicted and observed total phosphorus concentrations. The slope and intercept of the regression lines of the training data set, the validation data set and the testing data set were 0.78 and 4.1, 0.81 and 4.8, 0.8 and 4.2, respectively. For the total response, the coefficient of determination (R^2) was 0.91, RSR value was 0.42, and the slope and intercept of the regression line were 0.79 and 4.2.

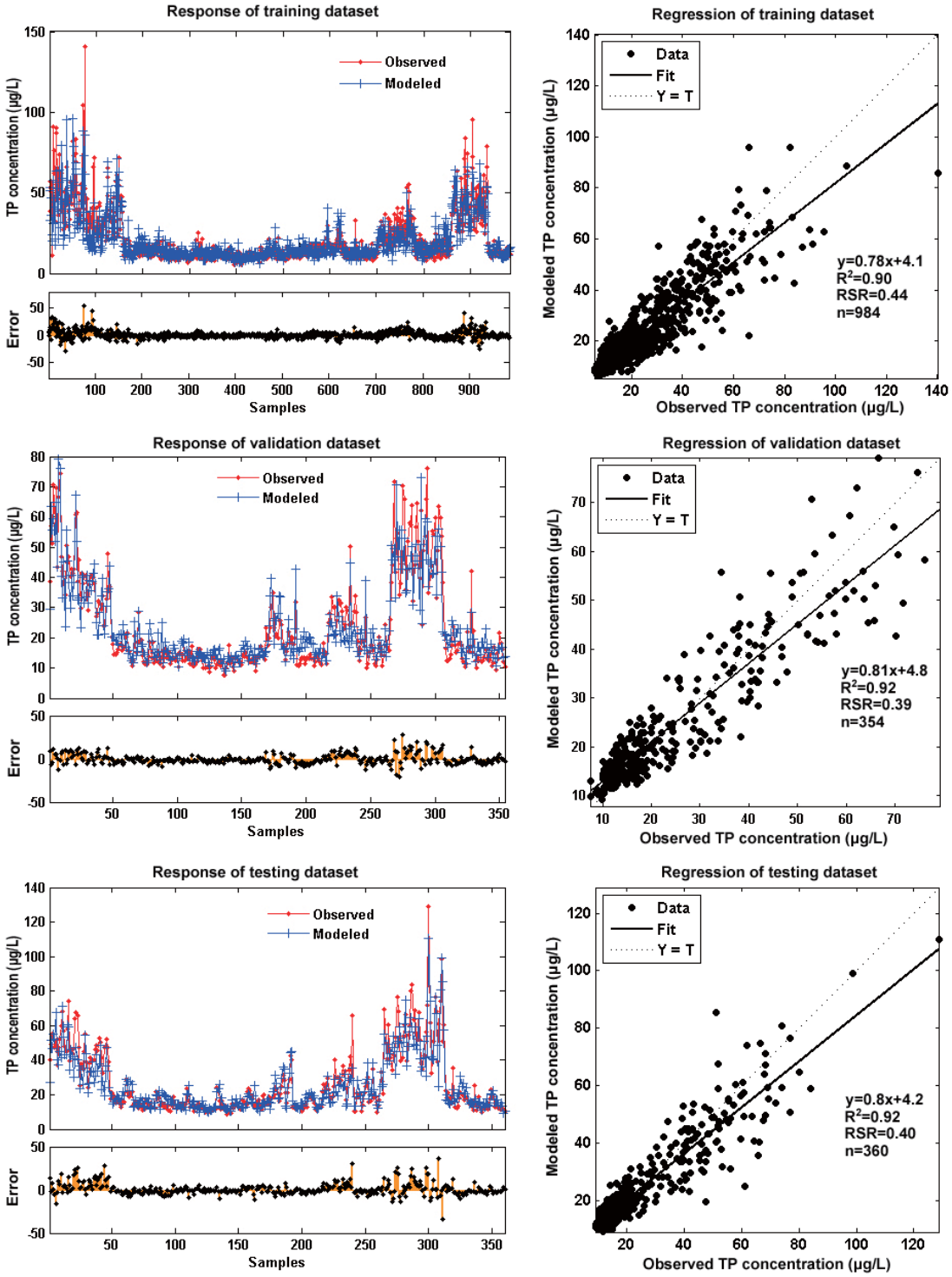


Figure 6-3 Comparison of the modeled and observed values of total phosphorus in training data set, validation data set and testing data set

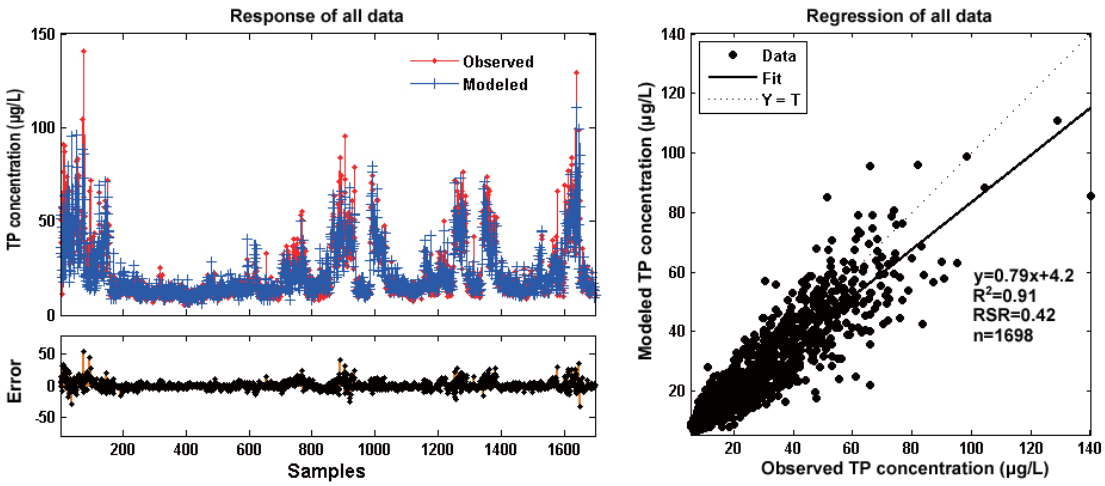


Figure 6-4 Comparison of the modeled and observed values of total phosphorus for all data

According to the above results, the validation and testing results are very similar, thus no over-fitting occurred during the training of the model. The R^2 values of the training, validation and testing data sets were all greater than 0.5, and all the RSR values were less than 0.7, which indicated that the predicted values of TP concentrations show good agreement with the actual values. Although some peaks were not predicted, the BPANN model established through the developed ANN module was able to satisfactorily simulate the variations of total phosphorus concentration in Lake Champlain.

As displayed in Figure 6-3 and Figure 6-4, the linear regression lines of the training, validation and testing data sets, as well as the total response of all data, intersected with the dot lines, which represent perfect fitting. The regression lines were mostly located below the dot lines. This indicated slight overestimation of the low values and underestimation of high values in the prediction of TP concentrations, which was similar to the predicted results of chlorophyll-a concentrations in the previous section. Analogously, the abrupt variation of data may account for the overestimation or underestimation problems. With the exception of inevitable errors

during the sampling and analysis process, the other possible cause was the variation of total phosphorus loads into Lake Champlain through tributaries, point sources, and non-point sources.

The results of TP prediction differed from that of Chl-a prediction in the comparison between the results of the three data sets. In the prediction of Chl-a concentrations, the ANN model performed better in the training data set than the other data sets due to the smaller range of data in the training data set. In contrast, the model performances in the prediction of TP concentrations were slightly better in the validation and testing data sets than in the training data set, which was proved by higher values of R^2 and the slopes of the linear regression lines, as well as the lower RSR values of the results of the validation and testing data sets. It is hypothesized that the reason for this was that the data range of TP concentrations in the training set, which was utilized as the target of training the model, was larger than that of the other two data sets, as displayed in Table 6-10. Therefore, by training the ANN model to fit the target data with a larger range, the model was capable of simulating the variations in the data set that has a smaller range.

Table 6-10 Statistical properties of TP concentrations in training, validation, and testing data sets

	Training data set	Validation data set	Testing data set
n	984	354	360
Mean (μg/L)	21.4222	23.7950	26.7089
Median (μg/L)	14.2250	16.9300	18.4000
Std. Deviation (μg/L)	16.05542	15.40850	18.42201
Minimum (μg/L)	5.33	7.57	8.90
Maximum (μg/L)	140.50	76.03	129.00

Additionally, the spatial distribution of the predicted concentrations of the selected month was presented in the “MapView” window. Take the results of May 2012 as an example: the generated contour map along with the range of the predicted values were shown in Figure 6-5. Compared with the maps of the monitoring

concentrations (Figure 5-10 (d)), the spatial distribution of the prediction TP concentrations agrees well with that of the monitoring values.

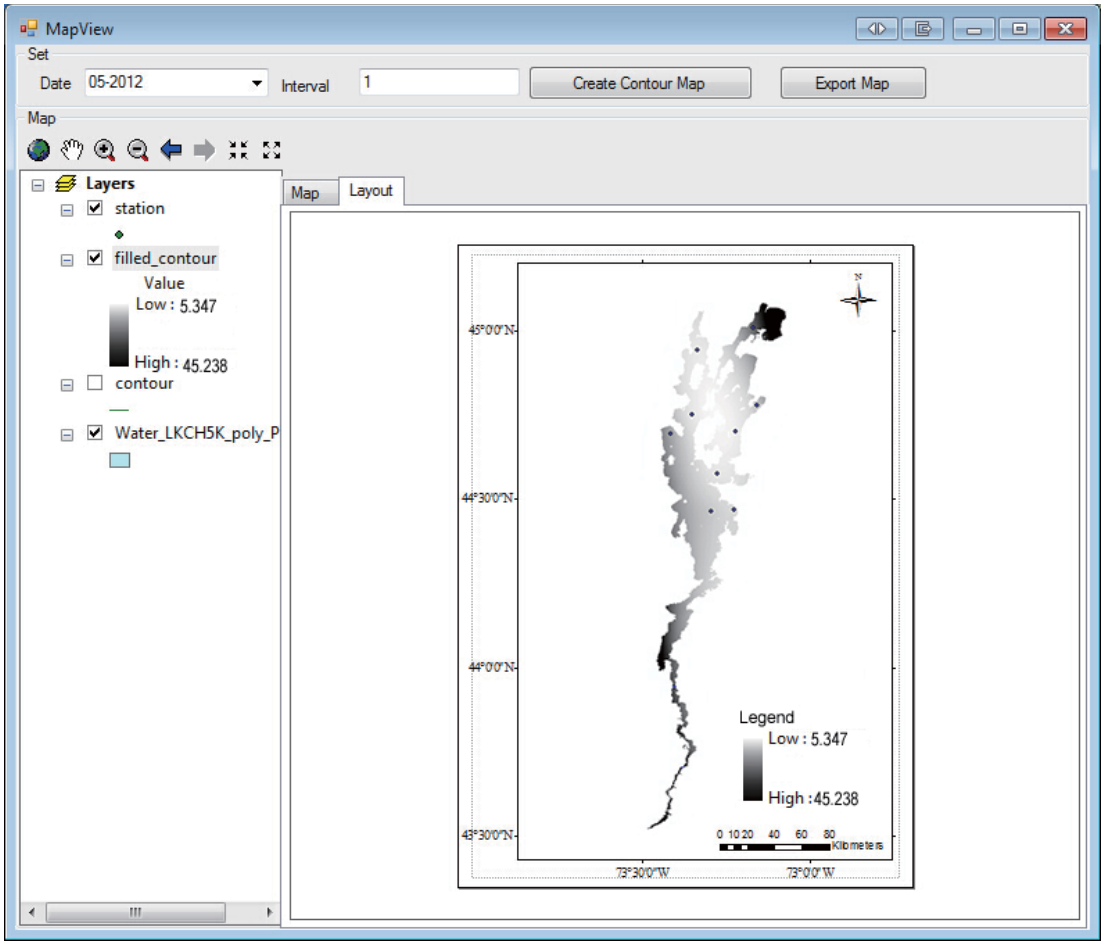


Figure 6-5 Interface of the spatial distribution map of the predicted total phosphorus concentration generated by the ANN module

6.4 Summary

This study explored the efficiency of the artificial neural network model in the prediction of water quality variables in Lake Champlain, which has a complex shoreline and geographical conditions. Chlorophyll-a concentrations were predicted through the ANN model provided by Matlab software. Subsequently, the algorithm of artificial neural network was compiled using C# code and integrated with the ArcGIS Engine and database to build the ANN module of the developed system. The ANN

module developed was applied to predict the total phosphorus concentration in Lake Champlain. The prediction results were verified with the lake water monitoring data in terms of R^2 and RSR, which indicate that the ANN models developed are able to provide satisfactory results in the prediction. Moreover, the spatial distribution maps of the predicted results generated agreed well with that of the monitoring values, which also provided an intuitive illustration of the prediction results. The developed ANN module could be used to provide information about the variation of water quality parameters.

CHAPTER 7 COMPARISON STUDY USING THE DEVELOPED ANN MODEL AND THE MIKE 21 MODEL FOR LAKE CHAMPLAIN

As one of the largest bodies of fresh water in the United States, the drainage basin of Lake Champlain includes portions of Vermont, New York, and Quebec. It stretches almost 200 km north-south, but is only 19 km wide and has complex shorelines and lake regions.

A phosphorus budget and mass balance model was developed using the BATHTUB program to identify phosphorus load reductions to attain the total phosphorus concentration criteria established in the water quality agreement (Smeltzer and Quinn, 1996). Daniel L. Mendelsohn and Henry Rines have conducted another phosphorus cycle water quality model based on the box model linked with the phosphorus kinetic rate equations taken from WASP5. The model was applied to study the dynamics and kinetics of the distribution and concentration of total phosphorus in the lake (Daniel L. Mendelsohn and Rines, 1995). Both of these two studies used the box model, which requires that the lake be represented by a number of boxes of known area, volume, river flow, and total phosphorus concentration. Since Lake Champlain has complex shorelines, the box model may not be able to fit the shoreline thus may reduce the simulation precision. The MIKE 21 model with a flexible mesh approach was able to fit better with the irregular boundary of Lake Champlain with the unstructured triangular grids.

In this chapter, considering the availability of lake and tributary water column data, sediment data and biological data, a hydrodynamic model and a simple water quality model (transport model) were established to simulate the monthly averaged total phosphorus concentration in Lake Champlain from May to October between

2009 and 2012. The time step interval of the model was set at 86400 s (1 day) and the monthly average values of the model results was further calculated. The simulation started from June 2008 to allow some time for model stabilization. The model calibration was performed by applying input data from June 2008 to December 2010 and the results were compared with the observed monthly average TP concentrations during the same period. Sequentially, the model validation was carried out by applying the model input data during January to October in 2011 and 2012 and the model results were cross-compared with the corresponding observed concentrations and the results of the ANN model in the developed GIS-SWQAM system in terms of the coefficient of determination and RSR.

7.1 Data Sources

The required data to establish the MIKE 21 flow model and water quality model include the terrain data, water chemistry data of the lake and tributaries, tributary inflow rate and outflow rate, and meteorology data. The sources of water chemistry data of the lake and tributaries and the flow rate data were illustrated in Section 5.2.

7.1.1 Terrain data

The bathymetry DEM and hydrography data for Lake Champlain were downloaded from the Vermont Center for Geographic Information under the Fresh Water Resources theme (http://vcgi.vermont.gov/warehouse/search_tools). The bathymetry DEM includes the lake depth at each point and the horizontal datum is NAD 1983. The horizontal datum and altitude datum of hydrography data are NAD 1983 and NGVD 1929, respectively.

7.1.2 Water intake

Water is withdrawn from Lake Champlain at many locations to supply municipal, industrial and private use. In the Lake Champlain Diagnostic-Feasibility Study (Vermont DEC and New York State DEC, 1997), six large water intakes in the lake were listed and other small facilities were ignored in the study. The total amount of water withdrawn from the six largest facilities was around 46.1 hm^3/yr, which equals to 126301 m^3/day. This value was assumed to be constant over time and applied in this model. Due to the lack of location information about these water withdrawal facilities, it was difficult to locate these intake points in the MIKE 21 model. Thus, the water withdrawal rate was converted to evaporation rate over the lake by dividing it over the surface area of Lake Champlain, which comes to 0.0995 mm/day as evaporation. This amount will be added to the natural evaporation rate as a simplified water withdrawal rate in the model.

7.1.3 Meteorology data

7.1.3.1 Wind speed and direction

It has been demonstrated that wind is an important forcing function for currents in Lake Champlain (Daniel Mendelsohn et al., 1996). Monthly averaged wind speed and direction data taken at Burlington Airport, Vermont were applied to the entire lake domain. The data was collected from the website of National Centers for Environmental Information (NCEI) of the National Oceanic and Atmospheric Administration (NOAA) (National Centers for Environmental Information). The wind rose plots of 2008 to 2012 were shown in Figure 7-1 a. - e.

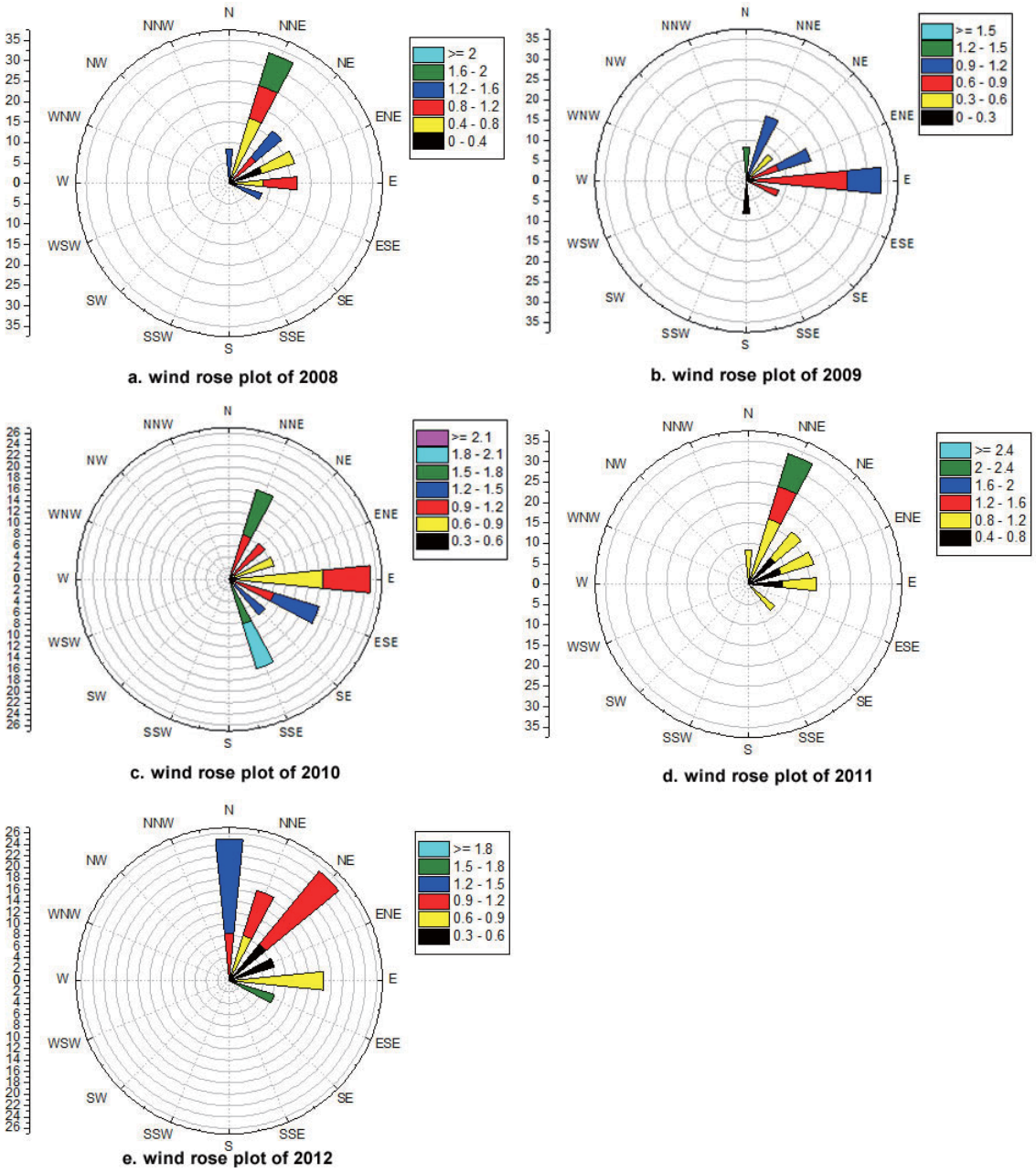
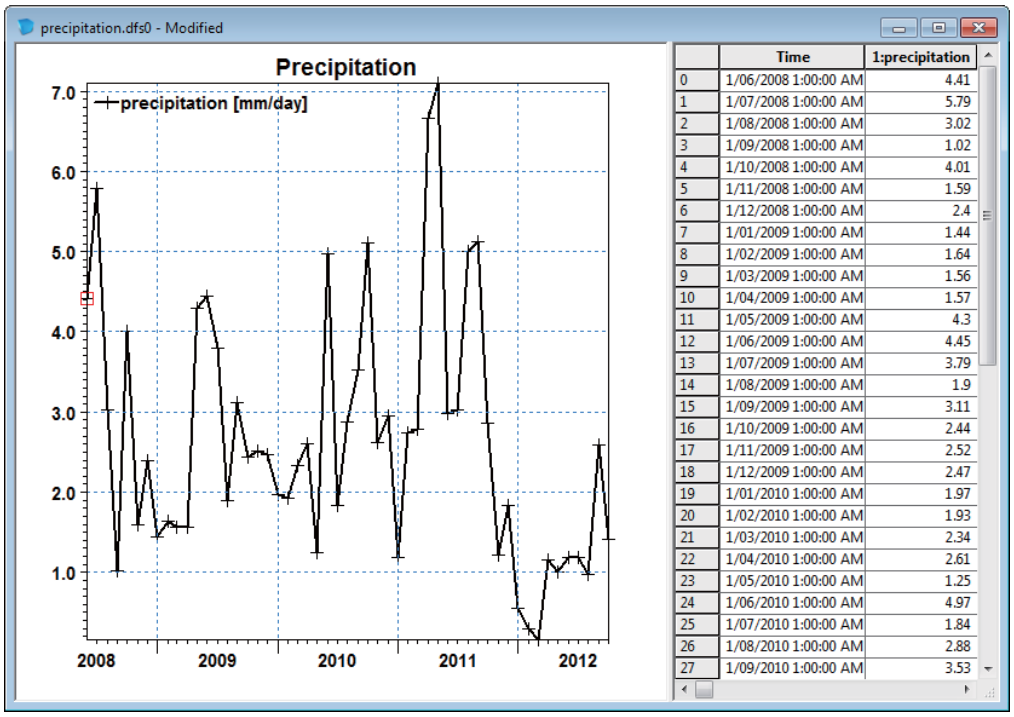


Figure 7-1 Wind rose plots of 2008 to 2012 according to the monthly wind data collected at Burlington Airport, Vermont

7.1.3.2 Precipitation

Precipitation data applied in this model were determined using monthly summaries of hourly precipitation data collected at Burlington Airport, Vermont. The

data was collected from the same source as the wind data. Figure 7-2 shows the variation of precipitation between June 2008 and October 2012.



	Time	1:precipitation
0	1/06/2008 1:00:00 AM	4.41
1	1/07/2008 1:00:00 AM	5.79
2	1/08/2008 1:00:00 AM	3.02
3	1/09/2008 1:00:00 AM	1.02
4	1/10/2008 1:00:00 AM	4.01
5	1/11/2008 1:00:00 AM	1.59
6	1/12/2008 1:00:00 AM	2.4
7	1/01/2009 1:00:00 AM	1.44
8	1/02/2009 1:00:00 AM	1.64
9	1/03/2009 1:00:00 AM	1.56
10	1/04/2009 1:00:00 AM	1.57
11	1/05/2009 1:00:00 AM	4.3
12	1/06/2009 1:00:00 AM	4.45
13	1/07/2009 1:00:00 AM	3.79
14	1/08/2009 1:00:00 AM	1.9
15	1/09/2009 1:00:00 AM	3.11
16	1/10/2009 1:00:00 AM	2.44
17	1/11/2009 1:00:00 AM	2.52
18	1/12/2009 1:00:00 AM	2.47
19	1/01/2010 1:00:00 AM	1.97
20	1/02/2010 1:00:00 AM	1.93
21	1/03/2010 1:00:00 AM	2.34
22	1/04/2010 1:00:00 AM	2.61
23	1/05/2010 1:00:00 AM	1.25
24	1/06/2010 1:00:00 AM	4.97
25	1/07/2010 1:00:00 AM	1.84
26	1/08/2010 1:00:00 AM	2.88
27	1/09/2010 1:00:00 AM	3.53

Figure 7-2 Precipitation data from June 2008 to October 2012

7.1.3.3 Evaporation

The evaporation rate from the lake surface obtained by Smeltzer (Smeltzer and Quinn, 1996) was assumed to be constant and applied to the entire lake domain in the MIKE 21 model herein, which was 0.65 m/yr (1.78 mm/day). The evaporation rate was estimated through the pan evaporation measurements at one NOAA station within the watershed by applying a regional pan-to-lake coefficient of 0.77.

7.2 Mesh Generation

The bathymetry DEM and hydrography data of Lake Champlain were processed using AutoCAD software to obtain the ASCII files that were used to generate mesh files in the mesh generator. The mesh generator is a powerful tool included in the

MIKE Zero software for creating, smoothing and adjusting the unstructured mesh files that contain the information for water depth and the boundary of the selected area. The generated mesh file for Lake Champlain is shown in Figure 7-3. Different colors represented different ranges of depth. The mesh density was increased in the narrow areas to increase the model accuracy. The final mesh file includes 5061 nodes and 7462 elements in the modeling area.

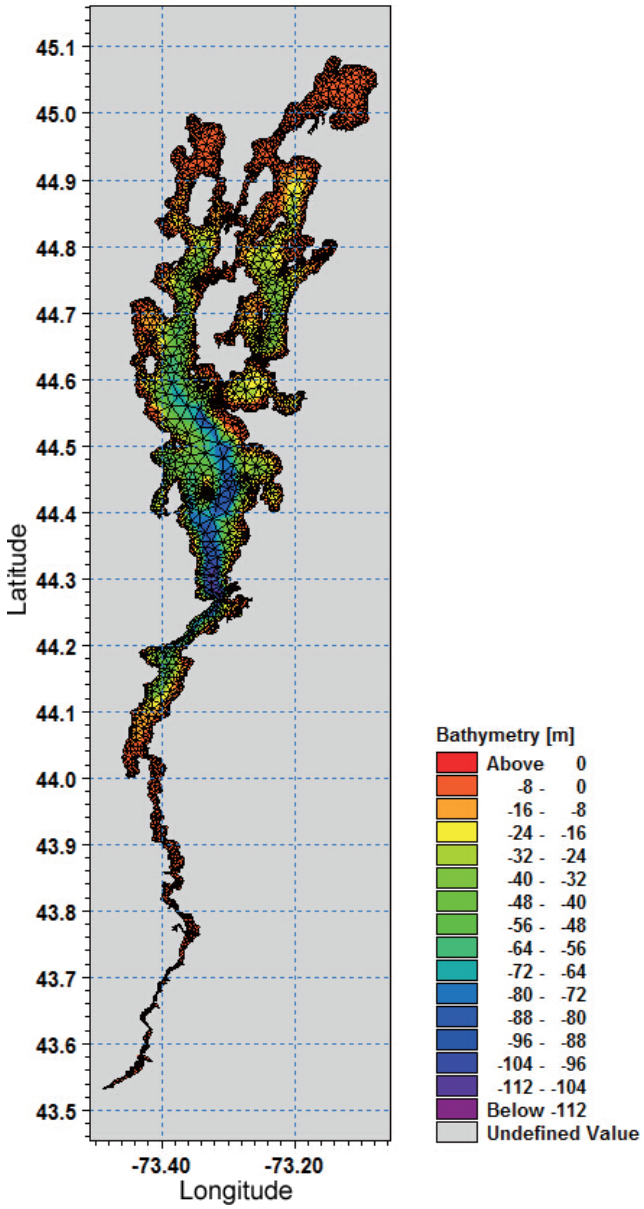


Figure 7-3 Mesh file of Lake Champlain for establishing the MIKE 21 model

7.3 Initial Conditions and Boundary Conditions

According to the surface elevation data of June 2008, initial surface elevation was set to 29.58m. Due to the lack of monitoring data for flow field and flow velocity, the initial velocity components (u and v) were determined according to studies of surface flow in Lake Champlain (McCormick et al., 2008), which indicated that the mean u-velocity and v-velocity were 1 cm/s and 5cm/s, respectively. U-velocity is positive to the east (x axis), and v-velocity is positive to the north (y axis).

The monthly average flow rates of the 23 tributaries around the lake in June 2008 were used as the boundary conditions in the hydrodynamic model.

For the transport model, the monthly average concentration of total phosphorus in June 2008 at the 15 lake stations was used as the initial condition for the calibration, as displayed in Figure 7-4 (a). The observed total phosphorus concentration in May 2011 was used as the initial condition for the model validation (Figure 7-4 (b)).

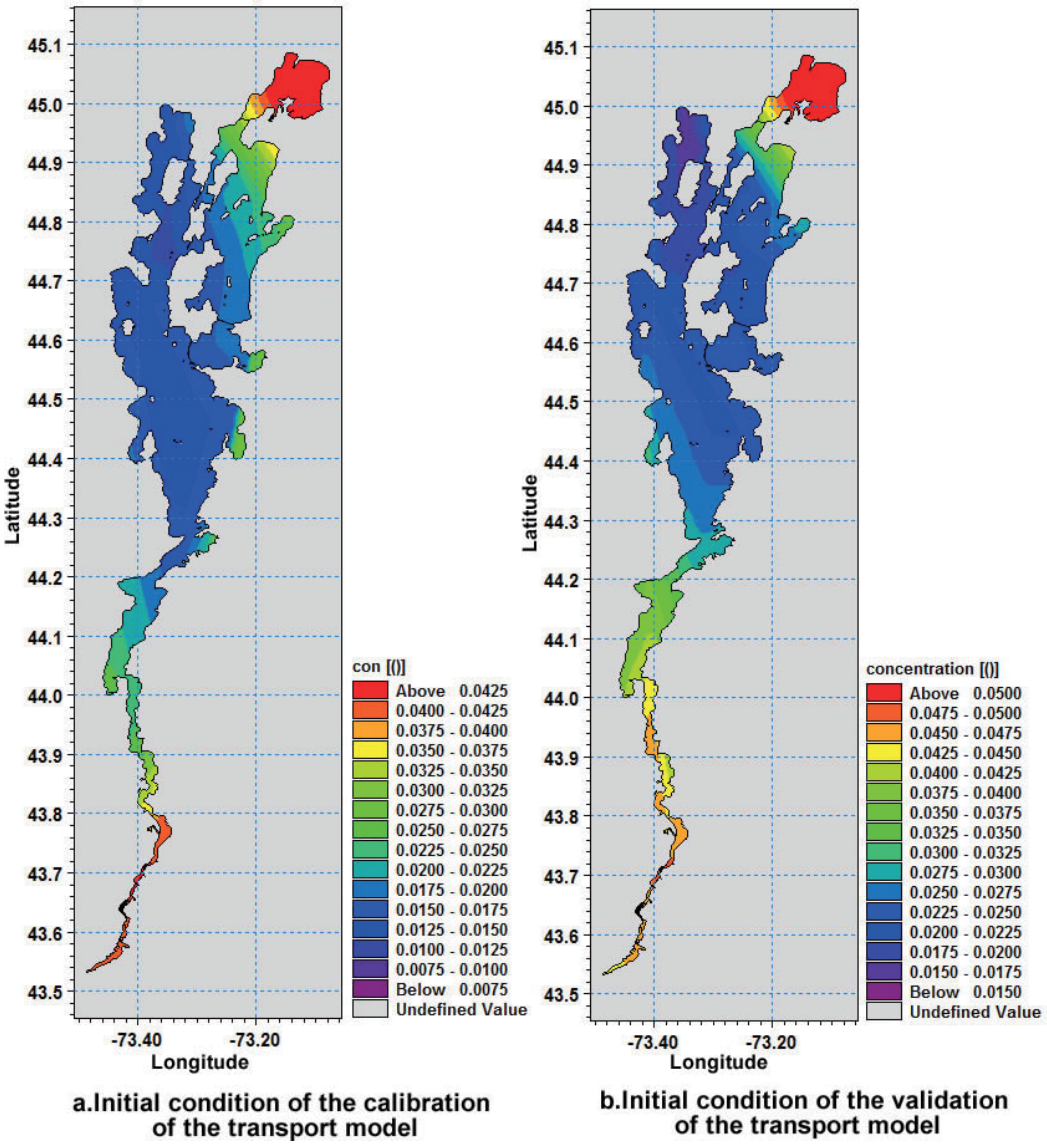


Figure 7-4 Initial conditions of the calibration and validation of the transport model

The monthly average TP concentrations in the discharge of each tributary in June 2008 were applied as the boundary conditions in the transport model.

7.4 Model Parameters and Calibration

The primary parameters included in the hydrodynamic model of MIKE 21 include the flood and dry depth, water density, eddy viscosity, bed resistance, Coriolis forcing, wind forcing, ice coverage, tidal potential, precipitation and evaporation,

wave radiation, and structures. Since the density gradient has minor effect on the two-dimensional shallow water equations, water density was not considered in this model. Lake Champlain is an inland lake, thus the tidal potential was excluded. Since the lake monitoring data used to validate the model results was collected during the non-frozen period, ice coverage was not considered herein. Structures were not taken into consideration because of the shortage of data about the structures in the lake area.

The flooding and drying is enabled to prevent the halting of the simulation when the total water depth becomes less than zero. The problem is reformulated when the water depth is less than the wetting depth. The reformulation is made by setting the momentum fluxes to zero and only considering the mass fluxes. The element is removed from the simulation if the water depth is less than the drying depth. The flooding depth is used to determine when an element is flooded and reentered into the calculation. The flooding depth, wet depth and drying depth in this study were set at 0.05m, 0.1m and 0.005m, respectively.

According to Equation 3.25 and Equation 3.26 in Section 3.6.2.2, the bed friction decreases with the increase of bed roughness value, i.e. Manning coefficient or Chezy coefficient. The suggested value for the Manning coefficient ($M = 32m^{1/3}/s$) corresponds to a flat bed of median grain size with the median grain diameter D_{50} = 0.1m. Variation in the Manning number for varying water depth and seabed type was summarized in Lambkin's report, which is shown in Figure 7-5.

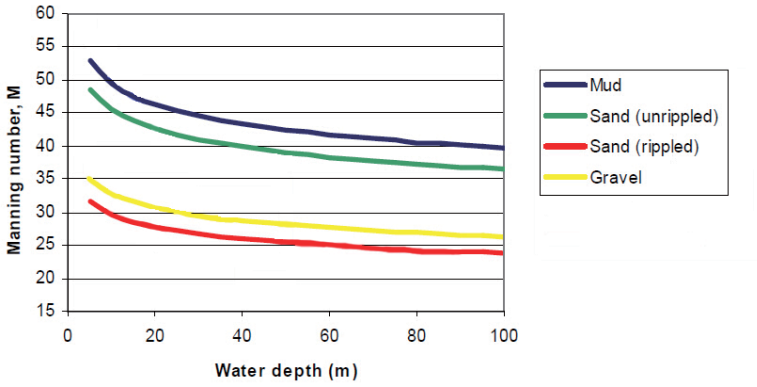


Figure 7-5 Variation in the Manning number for varying water depth and seabed type

The mean grain size of Lake Champlain sediments throughout the lake is generally in the fine silt range (10-25 microns) (Emily H. Dawson, 2008, Flora Weeks, 2012). Taking the water depth into consideration, the Manning number was initially set at 40 $m^{1/3}/s$ with reference to Figure 7-5 and then modified during calibration. The initial value of the Smagorinsky factor (Cs) of eddy viscosity was 0.25 for the whole lake domain and also modified through the calibration.

The Coriolis force is calculated by

$$f = 2\Omega \sin \varphi \tag{7.1}$$

where φ is latitude, $2\Omega = 1.458 \times 10^{-4} s^{-1}$. This parameter is constant at any fixed location. The Coriolis Forcing option in the hydrodynamic model was set as varying in the domain.

Wind speed and direction data collected at Burlington Airport, Vermont was loaded herein for the calculation of wind force. The wind friction coefficient is a weak function of wind speed. For medium and strong winds in open sea, the wind resistance coefficient is adopted at 0.0026 to get good results. Since the monthly average wind speed was relatively low, as shown in Figure 7-1, the wind friction coefficient was set to 0.0024 and calibrated in the simulation.

Time-series of precipitation collected at Burlington Airport, Vermont was loaded in the precipitation portion. By adding the evaporation rate converted from the water withdrawal rate to the natural evaporation rate, the total evaporation rate applied in the hydrodynamic model was 1.88 mm/day.

Since it is a 2D model, only horizontal dispersion was considered in the transport model. The dispersion coefficient formulation was applied herein with a constant dispersion coefficient over the model area, which was initially set to 0.03 m^2/sec and calibrated in the simulation. The decay rate was also set to a constant value ($4\times10^{-8}/s$) and calibrated in the simulation.

Atmospheric deposition rates of TP onto the lake surface through precipitation were determined based on the DFS BATHTUB modeling study of Lake Champlain (Vermont DEC and New York State DEC, 1997). The deposition rate of TP was 16.1 mg/m^2/yr and this amount was assumed to be constant over time in the current MIKE 21 modeling. The corresponding concentration of TP in each precipitation was obtained by dividing this deposition rate by the monthly precipitation rate. The average concentration of TP in precipitation during June 2008 to October 2012 was 0.03 mg/L, which was applied as the precipitation concentration of TP in the current model. The TP concentration in evaporation was assumed to be zero herein.

7.5 Model Calibration and Validation Results

The primary purpose of developing the MIKE 21 model was to simulate the total phosphorus concentration in Lake Champlain and compare the results with the BPANN model. Since the water environment capacity was not considered in this study and the flow field data was not available for the simulation period, the calibration and validation of water level and flow field in the lake were not considered. Total phosphorus concentration was the only consideration in this study.

7.5.1 Summary of model parameters

The hydrodynamic model and transport model were calibrated with input data from June 2008 to December 2010. The summary of setting values and calibrated values of model parameters were listed in Table 7-1.

Table 7-1 Summary of setting values and calibrated values of the hydrodynamic model and transport model

	Parameters	Values	
	No. of time steps	Calibration	944
		Validation	550
	Time step interval	86400 sec	
Hydrodynamic model	Drying depth	0.005m	
	Flooding depth	0.05m	
	Wetting depth	0.1m	
	Smagorinsky factor of eddy viscosity	0.27	
	Manning coefficient	38 $m^{1/3}/s$	
	Coriolis forcing	Varying in domain	
	Wind friction coefficient	0.0026	
	Precipitation	varying in time, constant in domain	
	Evaporation	1.88 mm/day	
Transport model	Dispersion coefficient	0.08 m^2/sec	
	Decay coefficient	$5.5*10^{-8}/s$	

7.5.2 Cross-comparison between the developed ANN model, MIKE 21, and field observation data

The transport model results contain the simulated daily concentrations of total phosphorus. The monthly average simulated concentrations of total phosphorus were sequentially calculated and compared with the observed total phosphorus concentrations as well as the results of the BPANN model developed in Section 5.5. The comparison of the results of MIKE 21 model, BPANN model and the observed concentrations at the 15 lake stations during the period of model calibration are displayed in Figure 7-6 and Figure 7-7, and the results of model validation are shown in Figure 7-8 and Figure 7-9. As displayed in these figures, both of the MIKE 21

model results and the BPANN model results were able to simulate the major trends of the total phosphorus concentration variations.

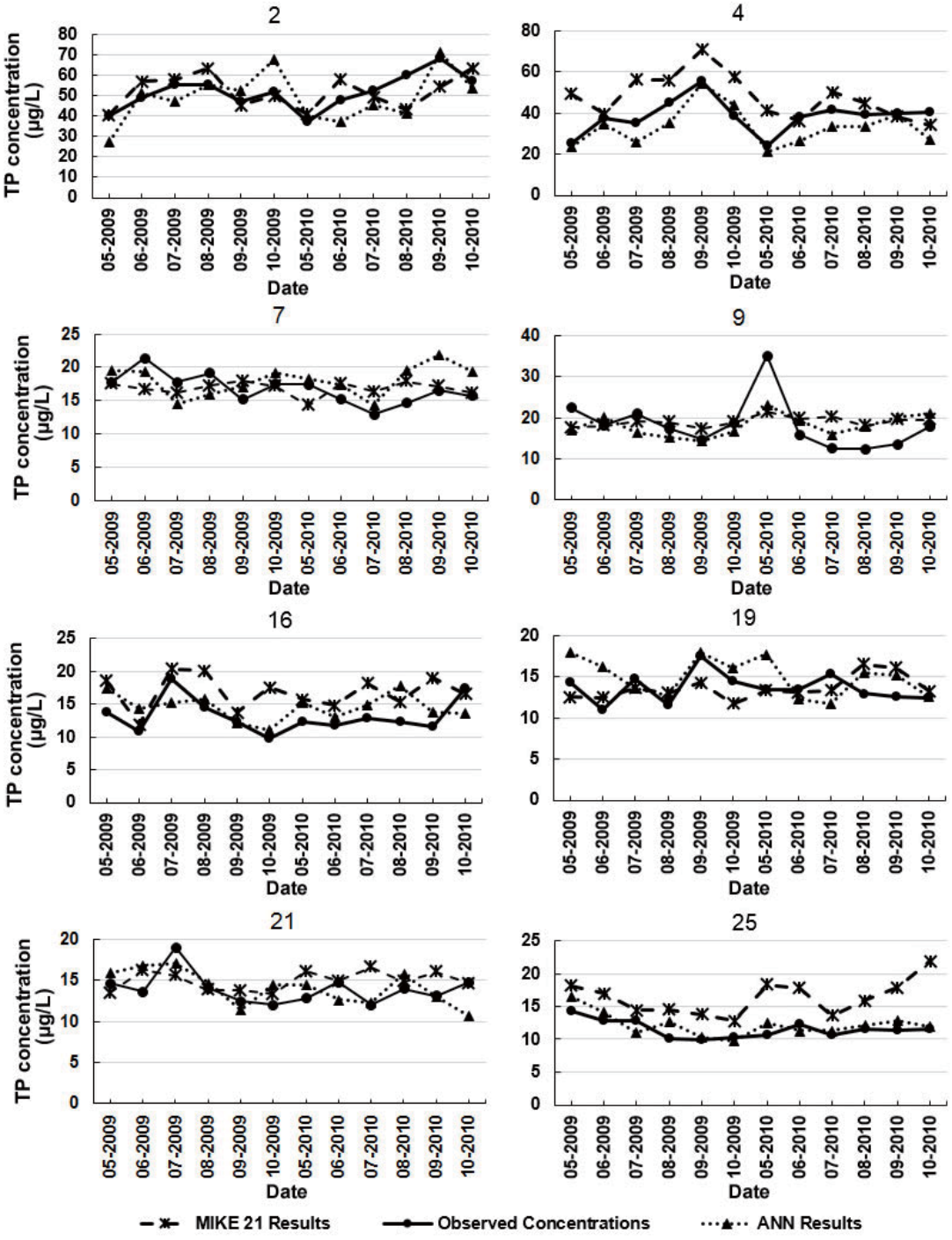


Figure 7-6 Comparison of results of MIKE 21 model calibration, BPANN model, and the observed concentrations (I)

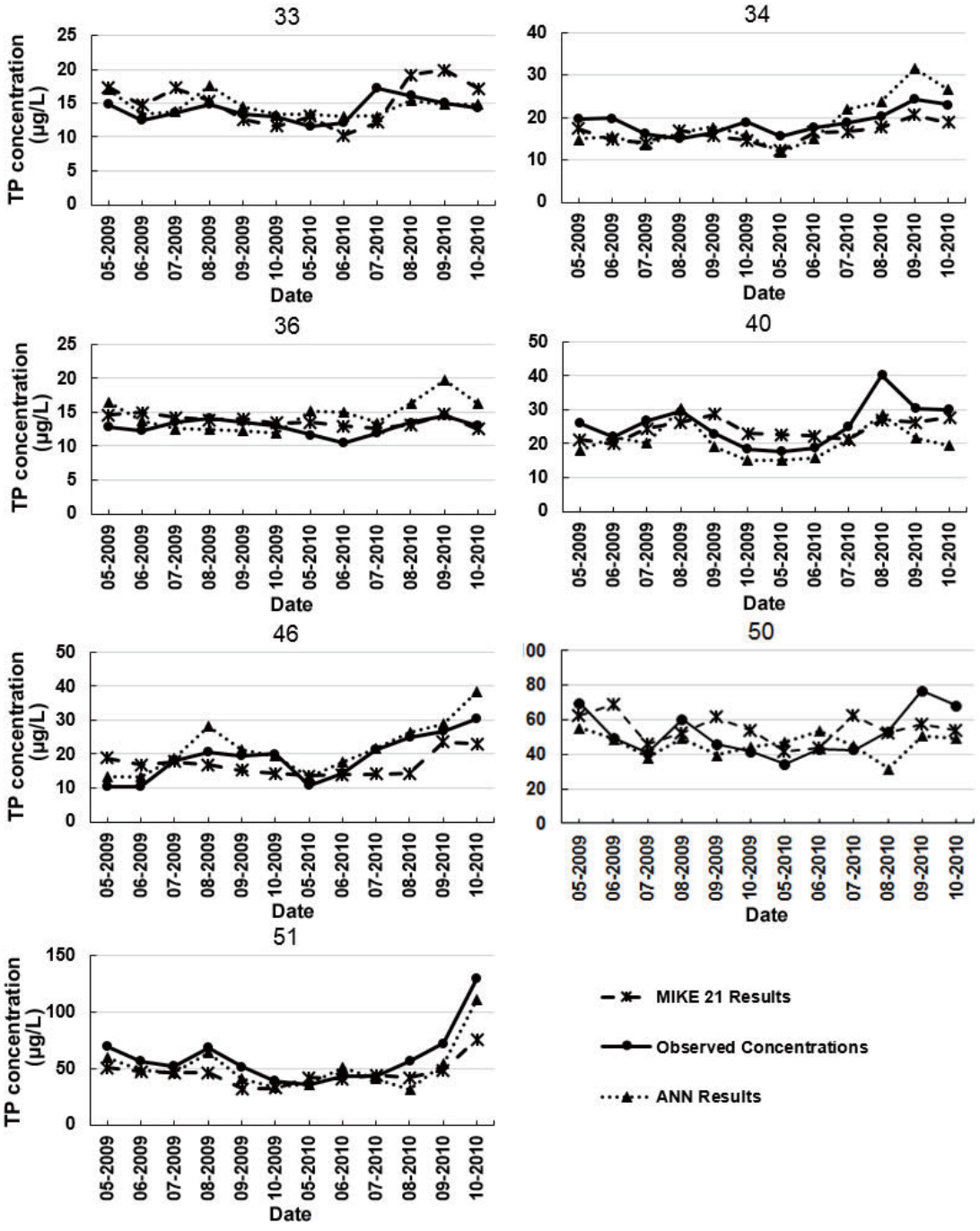


Figure 7-7 Comparison of results of MIKE 21 model calibration, BPANN model, and the observed concentrations (II)

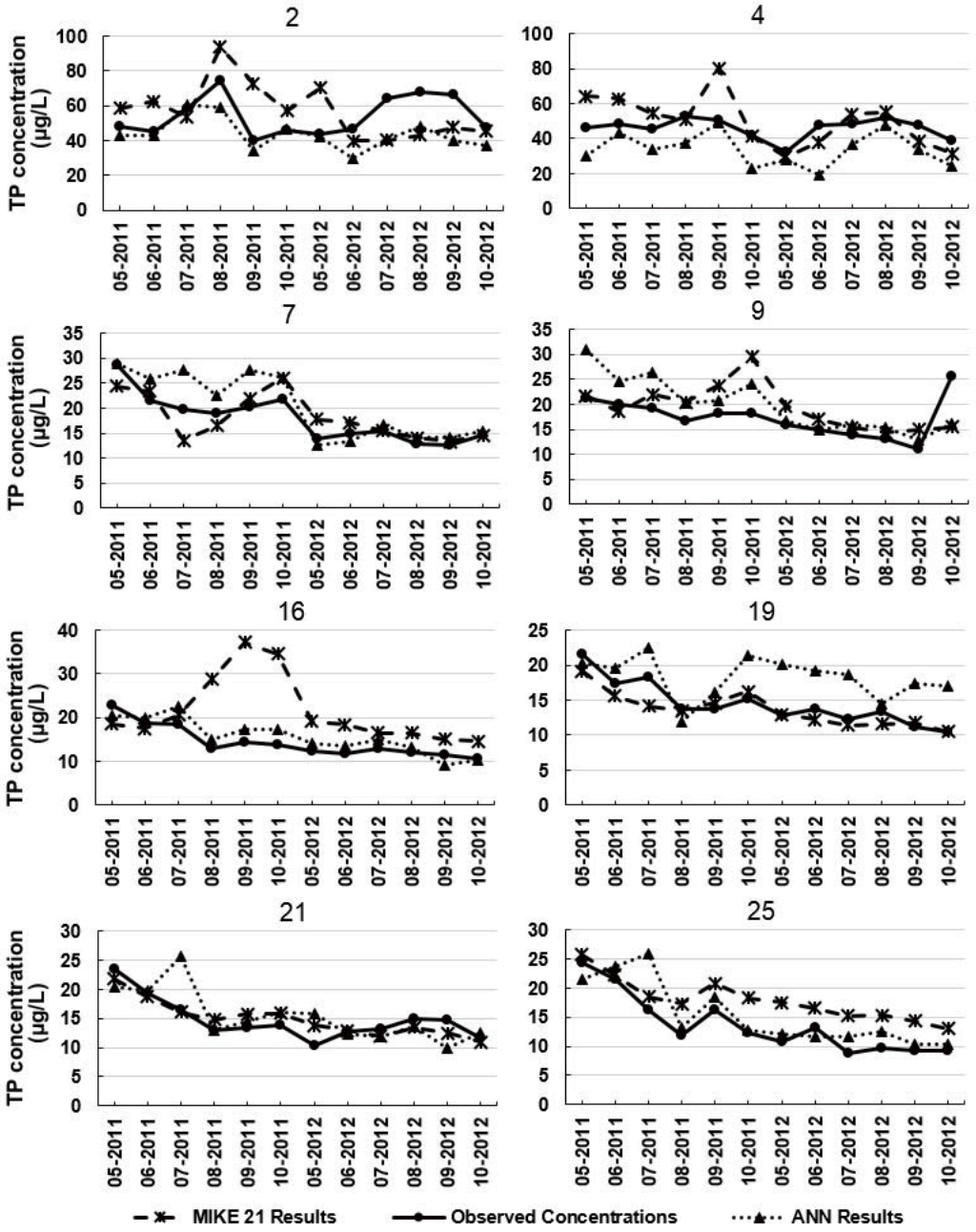


Figure 7-8 Comparison of results of MIKE 21 model validation, BPANN model, and the observed concentrations (I)

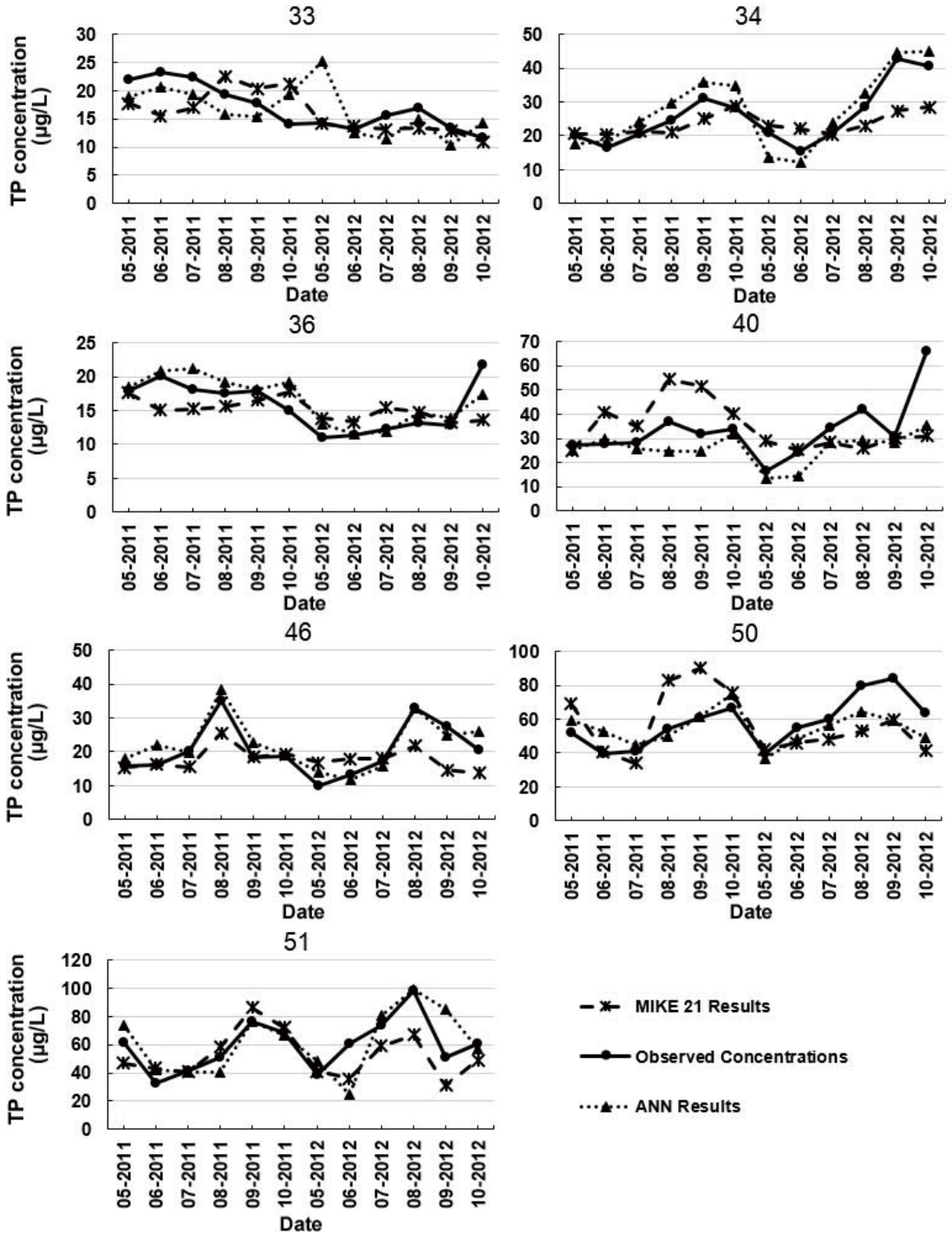


Figure 7-9 Comparison of results of MIKE 21 model validation, BPANN model, and the observed concentrations (II)

More statistical details regarding the comparison of model calibration and validation results are listed in Table 7-2 and Table 7-3, respectively. Coefficient of determination (R^2), linear regression equation, and the RSR calculated from model

results and observed TP concentrations were included in the statistical analysis. The coefficient of determination represents the percentage of variability of the observed concentrations that can be explained by the model results. RSR is an error index that combines the errors between the model results and the observed concentrations and the standard deviation of observed concentrations. The linear regression equation provides information regarding to the slope and intercept of the regression line.

Table 7-2 Statistical analysis of results of the MIKE 21 model calibration and the BPANN model

Station	MIKE 21 Results	ANN Results	Station	MIKE 21 Results	ANN Results
2	$R^2 = 0.49$ y = 0.48x+27 RSR = 1.0	$R^2 = 0.65$ y = 0.95x-0.26 RSR = 1.17	4	$R^2 = 0.48$ y = 0.64x+23 RSR = 1.72	$R^2 = 0.83$ y = 0.95x+3.4 RSR = 0.93
7	$R^2 =$ - 0.14 y = - 0.063x+18 RSR = 1.16	$R^2 = 0.21$ y = 0.21x+14 RSR = 1.39	9	$R^2 = 0.44$ y = 0.083x+18 RSR = 0.83	$R^2 = 0.50$ y = 0.21x+14 RSR = 0.81
16	$R^2 = 0.50$ y = 0.49x+10 RSR = 1.74	$R^2 = 0.33$ y = 0.24x+11 RSR = 1.16	19	$R^2 = 0.0036$ y = 0.0029x+14 RSR = 1.28	$R^2 = 0.28$ y = 0.35x+10 RSR = 1.02
21	$R^2 = 0.04$ y = 0.025x+15 RSR = 1.18	$R^2 = 0.43$ y = 0.48x+7.4 RSR = 1.20	25	$R^2 = 0.41$ y = 0.83x+6.8 RSR = 4.23	$R^2 = 0.67$ y = 0.91x+1.7 RSR = 1.14
33	$R^2 = 0.45$ y = 0.85x+3.2 RSR = 1.82	$R^2 = 0.41$ y = 0.38x+9.2 RSR = 1.09	34	$R^2 = 0.76$ y = 0.59x+5.2 RSR = 1.09	$R^2 = 0.83$ y = 1.7x-13 RSR = 1.33
36	$R^2 = 0.40$ y = 0.29x+10 RSR = 1.29	$R^2 = 0.12$ y = 0.26x+11 RSR = 2.81	40	$R^2 = 0.54$ y = 0.25x+18 RSR = 0.88	$R^2 = 0.79$ y = 0.6x+5.1 RSR = 1.03
46	$R^2 = 0.70$ y = 0.38x+11 RSR = 0.75	$R^2 = 0.94$ y = 1.1x+0.79 RSR = 0.60	50	$R^2 = 0.37$ y = 0.2x+43 RSR = 0.98	$R^2 = 0.40$ y = 0.2x+36 RSR = 1.04
51	$R^2 = 0.90$ y = 0.40x+22 RSR = 0.85	$R^2 = 0.93$ y = 0.8x+4.1 RSR = 0.5	Total response	$R^2 = 0.88$ y = 0.77x+6 RSR = 0.91	$R^2 = 0.94$ y = 0.8x+3.9 RSR = 0.72

Table 7-3 Statistical analysis of results of the MIKE 21 model validation and the BPANN model

Station	MIKE 21 Results	ANN Results	Station	MIKE 21 Results	ANN Results
2	R^2 = 0.016 y = 0.022x+56 RSR = 1.72	R^2 = 0.57 y = 0.45x+20 RSR = 1.26	4	R^2 = 0.64 y = 1.7x-27 RSR = 2.22	R^2 = 0.63 y = x - 14 RSR = 2.54
7	R^2 = 0.78 y = 0.74x+4.9 RSR = 0.65	R^2 = 0.90 y = 1.2x-1.7 RSR = 0.84	9	R^2 = 0.30 y = 0.34x+13 RSR = 1.37	R^2 = 0.55 y = 0.77x+6.5 RSR = 1.36
16	R^2 = 0.038 y = 0.078x+20 RSR = 3.00	R^2 = 0.86 y = 0.92x+2.4 RSR = 0.65	19	R^2 = 0.88 y = 0.69x+3.7 RSR = 0.56	R^2 = 0.48 y = 0.45x+12 RSR = 1.57
21	R^2 = 0.86 y = 0.73x+4.1 RSR = 0.51	R^2 = 0.59 y = 0.74x+4.5 RSR = 1.06	25	R^2 = 0.73 y = 0.47x+13 RSR = 1.48	R^2 = 0.84 y = 0.92x+2.8 RSR = 0.69
33	R^2 = 0.21 y = 0.24x+12 RSR = 1.40	R^2 = 0.39 y = 0.43x+9.2 RSR = 1.16	34	R^2 = 0.82 y = 0.3x+16 RSR = 0.79	R^2 = 0.94 y = 1.2x-3.7 RSR = 0.52
36	R^2 = 0.29 y = 0.13x+13 RSR = 0.98	R^2 = 0.83 y = 0.79x+4.2 RSR = 0.63	40	R^2 = 0.039 y = 0.032x+34 RSR = 1.28	R^2 = 0.75 y = 0.39x+13 RSR = 0.94
46	R^2 = 0.58 y = 0.25x+13 RSR = 0.90	R^2 = 0.93 y = 0.94x+2.9 RSR = 0.44	50	R^2 = 0.32 y = 0.4x+33 RSR = 1.35	R^2 = 0.65 y = 0.45x+29 RSR = 0.80
51	R^2 = 0.66 y = 0.59x+18 RSR = 0.88	R^2 = 0.70 y = 0.85x+10 RSR = 0.89	Total response	R^2 = 0.84 y = 0.83x+5.3 RSR = 1.07	R^2 = 0.90 y = 0.83x+4.6 RSR = 0.81

As for the results of each station, it is difficult to draw a conclusion about which model performed better, because the result of each station differed. Meanwhile, inconsistency of R^2 and RSR values occurred at some stations. For example, the result of MIKE 21 model calibration at station 2 showed a smaller R^2 value and lower RSR value than the ANN model results, which demonstrates that a lower percentage of variability of the observed values can be explained by the MIKE 21 model results, but the MIKE 21 model results had smaller errors compared to the observed concentrations. For the ANN model results at this station, both the R^2 and RSR values were higher than that of MIKE 21 model, indicating that the ANN model results have

greater similarity with variation trends in the observed values, however, the difference between them are larger than the results of the MIKE 21 model.

Since a small number of data was included in the comparison of model results and the observed concentrations, i.e., for each lake station, only 12 data were available for comparison. Furthermore, the average values of the observed concentrations of each station were included in the calculation of RSR values, and the statistical results in Table 7-2 and Table 7-3 tend to be sensitive to data variations, thus a slight difference in the data may result in different statistical results. Therefore, the RSR values of the BPANN model results of each station were larger than that of the results all over the lake, which were presented in Section 5.5.

For the calibration results of MIKE 21 all over the lake, the coefficient of determination was 0.88, RSR value was 0.91, the slope of the linear regression line was 0.77 and the intercept was 6. Meanwhile the BPANN model results of the same period have higher values of the coefficient of determination and the slope of the regression line, as well as lower values of RSR and the intercept of the regression line, indicating that the performance of the MIKE 21 model during the calibration period was not as good as the BPANN model. Similarly, the BPANN model also had better performance compared with the validation results of the MIKE 21 model. Even though the RSR values of the two models were larger than 0.7, taking all other statistical results into account, the model results were acceptable in the simulation of TP concentration in Lake Champlain.

In addition, linear regression analysis was performed between all the calibration and validation results of the MIKE 21 model and the observed concentrations. The regression result was compared with that of the testing data set of the BPANN model, which was conducted in the same period of data. As presented in Figure 7-10, the

coefficient of determination of the total response of the MIKE 21 model was 0.86, the slope and intercept of the linear regression line was 0.81 and 5.5, respectively. The RSR value was 0.51. The results of BPANN model showed a higher R^2 value, which was 0.92. The slope of the regression line was almost the same as that of the MIKE 21 model, while the intercept and the RSR value were smaller. Consequently, the performance of the BPANN model was marginally better than the MIKE 21 model regarding all the simulated results in model calibration and validation, which was consistent with the statistical results of the respective calibration and validation data set.

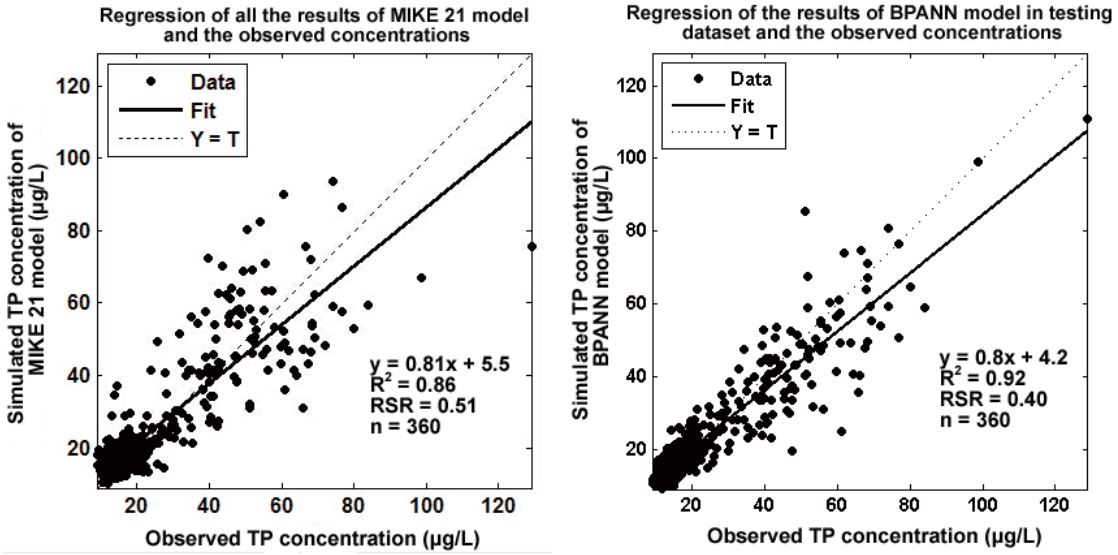


Figure 7-10 Comparison of simulated TP concentration of MIKE 21 model in the calibration and validation and results of the BPANN model in the testing data set

7.6 Discussion

The performance of the ANN model has been compared with many other models in various disciplines. The multilayer artificial neural network model performed better than the soil and water assessment tool (SWAT) model in the simulation of the monthly sediment yield in a watershed (Singh et al., 2012). The performance

efficiency of ANN in the prediction of stream water temperature was found to be the highest among the ANN model, the multiple regression analysis, and the chaotic non-linear dynamic algorithms (Sahoo et al., 2009). In the simulation of water levels of a river, the results obtained from the ANN model were much better than that of the MIKE 11HD model as indicated by the values of the goodness of fitness indices (Panda et al., 2010). On the contrary, for the prediction of wave height in Lake Superior, the error statistics of model trees and feed-forward back propagation ANNs were similar but model trees were marginally more accurate (Etemad-Shahidi and Mahjoobi, 2009).

In this study, the performance of the BPANN model was better than that of the MIKE 21 model as indicated by the error statistics. However, uncertainties were introduced into the MIKE 21 model since many assumptions and simplifications were applied in the model construction, which would possibly influence the model's performance. The MIKE 21 model is a 2D model that assumes the water is vertically well mixed. However, the water depth has a great amplitude of variation in the Lake Champlain, the 2D model may not be sufficient to simulate the internal current and water interactions between different layers.

Due to the lack of data about characteristics of sediment in the lake, the Manning number could not be interpolated according to the distribution of sediment particle sizes. Therefore, the Manning number was estimated and assumed to be constant in the whole lake domain. The biological interactions between plankton and total phosphorus, as well as the deposition and resuspension of particulate phosphorus were simplified to a single value as decay rate in the transport model.

The accuracy and sufficiency of the input data have significant influence on the results of the MIKE 21 model. In this study, five kinds of boundary conditions should

be considered in the hydrodynamic model, including tributary inflow into the lake, lake outlets, point and non-point discharge directly into the lake, water withdrawal for industry and agriculture, and the irrigation return water and rainfall and evaporation. Considering the availability of data, the non-point discharge and irrigation return water were ignored, and the point discharge was assumed to be included in the tributary inflow. Water intake rate and the evaporation rate measured in 1991 were applied in the model, which was not up to date and may not be able to reflect the current situation. Since wind force was the primary force of water flow in the MIKE 21 model in this study, an adequate wind field is of prime importance. The wind stress in MIKE 21 model is calculated using the wind speed 10 m above the water surface (MIKE by DHI, 2012). It is necessary to verify how well the lake surface wind can be represented by the observations at the Burlington Airport.

The performance of the MIKE 21 model might be improved if the above uncertainties could be addressed properly. However, the required time and effort to establish the model should also be considered in the comparison of the MIKE 21 model and the ANN model. As for the former model, many parameters that are related to the involved physical and biological processes in the model should be considered. The coefficients applied in the hydrodynamic and transport model need lots of calibration based on the errors between the model results and observed values. The process of data collection and model calibration need great effort and may be time-consuming. Moreover, a high-performance computer is required to reduce the run time of the MIKE 21 model. As for the ANN models, by contrast, no specific type of data is required and the acceptable ANN models can be established by training with the data set that is correlated with the target data. In addition, the results of ANN models can be achieved within minutes, which is much less than the running time of

the MIKE 21 model. Therefore, much less effort is needed in the application of ANN models.

In summary, taking all the factors discussed above into consideration, it is further verified that the artificial neural network model is more applicable for the water quality prediction of Lake Champlain in this study.

7.7 Summary

This chapter provided a case study of simulation of the total phosphorus concentrations in Lake Champlain by the MIKE 21 Flow Model FM. Water chemistry data, geographical data and meteorological data in the Lake Champlain area were collected and applied to the hydrodynamic model and transport model involved in the simulation. By comparing the simulation results with the observed total phosphorus concentrations and the results of the developed ANN model, the MIKE 21 model was able to provide acceptable simulation results, but the performance was slightly worse than the ANN model. Additionally, the required time and effort in applying the MIKE 21 model and the ANN model were compared and the ANN model developed was more applicable for Lake Champlain, which has complex geographic conditions that introduced more difficulties in establishing the hydrodynamic and transport models.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

This thesis was dedicated to developing a GIS-based system named as GIS-SWQAM, which includes a fuzzy-set based water quality assessment model, an ecological risk assessment model, an artificial neural network model for water quality prediction, and the integration of the spatial information technology with the surface water quality risk assessment models and water quality prediction model. An integrated system with a graphical user interface was developed in the C# programming language to combine the fuzzy synthetic assessment model, the ecological risk model, the ANN model, the GIS, and the database. In addition, the MIKE 21 hydrodynamic and transport model was established and the results were compared with the ANN model in the GIS-SWQAM system.

A series of research tasks were implemented in this thesis:

(a) GIS-based fuzzy synthetic water quality assessment models (MWQ and LWQ) have been developed through the joint coupling between GIS and a database. The models developed extend the traditional numerical results of the assessment models by delivering results with spatial references. By integrating GIS as the post processor of the models, the developed system enables engineers and decision makers to better understand the spatial distribution of water quality risks. A user-friendly GUI was further developed to manage input information, process data, and visually present the results.

The MWQ module integrates the ArcGIS Engine with a risk assessment model that includes fuzzy synthetic water quality evaluation, eutrophication risk assessment and heavy metal risk evaluation. The MWQ module was applied to assess the marine water quality near offshore platforms in the

northern part of Liaodong Bay, China. The LWQ module combines fuzzy synthetic water quality assessment and trophic status evaluation with the ArcGIS Engine, and it was applied in the water quality risk assessment of Lake Champlain by compiling with the local water quality standards. The results obtained through these two modules indicated the overall water quality conditions of the study area by both numerical and spatially referenced results. The water quality risk contour maps generated by the system depict the distribution of the integrated water quality conditions in the study area. The primary pollutants that contributed to the water quality variation were analyzed on the basis of the generated maps. The ecological risk of eutrophication and heavy metal contamination was also evaluated and analyzed.

The results of the MWQ module indicated that the overall water quality of the study area was good in Oct. 2005, May 2006, and Sept. 2007 and fair in May 2007. It was concluded that the high nutrient concentrations were the main reason that contributed to the water quality deterioration. The risk maps generated indicated elevated concentrations of oil and phosphorus near the platforms; higher concentration of inorganic nitrogen was found in the landward direction. The eutrophication risk level in Oct. 2005, May 2006, and Sept. 2007 corresponded to oligotrophic, light eutrophic and mesotrophic conditions, while mesotrophic and high eutrophic conditions occurred in May 2007. Only a slight heavy metal risk was identified for the four investigations.

The LWQ module results showed that the fuzzy water quality index at the fifteen lake stations in 2012 mostly fell into the interval of “Poor” classification. The fuzzy water quality index values were higher in July to September under the influence of temperature and dissolved oxygen concentration. Due to the great weight of total phosphorus concentration, the

distribution of the fuzzy water quality index was consistent with that of the total phosphorus concentration, i.e., the higher concentration of total phosphorus concentration and higher fuzzy water quality index representing worse water quality distributed in the northern and southern regions of Lake Champlain. The spatial distribution of trophic state index agreed fairly well with the fuzzy water quality index. The eutrophication arose in the lake regions that had worse water quality, which further proved that total phosphorus was the main pollutant in Lake Champlain.

The developed MWQ and LWQ modules provide reliable and intuitive results regarding the water quality distribution and offers an efficient tool for engineers and decision makers.

(b) Developed a prediction model that is able to provide variation information for selected water quality variables, the ability of a back-propagation artificial neural network model developed using Matlab software was initially investigated in this study to predict the chlorophyll-a concentration in Lake Champlain. The performance of the BPANN model was examined against observed chlorophyll-a concentration in terms of R^2 and RSR. Seven scenarios including different combinations of water quality variables were established to select the optimal set of input variables. The topology with seven neurons in the input layer, twelve neurons in the hidden layer, and one neuron in the output layer provided the best performance in the simulation of chlorophyll-a concentration among all the scenarios. The R2 values of the best performed model in the training set, validation set, and testing set were 0.82, 0.93, and 0.81, respectively. The corresponding RSR values of the three data sets were 0.62, 0.38, and 0.53, respectively. The results indicate that the developed BPANN model could provide satisfactory results in predicting the chlorophyll-a concentration in Lake Champlain.

Subsequently, the algorithm of the artificial neural network was compiled using C# programming code and integrated with the ArcGIS Engine and a database to build the ANN module for water quality prediction. The developed ANN module was applied to predict the total phosphorus concentration in Lake Champlain. The prediction results of TP concentrations are verified with the lake water monitoring data in terms of R^2 and RSR. The model with five input neurons, ten hidden neurons and one output neuron was determined to be the best performing model. For the training, validation, testing data sets and all data, the R^2 and RSR were correspondingly 0.90 and 0.44, 0.92 and 0.39, 0.92 and 0.40, 0.91 and 0.42, which indicate that the developed ANN model is able to provide satisfactory predictions and can be used to provide preliminary information on the variation of water quality parameters. The generated distribution map of predicted TP concentrations agrees well with that of the observed concentrations, providing a visualization support for lake managers.

(c) A hydrodynamic model in conjunction with a transport model was established to simulate the total phosphorus concentrations in Lake Champlain with the MIKE 21 Flow Model FM model, and the simulation results were cross-compared with the observed total phosphorus concentrations and the results of the developed ANN model. The R^2 and RSR values of the MIKE 21 model were 0.86 and 0.51, indicating that the MIKE 21 model was able to provide acceptable simulation results, but the performance was slightly inferior than the ANN model. Moreover, according to the comparison of required time and effort in the application of the MIKE 21 model and the ANN model, the developed ANN model was better for the application in Lake Champlain, which has complex geographic conditions that introduced more difficulties in establishing the hydrodynamic and transport models. This further verifies that

the developed ANN model is a reliable tool to assess the lake eutrophication and to help managing lake water quality.

In conclusion, the research work presented in this thesis has demonstrated that it is useful to apply the spatial information technology in effective surface water quality assessment and management. The research results indicate that the GIS-based integrated system (GIS-SWQAM) developed in this thesis provides effective support for the surface water quality management and control. The developed system and modeling tools can be also applied to surface water management in other regions.

8.2 Contributions

(a) Developed GIS-based water quality and ecological risk assessment models by joint coupling the ArcGIS Engine and a database. The developed MWQ module and LWQ module provide reliable and intuitive results regarding the spatial distribution of water quality risk to support decision-making and management on surface water quality.

(b) A back propagation artificial neural network model has been developed using Matlab software and a first attempt was made at exploring the efficiency of the developed ANN model in the production of chlorophyll-a concentration in Lake Champlain, where only some physical-based models were applied for the prediction of water quality parameters. The developed ANN model provided acceptable results in the chlorophyll-a prediction, which is an important contribution to the limited studies available on the application of data-driven models in Lake Champlain.

(c) Developed a GIS-based water quality prediction model that integrates an artificial neural network, the ArcGIS Engine and a database. The developed ANN module is capable of providing satisfactory results in the prediction of water quality parameters, as well as intuitively displaying the prediction results through spatial

distribution maps, which can be used to provide preliminary information about the variation of water quality parameters.

(d) Established a hydrodynamic model and a transport model using MIKE 21 software to simulate the nutrient concentration in Lake Champlain, and the results were compared with that of the developed ANN module, which contributes to the studies related to the comparison between physical-based models and data-driven models.

(e) Developed a graphical user interface which integrates the developed MWQ, LWQ, and ANN modules. The developed MWQ module, LWQ module and ANN module were successfully applied to case studies in China and Lake Champlain area. The developed system can also be applied to other countries.

8.3 Recommendations for Future Work

Future studies may focus on:

1. The data input functions of the MWQ and LWQ module can be further developed to achieve batch import of the data to provide more convenient use of the system.
2. In the current system, the water quality criteria in the MWQ and LWQ module can only be modified by editing the program codes, which is not convenient for the users not familiar with the programming language. In the future studies, a variety of water quality criteria from different countries could be included in the database of the system and able to be selected by users through the interface. So the system can be applied flexibly to other regions.
3. In the current study, four water quality parameters were included in the water quality assessment of Lake Champlain. More parameters could be considered

in the future, if applicable, to perform a more comprehensive assessment of the water quality in Lake Champlain.

4. The graphical user interface will need continuous improvement throughout the application of models to new case studies.

References:

Al-Sabhan, W., M. Mulligan and G. A. Blackburn (2003). "A real-time hydrological model for flood prediction using GIS and the WWW." Computers, Environment and Urban Systems 27 (1): 9-32.

Almeida, S. F. P., C. Elias, J. Ferreira, E. Tornés, C. Puccinelli, F. Delmas, G. Dörflinger, G. Urbanič, S. Marcheggiani, J. Rosebery, L. Mancini and S. Sabater (2014). "Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise." Science of The Total Environment 476–477: 768-776.

Bedri, Z., A. Corkery, J. J. O'Sullivan, M. X. Alvarez, A. C. Erichsen, L. A. Deering, K. Demeter, G. M. P. O'Hare, W. G. Meijer and B. Masterson (2014). "An integrated catchment-coastal modelling system for real-time water quality forecasts." Environmental Modelling & Software 61: 458-476.

Bowden, G. J., J. B. Nixon, G. C. Dandy, H. R. Maier and M. Holmes (2006). "Forecasting chlorine residuals in a water distribution system using a general regression neural network." Mathematical and Computer Modelling 44 (5–6): 469-484.

Brandmeyer, J. E. and H. A. Karimi (2000). "Coupling methodologies for environmental models." Environmental Modelling & Software 15 (5): 479-488.

Çalışkan, A. and Ş. Elçi (2009). "Effects of selective withdrawal on hydrodynamics of a stratified reservoir." Water resources management 23 (7): 1257-1273.

Carlson, R. E. (1977). "A trophic state index for lakes1." Limnology and oceanography 22 (2): 361-369.

Centre d'expertise hydrique du Québec. "History levels and flows of different hydrometric stations." Retrieved 02-15, 2015, from http://www.cehq.gouv.qc.ca/hydrometrie/historique_donnees/default.asp.

Champlain Lake long-term water quality and biological monitoring project. Retrieved 08/26, 2014, from http://www.watershedmanagement.vt.gov/lakes/htm/lp_longterm.htm.

Chen, W.-B., W.-C. Liu and M.-H. Hsu (2012). "Comparison of ANN approach with 2D and 3D hydrodynamic models for simulating estuary water stage." Advances in Engineering Software 45 (1): 69-79.

Chibole, O. K. (2013). "Modeling River Sosiani's water quality to assess human impact on water resources at the catchment scale." Ecohydrology & Hydrobiology 13 (4): 241-245.

Cho, S., B. Lim, J. Jung, S. Kim, H. Chae, J. Park, S. Park and J. K. Park (2014). "Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network." Measurement 53: 224-233.

Daniel L. Mendelsohn and H. Rines (1995). "Development and Application of a Full Phosphorus Cycle Water Quality Model to Lake Champlain." JOURNAL OF WATER MANAGEMENT MODELING (R183-15): 231-258.

Daniel Mendelsohn, Tatsusaburo lsaji and H. Rines (1996). Hydrodynamic and Water Quality Modeling of lake Champlain, Applied Science Associates, Inc.

Debaine, F. and M. Robin (2012). "A new GIS modelling of coastal dune protection services against physical coastal hazards." Ocean & Coastal Management 63: 43-54.

Department of Environmental Conservation, A. O. N. R., State of Vermont (2014). Vermont Water Quality Standards Environmental Protection Rule Chapter 29(a).

Deus, R., D. Brito, M. Mateus, I. Kenov, A. Fornaro, R. Neves and C. N. Alves (2013). "Impact evaluation of a pisciculture in the Tucuruí reservoir (Pará, Brazil) using a two-dimensional water quality model." Journal of Hydrology 487: 1-12.

DHI Software. (2014). "MIKE 21." Retrieved July 2nd, 2014, from http://www.mikepoweredbydhi.com/products/mike-21.

DHI Software for Water Environment (2014), MIKE 21 & MIKE 3 Flow Model FM Transport Module Short Description. Retrieved July 2nd, 2014, from https://www.mikepoweredbydhi.com//media/shared%20content/mike%20by%20dhi/flyers%20and%20pdf/productdocumentation/short%20descriptions/mike213_fm_tr_short_description.pdf.

DHI Water & Environment (2006). "MIKE 21 & MIKE 3 FLOW MODEL FM Hydrodynamic Module Short Description ". Retrieved July 4th, 2014, from http://www.mikepoweredbydhi.com/~/media/Microsite_MIKEbyDHI/Publications/PDF/Short%20descriptions/MIKE213_FM_HD_Short_Description.ashx.

Dieterle, F. J. (2003). Multianalyte quantifications by means of integration of artificial neural networks, genetic algorithms and chemometrics for time-resolved analytical data. Doctoral thesis, University of Tübingen.

Dogan, E., B. Sengorur and R. Koklu (2009). "Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique." Journal of Environmental Management 90(2): 1229-1235.

Doulgeris, C., P. Georgiou, D. Papadimos and D. Papamichail (2012). "Ecosystem approach to water resources management using the MIKE 11 modeling system in the Strymonas River and Lake Kerkini." Journal of Environmental Management 94(1): 132-143.

Emerson, D. G., A. V. Vecchia and A. L. Dahl (2005). Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota, US Department of the Interior, US Geological Survey.

Emily H. Dawson (2008). Broad climate variability based on seismic stratigraphy and sediment cores in Willsboro Bay, Lake Champlain. Middlebury, VT, Middlebury College**:** 83.

Environment Canada (2004). Canadian Guidance Framework for the Management of Phosphorus in Freshwater system, National Guidelines and Standards Office.

Etemad-Shahidi, A. and J. Mahjoobi (2009). "Comparison between M5′ model tree and neural networks for prediction of significant wave height in Lake Superior." Ocean Engineering 36(15–16): 1175-1181.

Facey, D. E., J. E. Marsden, T. B. Mihuc and E. A. Howe (2012). "Lake Champlain 2010: A summary of recent research and monitoring initiatives." Journal of Great Lakes Research 38: 1-5.

Flora Weeks (2012). Multifaceted Analysis of Drift B, Lake Champlain: Unpublished Senior Thesis, Middlebury College, Middlebury, VT, 135 p.

Galvez-Cloutier, R. and M. Sanchez (2007). "Trophic status evaluation for 154 lakes in Quebec, Canada: monitoring and recommendations." Water Qual. Res. J. Can 42 (4): 252-268.

Gharibi, H., A. H. Mahvi, R. Nabizadeh, H. Arabalibeik, M. Yunesian and M. H. Sowlat (2012). "A novel approach in water quality assessment based on fuzzy logic." Journal of Environmental Management 112: 87-95.

Ghebremichael, L., T. Veith and M. Watzin (2010). "Determination of critical source areas for phosphorus loss: Lake Champlain basin, Vermont." Transactions of the ASABE 53 (5): 1595-1604.

Great Lakes Science Advisory Board, Great Lakes Water Quality Board and International Air Quality Advisory Board and Health Professionals Task Force to the International Joint Commission (IJC) (2009). The Impact of Urban Areas on Great Lakes Water Quality. Windsor, Ontario, Canada.

Halls, J. N. (2003). "River run: an interactive GIS and dynamic graphing website for decision support and exploratory data analysis of water quality parameters of the lower Cape Fear river." Environmental Modelling & Software 18(6): 513-520.

HaSS. (2014). "Modelling Data Using Correlation and Regression." Retrieved September 20th, 2014, from http://www.strath.ac.uk/aer/materials/4dataanalysisineducationalresearch/unit4/correlationsdirectionandstrength/.

He, B., T. Oki, F. Sun, D. Komori, S. Kanae, Y. Wang, H. Kim and D. Yamazaki (2011). "Estimating monthly total nitrogen concentration in streams by using artificial neural network." Journal of Environmental Management 92(1): 172-177.

Heaton Research (2010). Introduction to Encog 2.5 for C#. Retrieved September 21st, 2014 from http://www.heatonresearch.com/dload/ebook/IntroductionToEncogCS.pdf.

Heaton Research. (2014). "Encog Machine Learning Framework." Retrieved September 21st from http://www.heatonresearch.com/encog.

Helmer, R. and I. Hespanhol (2011). Water Pollution. Environmental Health Hazards. J. M. S. Encyclopedia of Occupational Health and Safety. Geneva, International Labour Organization, Encyclopaedia of Occupational Health & Safety. 53.

HJ-442 (2008). Specification for offshore environmental monitoring, Ministry of Enviriomental Protection of the People's Republic of China.

Huang, B. and B. Jiang (2002). "AVTOP: a full integration of TOPMODEL into GIS." Environmental Modelling & Software 17(3): 261-268.

Huo, S., Z. He, J. Su, B. Xi and C. Zhu (2013). "Using Artificial Neural Network Models for Eutrophication Prediction." Procedia Environmental Sciences 18(0): 310-316.

Hurley, T., R. Sadiq and A. Mazumder (2012). "Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality." Water Research 46(11): 3544-3552.

Icaga, Y. (2007). "Fuzzy evaluation of water quality classification." Ecological Indicators 7(3): 710-718.

Jeong, S., K. Yeon, Y. Hur and K. Oh (2010). "Salinity intrusion characteristics analysis using EFDC model in the downstream of Geum River." Journal of Environmental Sciences 22(6): 934-939.

Jiang, J., P. Wang, W.-s. Lung, L. Guo and M. Li (2012). "A GIS-based generic real-time risk assessment framework and decision tools for chemical spills in the river basin." Journal of Hazardous Materials 227–228: 280-291.

John Zambrano and S. Stoner (1998). Ambient water quality standards and guidance values and groundwater effluent limitations. Division of Water Technical and Operational Guidance Series (1.1.1).

Kannel, P. R. and T. Y. Gan (2013). "Application of WASP for Modelling and Management of Naphthenic Acids along Athabasca River, Alberta, Canada." Water Air and Soil Pollution 224(11): 16.

Karakaya, N., F. Evrendilek, K. Gungor and D. Onal (2013). "Predicting Diel, Diurnal and Nocturnal Dynamics of Dissolved Oxygen and Chlorophyll-a Using Regression Models and Neural Networks." Clean-Soil Air Water 41(9): 872-877.

Kazi, T. G., M. B. Arain, M. K. Jamali, N. Jalbani, H. I. Afridi, R. A. Sarfraz, J. A. Baig and A. Q. Shah (2009). "Assessment of water quality of polluted lake using multivariate statistical techniques: A case study." Ecotoxicology and Environmental Safety 72(2): 301-309.

Kim, M. E., T. S. Shon and H. S. Shin (2013). "Forecasting algal bloom (chl-a) on the basis of coupled wavelet transform and artificial neural networks at a large lake." Desalination and Water Treatment 51(19-21): 4118-4128.

Kirillov, A. (2013). AForge .NET framework. Retrieved September 25th from http://www.aforgenet.com/.

Kratzer, C. R. and P. L. Brezonik (1981). "A CARLSON-TYPE TROPHIC STATE INDEX FOR NITROGEN IN FLORIDA LAKES1." JAWRA Journal of the American Water Resources Association 17(4): 713-715.

Lake Champlain Basin Program-nutrients. "Phosphorus Reduction Progress." Retrieved January 12th, 2015, from http://www.lcbp.org/water-environment/water-quality/nutrients/.

Lake Champlain Basin Program. (2014). "Lake Champlain Long-term Water Quality and Biological Monitoring Project." Retrieved May 26th, 2014, from http://www.watershedmanagement.vt.gov/lakes/htm/lp_longterm.htm.

Lee, J. H., Y. Huang, M. Dickman and A. Jayawardena (2003). "Neural network modelling of coastal algal blooms." Ecological Modelling 159(2): 179-201.

Li, H., G. Hou, F. Dakui, B. Xiao, L. Song and Y. Liu (2007). "Prediction and elucidation of the population dynamics of Microcystis spp. in Lake Dianchi (China) by means of artificial neural networks." Ecological Informatics 2(2): 184-192.

Li, J., J.-h. Cheng, J.-y. Shi and F. Huang (2012). Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. Advances in Computer Science and Information Engineering, Springer**:** 553-558.

Liao, Y., J. Xu and W. Wang (2011). "A Method of Water Quality Assessment Based on Biomonitoring and Multiclass Support Vector Machine." Procedia Environmental Sciences 10, Part A: 451-457.

Liu, W.-C. and W.-B. Chen (2012). "Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models." Computers & Geosciences 45: 13-25.

Lopes, J. F., C. I. Silva and A. C. Cardoso (2008). "Validation of a water quality model for the Ria de Aveiro lagoon, Portugal." Environmental Modelling & Software 23(4): 479-494.

Ma, Z., X. Song, R. Wan, L. Gao and D. Jiang (2014). "Artificial neural network modeling of the water quality in intensive Litopenaeus vannamei shrimp tanks." Aquaculture 433: 307-312.

Maier, H. R. and G. C. Dandy (2000). "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications." Environmental modelling & software 15(1): 101-124.

Maier, H. R., A. Jain, G. C. Dandy and K. P. Sudheer (2010). "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions." Environmental Modelling & Software 25(8): 891-909.

MathWorks-data division. (2014). "Divide Data for Optimal Neural Network Training." Retrieved July 23rd, 2014, from

http://www.mathworks.com/help/nnet/ug/divide-data-for-optimal-neural-network-training.html.

MathWorks-trainlm. (2014). "Levenberg-Marquardt backpropagation." Retrieved August 27th, 2014, from http://www.mathworks.com/help/nnet/ref/trainlm.html.

MathWorks. "MATLAB Compiler SDK-Build software components from MATLAB programs." Retrieved May 21st, 2015, from http://www.mathworks.com/products/matlab-compiler-sdk/.

McCormick, M. J., T. O. Manley, D. Beletsky, A. J. Foley, III and G. L. Fahnenstiel (2008). "Tracking the Surface Flow in Lake Champlain." Journal of Great Lakes Research 34 (4): 721-730.

McKinney, D. C. and X. Cai (2002). "Linking GIS and water resources management models: an object-oriented method." Environmental Modelling & Software 17(5): 413-425.

MIKE by DHI (2012). MIKE 21 & MIKE 3 FLOW MODEL FM Hydrodynamic and Transport Module Scientific Documentation.

MIKE by DHI (2013). MIKE 21 and MIKE 3 Flow Model FM Hydrodynamic Module Short Description. The Modules of the Flexible Mesh Series.

Moriasi, D., J. Arnold, M. Van Liew, R. Bingner, R. Harmel and T. Veith (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." Trans. ASABE 50 (3): 885-900.

National Centers for Environmental Information. "Local Climatological Data Publication at Burlington." Retrieved Feburary 10th, 2015, from http://www.ncdc.noaa.gov/IPS/lcd/lcd.html?_page=1&state=VT&stationID=14742&_target2=Next+%3E.

Ocampo-Duque, W., N. Ferré-Huguet, J. L. Domingo and M. Schuhmacher (2006). "Assessing water quality in rivers with fuzzy inference systems: A case study." Environment International 32 (6): 733-742.

OhioEPA (2010). Appendix A: Streamflow Estimation Techniques. TMDLs for the White Oak Creek Watershed, State of Ohio Environmental Protection Agency.

Olu-Owolabi, B. I., F. O. Agunbiade, E. O. Oseghe and K. O. Adebowale (2012). "Fuzzy Logic Modeling of Contamination Degree of Ni and V Metal Species in Sediments from the Crude Oil Prospecting Area of the Ondo Coast, Nigeria." Human & Ecological Risk Assessment 18(4): 902-918.

Open HUB. "Accord. net framework." Retrieved 04/03, 2015, from https://www.openhub.net/p/Accord-NET?ref=sample.

Palani, S., S.-Y. Liong and P. Tkalich (2008). "An ANN application for water quality forecasting." Marine Pollution Bulletin 56(9): 1586-1597.

Panda, R. K., N. Pramanik and B. Bala (2010). "Simulation of river stage using artificial neural network and MIKE 11 hydrodynamic model." Computers & Geosciences 36(6): 735-745.

Park, K., H.-S. Jung, H.-S. Kim and S.-M. Ahn (2005). "Three-dimensional hydrodynamic-eutrophication model (HEM-3D): application to Kwang-Yang Bay, Korea." Marine Environmental Research 60(2): 171-193.

Park, K., A. Y. Kuo, J. Shen, J. M. Hamrick and I. Tetra Tech (2000). A three-dimensional hydrodynamic-eutrophication model (HEM-3D): description of water quality and sediment process submodels (EFDC water quality model), Virginia Institute of Marine Science Gloucester Point, Virginia.

Patil, S. L., H. J. Tantau and V. M. Salokhe (2008). "Modelling of tropical greenhouse temperature by auto regressive and neural network models." Biosystems Engineering 99(3): 423-431.

Peng, S., G. Y. Z. Fu and X. H. Zhao (2010). "Integration of USEPA WASP model in a GIS platform." Journal of Zhejiang University-Science A 11(12): 1015-1024.

Piotrowski, A. P., M. Osuch, M. J. Napiorkowski, P. M. Rowinski and J. J. Napiorkowski (2014). "Comparing large number of metaheuristics for artificial neural networks training to predict water temperature in a natural river." Computers & Geosciences 64: 136-151.

Rehana, S. and P. P. Mujumdar (2009). "An imprecise fuzzy risk approach for water quality management of a river system." Journal of Environmental Management 90 (11): 3653-3664.

Sahoo, G. B. and C. Ray (2006). "Flow forecasting for a Hawaii stream using rating curves and neural networks." Journal of Hydrology 317(1–2): 63-80.

Sahoo, G. B., S. G. Schladow and J. E. Reuter (2009). "Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models." Journal of Hydrology 378(3–4): 325-342.

Sánchez-Avila, J., J. Bonet, G. Velasco and S. Lacorte (2009). "Determination and occurrence of phthalates, alkylphenols, bisphenol A, PBDEs, PCBs and PAHs in an

industrial sewage grid discharging to a Municipal Wastewater Treatment Plant." Science of The Total Environment 407(13): 4157-4167.

Santini, M., G. Caccamo, A. Laurenti, S. Noce and R. Valentini (2010). "A multi-component GIS framework for desertification risk assessment by an integrated index." Applied Geography 30(3): 394-415.

Sedki, A., D. Ouazar and E. El Mazoudi (2009). "Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting." Expert Systems with Applications 36(3, Part 1): 4523-4527.

Seo, D., M. Kim and J. H. Ahn (2012). "Prediction of Chlorophyll-a Changes due to Weir Constructions in the Nakdong River Using EFDC-WASP Modelling." Environmental Engineering Research 17(2): 95-102.

Sharma, A., M. Naidu and A. Sargaonkar (2013). "Development of computer automated decision support system for surface water quality assessment." Computers & Geosciences 51: 129-134.

Shrestha, S. and F. Kazama (2007). "Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan." Environmental Modelling & Software 22(4): 464-475.

Singh, A., M. Imtiyaz, R. Isaac and D. Denis (2012). "Comparison of soil and water assessment tool (SWAT) and multilayer perceptron (MLP) artificial neural network for predicting sediment yield in the Nagwa agricultural watershed in Jharkhand, India." Agricultural Water Management 104: 113-120.

Singh, K. P., A. Basant, A. Malik and G. Jain (2009). "Artificial neural network modeling of the river water quality—A case study." Ecological Modelling 220(6): 888-895.

Smeltzer, E., F. Dunlap and M. Simoneau (2009). "Lake Champlain phosphorus concentrations and loading rates, 1990-2008." Lake Champlain Basin Program. Grand Isle, VT.

Smeltzer, E. and S. Quinn (1996). "A phosphorus budget, model, and load reduction strategy for Lake Champlain." Lake and Reservoir Management 12 (3): 381-393.

Svozil, D., V. Kvasnicka and J. í. Pospichal (1997). "Introduction to multi-layer feed-forward neural networks." Chemometrics and Intelligent Laboratory Systems 39 (1): 43-62.

Tang, P.-K., Y.-C. Huang, W.-C. Kuo and S.-J. Chen (2014). "Variations of model performance between QUAL2K and WASP on a river with high ammonia and organic matters." Desalination and Water Treatment 52(4-6): 1193-1201.

Tetra Tech, I. (2002). User's manual for Environmental Fluid Dynamics Code Hydro Version (EFDC-Hydro) Release 1.00.

The Canadian Encyclopedia (2014). "Lake Champlain." Retrieved August 25th, 2014, from http://www.thecanadianencyclopedia.ca/en/article/lake-champlain/.

The Lake Champlain Basin (2009). The Lake Champlain Basin Waterbody Inventory and Priority Waterbodies List, Bureau of Watershed Assessment and Management,Division of Water,NYS Department of Environmental Conservation.

The Ministry of Sustainable Development, Environment and the Fight against Climate Change. Retrieved December 21st, 2014, from http://www.mddelcc.gouv.qc.ca/index.asp.

Tufford, D. L. and H. N. McKellar (1999). "Spatial and temporal hydrodynamic and water quality modeling analysis of a large reservoir on the South Carolina (USA) coastal plain." Ecological Modelling 114(2–3): 137-173.

United States Environmental Protection Agency. (2013). "Water Quality Analysis Simulation Program (WASP)." Retrieved March 15th, 2015, from http://www.epa.gov/athens/wwqtsc/html/wasp.html.

United States Geological Survey. "Explore Map of USGS Gages in the Lake Champlain Watershed." Retrieved December 5th, 2014, from http://vt.water.usgs.gov/echo_gage/basin_map.htm.

Vairavamoorthy, K., J. Yan, H. M. Galgale and S. D. Gorantiwar (2007). "IRA-WDS: A GIS-based risk analysis tool for water distribution systems." Environmental Modelling & Software 22(7): 951-965.

Vermont DEC and New York State DEC (1997). A phosphorus budget, model, and load reduction strategy for Lake Champlain (Lake Champlain Diagnostic-Feasibility Study final report). Waterbury, VT and Albany, NY.

Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation (2012). Lake Champlain Long-Term Water Quality and Biological Monitoring Program Description. Vermont Department of Environmental Conservation Water Quality Division and New York State Department of Environmental Conservation.

VishnuRadhan, R., P. Vethamony, Z. Zainudin and K. V. Kumar (2014). "Waste Assimilative Capacity of Coastal Waters along Mumbai Mega City, West Coast of India Using MIKE-21 and WASP Simulation Models." Clean-Soil Air Water 42(3): 295-305.

Wang, C., C. Shen, P. F. Wang, J. Qian, J. Hou and J. J. Liu (2013). "Modeling of sediment and heavy metal transport in Taihu Lake, China." Journal of Hydrodynamics 25(3): 379-387.

Warren, I. R. and H. K. Bach (1992). "MIKE 21: a modelling system for estuaries, coastal waters and seas." Environmental Software 7 (4): 229-240.

Wool, T., R. Ambrose, J. Martin and E. Comer (2007). Water Quality Analysis Simulation Program (WASP) Version 6.0 Draft: User's Manual, US Environmental Protection Agency–Region 4, Electronic file provided with installation of WASP version.

Wool, T. A., S. R. Davie and H. N. Rodriguez (2003). "Development of three-dimensional hydrodynamic and water quality models to support total maximum daily load decision process for the Neuse River Estuary, North Carolina." Journal of Water Resources Planning and Management 129(4): 295-306.

Wu, C. B., X. W. Den, X. Z. Yuan, H. Chen, L. Y. Zhang and Y. J. Jiao (2014). Simulation of salinity in artificial Lake in coastal region. 2nd International Conference on Advances in Computational Modeling and Simulation, ACMS 2013, July 17, 2013 - July 19, 2013, Kunming, China, Trans Tech Publications Ltd.

Wu, G. and Z. Xu (2011). "Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake." Ecological Modelling 222 (6): 1245-1252.

Zhang, M.-l., Y.-m. Shen and Y. Guo (2008). "Development and application of a eutrophication water quality model for river networks." Journal of Hydrodynamics, Ser. B 20(6): 719-726.

Zhang, X., K. R. Rygwelski, R. Rossmann, J. J. Pauer and R. G. Kreis Jr (2008). "Model construct and calibration of an integrated water quality model (LM2-Toxic) for the Lake Michigan Mass Balance Project." Ecological Modelling 219 (1–2): 92-106.

Zhao, L., Y. Li, R. Zou, B. He, X. Zhu, Y. Liu, J. Wang and Y. Zhu (2013). "A three-dimensional water quality modeling approach for exploring the eutrophication responses to load reduction scenarios in Lake Yilong (China)." Environmental Pollution 177: 13-21.

Zhao, L., X. Zhang, Y. Liu, B. He, X. Zhu, R. Zou and Y. Zhu (2012). "Three-dimensional hydrodynamic and water quality model for TMDL development of Lake Fuxian, China." Journal of Environmental Sciences 24 (8): 1355-1363.

Zhao, S., L. Da, Z. Tang, H. Fang, K. Song and J. Fang (2006). "Ecological consequences of rapid urban expansion: Shanghai, China." Frontiers in Ecology and the Environment 4 (7): 341-346.

Zheng, L., W. Cui and Y. Jia (2007). "Evaluation on seawater quality by fuzzy comprehensive evaluation method in Qingdao dumping area." Marine Environmental Science 26 (1): 38-41.

Zhu, C., Q. Liang, F. Yan and W. Hao (2013). "Reduction of Waste Water in Erhai Lake Based on MIKE21 Hydrodynamic and Water Quality Model." The Scientific World Journal 2013: 1-9.