Monitoring Lake Simcoe Water Quality using Landsat TM Images

By

Xian Guan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Geography

Waterloo, Ontario, Canada, 2009

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Inland lakes are important resources to humans, while the eutrophication effect caused by an overload of nutrients is a significant problem. This study focuses on utilizing the satellite remote sensing to monitor the water quality of Lake Simcoe, Ontario, Canada, which has been suffering from the overload of Total Phosphorus (*TP*) and therefore eutrophication for decades. The data employed in this study includes 22 cloud-free Landsat 5 TM images, as well as the nearly simultaneous in-situ data from 15 observation stations on the lake. Compared to the generally used model, an improved model is developed in this study to estimate the Secchi Disk Transparency (SDT), a parameter for water clarity measurements, using the TM images. Models based on different band combinations are compared to estimate the chlorophyll-*a (*chl-*a)* concentration. The results of these estimations are validated using the in-situ data by the linear regression analysis, and the accuracies are measured by the correlation coefficients R^2.

The results reveal that the improved SDT model provides higher prediction accuracies than the general model when applied to 68.2% (15 out of 22) of the images. The majority of the SDT predictions show high R^2, whereas some of the estimated chl-*a* concentrations have weak relationships with the in-situ data. The possible reasons for this are the geo-location of stations, as well as the influences of chl-*a* and Dissolved Organic Carbon (DOC). The

resultant concentration maps indicate that the eutrophic water is normally distributed at the near-shore areas and the northeastern part of Lake Simcoe. In addition, the southern Cook's Bay has always been suffering from an extremely serious water quality problem even until now. Meanwhile, the water quality of the southwestern part of Lake Simcoe is much better than the other parts of this lake. The results also show that the water quality of Lake Simcoe was at its worst in August and September for the past 22 years while it was much better in the other sampling seasons. According to the trend of the monthly averaged SDT, on an overall scale, the SDT dropped from 1980 to 1982 and then kept relatively stable until the fall of 1992, followed by a gradual increase until 2000, and then stayed constant until the summer of 2008. The chl-*a* concentration reveals an inverse trend, i.e., the higher the chl-a concentration, the more turbid the water.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Dr. Jonathan Li for his valuable guidance, continuous encouragement and financial support during my masters study period. He taught me not only the approach to the specific problem in this thesis, but also his insights, inspirations and enthusiasm to conduct research.

Secondly, I would like to thank Dr. Bill Booty, who serves on my Thesis Committee. He provided me the great opportunity to do the research at the Canada Centre for Inland Waters (CCIW), Burlington, Ontario, and numerous valuable advices. I also would like to thank Dr. Robert Bukata, Mr. John Jerome and Dr. Caren Binding at CCIW, who helped me a lot to collect the data and discuss with me. I am grateful to Dr. Jennifer Winter and Mr. Hamdi Jarjanazi, who offered me all the in-situ measurements.

Thirdly, I would like to thank the two readers of this thesis, Professors Dr. Claude Duguay and Dr. Su-Yin Tan, for taking their valuable time to read my thesis and for constructive suggestions.

I would like to thank Leif Olmanson, who is a Ph.D candidate at the University of Minnesota, to provide me a number of valuable suggestions. I also would like to thank my colleagues Qingxu Huang, Yao Lu, Yu Li, as well as the remote sensing specialist Anne

Grant from Mapping, Analysis, and Design at the Faculty of Environment, for helping me get through a lot of technical problems. Thanks also go to Mrs. Lynn Finch, for her administrative support.

I must express my sincere thanks to Professor Dr. Geoff Wall, who brought me to this university.

Last but not least, I am deeply indebted to my family, for their endless love and support. To them I dedicate this thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
AOI	Area of Interest
AVHRR	Advanced Very High Resolution Radiometer
ASCII	American Standard Code for Information Interchange
CASI	Compact Airborne Spectrographic Imager
CCIW	Canada Centre for Inland Waters
CDOM	Coloured Dissolved Organic Matter
Chl-a	Chlorphyll-a
DN	Digital Number
DOC	Dissolved Organic Carbon
ETM^+	Enhauced Thematic Mapper Plus
GCPs	Ground Control Points
LR	Linear Regression
LSCF	Lake Simcoe Conservation Foundation
LSEMS	Lake Simcoe Environmental Management Strategy
LSRCA	Lake Simcoe Region Conservation Authority
MOE	Ontario Ministry of Environment
MERIS	Medium Resolution Imaging Spectrometer Instrument

MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multispectral Scanner
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NLR	Non-Linear Regression
SAV	Submergent Aquatic Vegetation
SDT/SDD	Secchi Disk Transparency/Secchi Disk Depth
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SEE	Standard error of estimate
TIR	Thermal Infrared
TM	Thematic Mapper
TP	Total Phosphorus
TSIs	Trophic State Indices
TSM	Total Suspended Matter
TSS	Total Suspended Solids
USGS	United States Geological Survey
UTM	Universal Transverse Mercator Coordinates
WGS	World Geodetic System

Chapter 1

Introduction

Inland water is critical to our citizens. The eutrophication problem caused by the overload of nutrients strongly affects the ecosystem in the lakes and therefore people's lives. Section 1.1 introduces the water quality problems existing in inland lakes. Section 1.2 introduces the traditional way to monitor water quality and its weakness. Section 1.3 explains the reasons for this study, using remote sensing techniques to monitor lake water quality, plus it contains the four overall objectives of this study. The organization of this thesis is briefly described in Section 1.4.

1.1 Water Quality Problems of Inland Waters

Clean and adequate fresh water is essential to the welfare of human beings, as the water quality directly affects natural ecosystems, human health, and economic activities. Meanwhile, human activities directly affect water quality. Concentrated pollutants entering surface waters from specific locations (point-source discharges) and dispersed pollutants generated from local or small-scale activities (non-point source discharges) add large quantities of nutrients, pathogens, and toxins to the lakes. A Canadian government report in 2001 stated that human activity resulted in more than 12 thousand tonnes of Phosphorus (P)

and 303 thousand tonnes of Nitrogen (N) entering Canadian fresh, ground, and coastal waters in 1996. It also reported that municipal sewage, which adds approximately 5.6 thousand tones P, is the largest point source, while the discharge of industrial wastewater is also a considerable source (Chambers et al., 2001).

These overloads of nutrients result in eutrophication, one cause of ecological problems in the world today. The excessive phosphorus, together with some other chemicals, can accelerate the growth of algae and other aquatic plants. In response to over-enrichment with nutrients, the phytoplankton community may shift to bloom-forming nuisance algae, which are harmful to other organisms (Smith, 1990). Decomposition of algal blooms can lead to foul odors and oxygen depletion, which can in turn, lead to fish deaths (Carpenter et al., 1998; Smith, 1998). These factors, together with the heightened concentration of suspended and dissolved matter caused by humans, lead to a rapid transition to a higher trophic level, which entails algal bloom events, anoxia, and eventually a dramatic deterioration of water quality (Henderson-Sellers and Markland, 1987; Kondratyev et al., 1998).

1.2 Water Quality Monitoring Methods

These problems that inland lakes are suffering from make protecting and monitoring lake water quality a major concern for many local, provincial, and federal government agencies. Many lake-water monitoring programs have been set up to keep records of the change in

water quality parameters. These programs usually include on-site sampling and laboratory analysis. Three water quality variables that have been most commonly used to indicate trophic state are total phosphorus (TP), chlorophyll *a* (chl-*a*), and Secchi disk transparency (SDT). Measurement of these water quality variables, along with various transformations, such as the Trophic State Indices (TSIs), has been widely used by lake management agencies and organizations. Generally, this ground-based monitoring achieves high accuracy of lake water quality measurements. However, the expense, time, and sampling frequency make ground-based monitoring impractical to be applied to large areas. The resolution is a compromise between spatial and temporal details (Kloiber et al., 2002).

Although it has not been officially admitted as a government policy tool, remote sensing techniques have been proven useful for monitoring water quality for decades. The use of this satellite-based technology is a cost-effective way to gather the needed information for regional water quality assessments in lake-rich countries such as Canada. Among the several earth observation satellites, the relatively low cost, temporal coverage, spatial resolution, and data availability of the Landsat Thematic Mapper (TM) imagery make it particularly useful for assessment of inland lakes. Several studies have demonstrated a strong relationship between Landsat Multispectral Scanner (MSS) or TM data and ground observations of water clarity and chl-*a* (e.g., Brown et al., 1977, Lillesand et al., 1983; Lathrop and Lillesand, 1986; Lathrop, 1992; Cox et al.,1998). In 1999, with the launch of

Landsat 7, data distribution was returned to the public sector, and the cost of data acquisition dropped significantly. In particular, TM data is available free of charge. Along with today's powerful desktop computers and sophisticated software, processing and analysis of satellite imagery has become relatively inexpensive and easy to perform.

1.3 Objectives of the Study

This study is motivated by the water quality deterioration in the Canadian inland waters, as well as the weakness of conventional water quality programs. One of the eutrophic inland lakes, Lake Simcoe, is selected as the study site. This study has four overall objectives:

1. To utilize and develop different models for estimating two useful water quality parameters, SDT and chl-*a* using the Landsat 5 TM images.
2. To establish empirical relationships between these two water quality parameters and spectral information by using linear regression analysis.
3. To identify both temporal and spatial water quality patterns of the study area and to map the excessive pollution regions in reference to the inland lake classification criteria.
4. To graph the overall water quality trends of the study area for a period of 22 years, from 1987 to 2008.

Despite past efforts to improve the water quality of the study area, Lake Simcoe, the ecosystem remains stressed, particularly in the southern part. This, coupled with the reduction of observation stations, creates a greater urgency to monitor this area more effectively. It is necessary to generate both the temporal change and spatial distribution maps of the lake. Furthermore, if the relationship between the estimated data and in-situ data is reliable enough, it is possible to estimate and chart water quality of this lake back to the late 1970s. This study proves that remote sensing is a reliable technique that can be used as a supplement for ground-based water quality observation and could possibly help the managers and policy makers at Environment Canada and Ontario Ministry of Environment (MOE) to improve the monitoring, and thus the water quality, of Lake Simcoe.

1.4 Thesis Structure

The remaining parts of this thesis are organized into four chapters:

Chapter 2 reviews the related work done by other researchers. It starts with the importance of monitoring water quality by utilizing remote sensing techniques. Then an overview of current water quality models is introduced and explained. Next is a review and discussion of the specific research concerning the use of Landsat images for water quality estimation. Finally, Chapter 2 is summarized and the research gaps are discussed.

Chapter 3 details the proposed methodology, including a description of the study area and the in-situ and satellite data, and presents the models for both the SDT and chl-*a* parameters, and the regression analysis technique.

Chapter 4 details and explains the experimental results from this study, which consists of the regression equation for each of the parameters, the temporal trends of the water quality of Lake Simcoe over the past 22 years, the spatial distribution of the water pollution, and the water quality classification based on the Carlson's criteria (1996) . The influence of dissolved organic carbon is also discussed in this chapter.

Chapter 5 draws conclusions from this research and suggests further areas of study.

Chapter 2

Remote Sensing of Water Quality: An Overview

2.1 Importance of Water Quality Monitoring by Remote Sensing

Inland lakes provide an important recreational, commercial, and aesthetic resource for the public (Nelson et al., 2003). However, a continuous deterioration of the ecological status of inland water basins, salt-water coastal zones, and oceanic gulfs makes water quality problems urgent (NASA, 1986). The reason for this deterioration is largely because of an unprecedented intensification of natural eutrophication in the water bodies. This 'aging' of water bodies is triggered and accelerated by the extensive anthropogenic inputs of nutrients, usually phosphorus and nitrogen. Additionally, the increase in water temperature caused by thermal industrial inflows can also contribute to this effect (Kondratyev et al., 1998). Those factors, together with the human-induced heightening of concentrations of suspended and dissolved matter, lead to a rapid transition of the basin to a higher trophic level, which entails algal bloom, anoxia, and eventually a dramatic deterioration of water quality parameters (Henderson-Sellers and Markland, 1987; Kondratyev et al., 1998).

The eutrophic problem existing in the inland lakes makes routine water quality monitoring programs of a great necessity. Government agencies usually judge the quality of lake

waters by monitoring the water quality indicators, or parameters. One of them is called Secchi Disk Transparency (or SDT), which is normally considered an effective and economic parameter to measure water clarity. The Secchi Disk is a 20cm diameter white disk which is lowered into the water column until it can no longer be seen. The distance between the disk and the surface of the water is called the SDT.

Another most common indicator of trophic state is chl-a, which is the green pigment found in all plants allowing them to photosynthesize. If populations of algae grow too fast in relation to the rest of the system, though, they may create water quality problems in the lake ecosystem, such as algal blooms. The absorption peak and the reflection peaks are near 440nm and 550nm, respectively. The chl-*a* concentration can be a very good indicator for the lake water quality. So the lake scientists often test lakes for chlorophyll to get an estimation of the amount of algae growing in a lake (Li et al., 2007).

Dissolved Organic Carbon (DOC) is a broad classification for organic molecules of varied origin and composition within aquatic systems. In general, organic carbon compounds are a result of the decomposition process of higher organic matter, such as plants. Also, when water comes into contact with high organic soils, a certain amount of these components can drain into rivers and lakes as DOC. DOC is normally colourless, but it absorbs light across the UV (280-400nm) and visible (400-700nm), and therefore reducing the potential of reflectance (Hirthle and Rencz, 2003).

SDT, together with chlorophyll pigment and total phosphorus are believed to be the three main parameters to calculate the Carlson Trophic State Index (Carlson, 1977). Of these three, SDT and chl-*a* can be estimated by satellite imagery (Kloiber et al., 2002).

Traditional methods, which are the ground-based measurements, used in water quality classification include the collection of water samples from lakes and their analysis in laboratory conditions (Koponen et al., 2004). Ground truth measurements, usually include: 1) water sampling for laboratory analysis, 2) the on-site measurements, and 3) weather observations (Koponen et al., 2002). Normally these existing measurement techniques are sophisticated (Nelson et al., 2003). For example, in the chl-*a* retrieval procedure, the sum of chl-*a* is determined by a spectrophotometer after extraction with hot ethanol and turbidity by nephelometric method, then scattered within a 90° angle from the beam directed at the water sample (Koponen et al., 2002).

Obviously this in-situ point measurement can give accurate results in the laboratory, and there are many studies proving that the quality of data collected by volunteers is comparable to data collected by professional monitoring personnel, with no statistical difference in summer averages of SDT between volunteer and professional measurements (Heiskary et al., 1994; Kerr et al., 1994; Obrecht et al., 1998; Canfield et al., 2002; Nelson et al., 2003). However, the vast area of a lake makes it difficult to evaluate water quality of

the lake as a whole, especially for a country with thousands of lakes such as Canada. Also, a dense water quality monitoring network would not be cost effective. Currently, most of the sampling work is fulfilled by volunteers, and it is doubtful that there are sufficient volunteers to sample all the lakes, some of which are very large. Furthermore, there are some extreme conditions which are so dangerous that human beings cannot access them readily. Most importantly, it is difficult to generate the up-to-the-minute spatial overview data that is necessary for global assessment and monitoring of water quality (Brivio et al., 2001; Wang et al., 2004). Meanwhile, the design of regional-scale monitoring programs is frequently based on a compromise between spatial and temporal detail. Considering the cost and logistical problems, the ground-based monitoring programs usually sacrifice spatial coverage to monitor fewer lakes in favor of more frequent sampling (Kloiber et al., 2002). New approaches and techniques are required to collect the water quality parameters on larger lakes and more lakes.

In recent years there is growing interest in evaluating lake water quality using a remote sensing technique, which has been proven as a useful tool for monitoring water quality for decades, although the routine application of this technique has been slow to catch on. This may be due to a lack of familiarity with the technique among inland aquatic scientists and also due to a perception that the data are expensive and difficult to process. In 1999 with the launch of Landsat 7 with Enhauced Thematic Mapper Plus (ETM^{+}), data distribution

was made available to the public, dropping the cost of data acquisition significantly, with TM data available free of charge. With today's advanced hardware and software, processing and analyzing satellite images are easier and less costly.

These developments, combined with the improvement of spatial and temporal resolution, as well as the optimization of algorithms, satellite remote sensing methods began to hold more potential for improving water quality monitoring (Wang et al., 2004).

2.2 An Overview of Water Quality Models

The remote sensing technique has been utilized for water quality monitoring for over twenty years, and during this time, a variety of approaches for water quality parameters estimation were developed. Giardino et al. (2007) categorized the current methodologies into three main types, which are: 1) Empirical, 2) Semi-empirical, and 3) Analytical.

2.1.1 Empirical Approach

The first category, which is the simplest, is the empirical approach. This methodology is based on developing bi-variate or multivariate regressions between the remotely sensed reflectance data and the ground measurement water quality parameters. This approach was used by Neilson et al. (2003), Li et al. (2007), Koponen et al. (2001), Brezonik et al. (2007), who investigated the relationship between digital numbers of single band or the band combinations, and the in-situ water quality data collected in coincidence with the sensor

overpass. The simple empirical method has been applied in the water quality area for the longest period of time, but the most significant weakness of this approach is that it relies too much on the in-situ data. If the ground measurements are limited, or not available, the empirical approach may be inaccurate. In addition, the empirical approach is often site-specific or scene-dependent, which means one model is possibly only being applied to one image.

2.2.2 Semi-Empirical Approach

A second approach, called semi-empirical, is a method where spectral characteristics of the parameters of interest are known, and this knowledge is included in the statistical analysis, which is based on the well-chosen spectral areas which correlate with appropriate wavebands (Giardino et al., 2007). Harma et al. (2001) investigated the feasibility of band combination of the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Medium Resolution Imaging Spectrometer Instrument (MERIS) satellite data to monitor the lakes and coastal waters in Finland by utilizing this semi-empirical algorithm, and then tested the accuracy of the algorithm by Landsat TM data. Ostlund et al. (2001) employed the hyperspectral data collected by the Compact Airborne Spectrographic Imager (CASI) and Landsat TM data, as well as the in-situ data, to build algorithms to extract the chl-*a* concentration and the suspended sediments. The semi-empirical approach is currently the most often used method, but this approach has the same problem of scene-dependence.

2.2.3 Analytical Approach

The third approach is called the analytical approach, which involves inverting four relationships to determine water quality parameters from the remote sensed data, in accordance with the order of the water quality parameters, the specific Inherent Optical Properties, the bulk Inherent Optical Properties, the Apparent Optical Properties and the Top of Atmosphere radiance (Giardino et al., 2007). An example of scholars using such an approach is Dekker et al. (2001), who used a 1-Dimensional network model, based on analytical optical modeling, utilizing the in-situ inherent optical properties. The method they used allows comparisons of multiple Total Suspended Matter (TSM) maps, both temporal and in-situ, extracted from Landsat TM and SPOT images. Thus they claimed that the remotely sensed data can become an independent measurement tool for water management. However, in their study, the model is prove to be less sensitive when TSM concentration is low. Different from the empirical and semi-empirical approaches, the analytical one is a well-calibrated and physically-based model, which can thus theoretically be applied to every satellite scene of the selected lake. Although the results of the analytical approach are more reliable and applicable than the previous two methods, it is relatively complicated and involves many optical characteristics of the body of water; therefore, the use of this approach is still limited. Most studies prefer to employ the empirical and semi-empirical approaches.

2.3 TM Images for Water Quality Monitoring

2.3.1 Rationale for Choosing Landsat TM Images

Satellites images of different spatial resolution are employed to estimate the water quality parameters. Chen et al. (2007b), using the Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) satellite imagery from 1997 to 2005, examined both the spatial and temporal variability of the SDT within Tampa Bay. Prangsma and Roozekrans (1989) utilized the Advanced Very High Resolution Radiometer (AVHRR) instrument, providing the observational data, to interpret the parameters of water quality of the coastal zone. Koponen et al. (2004) extracted six water quality parameters of Finnish lakes from the Earth Observing System Terra/Aqua MODIS, using a bio-optical reflectance model. The accuracy of classification turned out to be 80%.

Although the SeaWiFS, AVHRR and MODIS images have a good temporal resolution and data availability, the low spatial resolution of these satellites leads to a missing of water quality details in lakes. The open ocean water is defined as Case 1 water, while the coastal, estuary or inland water bodies are classified as the Case 2 water (Morel and Prieur, 1977). The remote sensing of Case 2 water has been far less successful than that of Case 1 water, mainly due to the complex interactions among optically active substances (e.g., phytoplankton, suspended sediments, coloured dissolved organic matter, and water) in the

former (Oyama et al., 2009). Therefore, the datasets such as MODIS and SeaWiFS, which are more frequently used in Ocean and large scale lakes, are generally regarded as not suitable for the smaller and more complex inland lakes. Instead, the satellite images with medium or high spatial resolution (0.6m-1m spatial resolution) are required. Indeed, there have been quite a few studies that utilize high-resolution images to monitor inland water quality. Sawaya et al. (2003) employed IKONOS images to evaluate the SDT and urban impervious areas. They proved that the modeling and estimation methodologies designed for medium-resolution satellite image data could also be employed to the high-resolution data. Wolter et al (2005) used QuickBird images to map the Submergent Aquatic Vegetation (SAV) of the Great Lakes. Nevertheless, the result of both studies shows that the shadow associated with high-resolution images is a limitation that would possibly cause interference when extracting the water features and could not be neglected. Given this drawback and the high price of high-resolution images, it may be not appropriate to use high-resolution images in the study area with large spatial coverage or the long trend monitoring.

Combining all the above, the relatively low cost, wide temporal coverage, fair spatial resolution, and data availability of the Landsat TM imagery make it particularly useful for the assessment of inland lakes (Kloiber et al., 2002).

2.3.2 Related Work using Landsat TM images

Landsat TM images have been explored in several studies as a method of reducing the cost and labor of sampling water clarity in the field (Khorram and Cheshire, 1985; Lathrop, 1992; Kloiber et al., 2000; Dewider and Khedr, 2001; Nelson et al., 2003). It has many advantages such as good spatial coverage and the possibility of measuring many lakes simultaneously, which would provide valuable data on the seasonal variability of water quality. Also, it can reduce water sampling and lakes not monitored by traditional methods (Koponen et al., 2002). In addition, sensors such as Landsat provide an excellent tool to examine changes over time since Landsat images have been collected over large areas of the Earth's surface for decades. The first Landsat satellite was launched in 1972, and Landsat 5, which is the sensor for all the images used in this study, was launched in 1984 and is still functioning now.

Early in the 1980s, Khorram et al. (1989) used Landsat 4 TM imagery and the in-situ data around the date of satellite's overpass to estimate SDT, chl-*a*, Turbidity and Temperature of Augusta Bay, Italy. In this early stage, the model developed only used averaged digital number (DN) values extracted from the satellite bands. Only one image was employed, thus no change of water quality was presented. In addition, they did not give details of the resultant water quality distribution.

Wang et al. (2006) employed one Landsat 5 TM image to establish its relationship with four water quality indicators, SDT, chl-*a*, Total Suspended Solids (TSS) and turbidity in the Reelfoot Lake, Tennessee. They compared various combinations of TM bands to retrieve these parameters and generated the distribution maps of them respectively. Their results show that the TM2 is positively correlated to the water quality data while TM3 is negative due to the strong absorption of algae bloom, and the poor water quality may result in high reflectance of TM2 and low reflectance of TM3. They have noted that the TM2 and TM3 are the best predictors for water quality.

Kallio et al. (2008) investigated the use of Landsat 7 ETM^+ images to monitor the turbidity, coloured dissolved organic matter (CDOM) and SDT in two river basins in Finland. They pointed out that the atmospheric parameters can possibly improve the accuracy of water quality estimates without concurrent in situ measurements.

Li et al. (2007) compared inverse models built by other scholars, using a semi-empirical approach, to extract seasonal chl-*a* concentration from 10 TM images of Lake Taihu, China. The results of their study showed that the model equations of different seasons turned out to be the same or similar, while the coefficients were changing a lot. They also proved that the higher the concentration of chl-a, the more accurate the model would be.

Wu et al. (2008) compared 22 MODIS and 5 Landsat-5 TM datasets within two years for mapping spatial-temporal dynamics of SDT in Poyang Lake, China. They used two multiple regression models including the blue and red bands of Landsat images. The results show that both datasets can achieve high accuracies with the in-situ data with the R^2 of 0.94. The higher the SDT is, the less signal reflectance would be returning to Landsat TM sensor.

Although many studies show their algorithms of retrieving water quality parameters from satellite images have high accuracies, one weakness exists among them. That is they usually utilize only one or two satellite images and the corresponding ground data as their research scales, which means that their methodologies are probably time-specific or site-specific. Although the Landsat images have been used for estimating water quality for more than 30 years, few studies have ever been done on the long-time monitoring of inland lakes (Olmanson et al., 2008). As the water quality of lakes would not have a dramatic change during a very short time period, one algorithm may possibly be applied to many images due to this lack of variation. One exception for this is the recent study of Olmanson et al. (2008), which completed a 20-year comprehensive water clarity database of over 10,000 lakes of Minnesota. The results show that there is a close relationship between the satellite reflectance value and the in-situ SDT data; they also provide evidence that the SDT has a strong geographic pattern in Minnesota. However, this research just simply continues their work of year 2002 (Olmanson et al., 2002) by adding some image data with no new

models or new water quality parameters. As the SDT is the simplest indicator which can be estimated by satellite images, adding more parameters is more challenging and difficult, especially for a long period.

2.3.3 Assessment of Water Quality Estimation Models

Several investigations have found that reliable empirical relationships can be developed between the Landsat data and the in-situ measured water clarity and chlorophyll concentration data (Kloiber et al., 2002). The relationship between remotely sensed imageries and water quality parameters are usually examined through linear regression of the in-situ measured data and the spectral values retrieved from the sensor (Lillesand et al., 1983; Lathrop and Lillesand, 1986; Kloiber et al., 2002; Nelson et al., 2003; Onderka and Pekaroua, 2008). Kondratyev et al. (1998) mentioned in their article that unlike the clear marine or oceanic waters, for which linear regression retrieval relationship are valid, inland and coastal zone water masses have a high degree of optical complexity.

According to the literature, most of the researchers use the regression models as the statistics analysis to calibrate the algorithms of extracting information and look for relationships between the image value and the ground data. One example is the study of Nelson et al. (2003), which used a linear regression model to relate the band ratio of ETM1/ETM3 with the in situ SDT. Fuller et al. (2004) used the multiple regression method

to derive the equation for extracting SDT from Landsat 7 ETM+ data and the stepwise regression method to identify the best-fit regression equation for predicting chl-a. Similar methodologies have been applied by Koponen et al. (2002), Kloiber et al. (2002), Ruhl et al. (2001), Sawaya et al. (2003), Chen et al. (2007), etc. The key in regression analysis is to select an appropriate regression method and independent variables of high R^2, which indicates good correlation between the regression equation and existing data when its value approaches one (Wang et al., 2006). There are some scholars who used other methodologies to validate the data. Lathrop (1992) applied a bivariate regression in his study. Gan et al. (2004) employed the linear regression (LR) and the nonlinear regression (NLR), as well as the artificial neural network (ANN) called modified counter propagation network. They found that the NLR described the relationship between spectral data and ground data more closely than the LR. They also concluded that the ANN is more suitable to learn and generalize information from complex or poorly understood systems, in which case it will better represent the relationship than NLR. However this calibration method is normally utilized on the condition that the in-situ measurements of water quality characteristics are limited in spatial and temporal domains (Sudheer et al., 2006). In addition, the input information selection procedure is quite complex and requires considerable computational time. Therefore, if sufficient ground measurements exist, the regression methods are more frequently used.

2.4 Chapter Summary

To summarize, the eutrophic problems existing in the inland lakes make the routine water quality monitoring programs of great necessity and importance. As the spatial coverage and sampling frequency make these traditional sampling programs limited, remote sensing is widely used in water quality studies nowadays.

The current water quality estimation models based on satellite images are classified into three main categories: empirical approach, semi-empirical approach and analytical approach. Both the empirical and semi-empirical approaches have the merits of simplicity, whereas both of them have the weakness of overly relying on in-situ data. The analytical approach is physically-based and relatively independent from the ground measurements data but it involves too many optical constituents.

Images with different spatial resolutions have been applied to the area of water quality studies. Satellite data with relatively low spatial resolutions such as the SeaWiFS and MODIS images have very good temporal resolutions so they are widely utilized in trend analysis; however, they are often applied in the context of oceanic water rather than lake water studies owing to their limited spatial resolutions. High resolution satellite data such as IKONOS and QuickBird can better represent the spatial details of lake water quality,

whereas these studies are often limited by the influence of shadow in the images, as well as the expensive data acquisition. The 30m medium spatial resolution image such as Landsat TM data is considered as a compromise between these weaknesses and it can be used in both trend analysis and for displaying spatial details of lake water.

It is also noticeable that several studies of water quality estimation by remote sensing focus on the algorithms for retrieving water quality themselves, while ignoring the lake water characteristics. Afterall, remote sensing serves as a tool or supplement for water quality monitoring in order to reduce human's burden, whether it will be treated as a policy or not. At the moment, few investigations are aiming at the long trend water quality monitoring of inland lakes, although many scholars choose to select one specific study area and focus on the water pollution problem itself.

In addition to building a relationship between the TM images and water quality parameters, this study aims to analyze water quality (SDT and chl-*a*) change of one inland lake, Lake Simcoe, in the last 22 years, as well as the geographic pattern of these water quality parameters. As both the image data and in-situ data involved in this study are quite complex and from different time periods of each year, the empirical approach is utilized to build reliable relationships between each pair of data (the image data and the almost simultaneous in-situ data). To avoid the scene-specific problems of the empirical approach,

one best model is chosen for each image after the comparison of different band combinations. These empirical equations calculated from regression analysis are then utilized to estimate the SDT and chl-*a* concentration maps.

Chapter 3

Satellite-based Water Quality Monitoring Method

This chapter provides a detailed description of the methodology developed in this study. Section 3.1 introduces the study site of this research, which includes the reasons for choosing Lake Simcoe, the problem statement and the related work about this lake. Then Section 3.2 describes both the image data and the in-situ data. In Sections 3.3 and 3.4, interpretations of the image process procedures and the explanation of models are given. Section 3.5 describes how the linear regression is used in this study to calibrate the estimated water quality parameters, followed by the spatial and temporal analysis in Section 3.6. Lastly this chapter is summarized in Section 3.7.

3.1 Study Area

3.1.1 Lake Simcoe and Its Water Quality Problem

Lake Simcoe, located in the southern part of Ontario, is regarded as the 12^{th} largest lake in the province. It is also a large freshwater lake to freeze over completely during the winter. With the latitude of 44°25′N and the longitude of 79°20′W, Lake Simcoe has a surface area of 722km^2 in total, a maximum depth of 41m, length of 30km, and width of 25km. This lake consists of the "main" basin and two large bays, which are Cook's Bay at the south and Kempenfelt Bay on the west side (Johnson and Nicholls, 1989).

Great significance has been empowered to this lake by the residents living around it. In this area, the fishery is known as an important industry, which generates more than 200 million dollars a year (Environment Canada, 2009). However, there have been water-quality problems in Lake Simcoe since the 1970s, when the phosphorus load to this lake dramatically increased and subsequently contributed to the recruitment failure of cold-water fish, as well as the excessive growth of aquatic macrophytes and algae (Winter et al., 2007).

This overload of phosphorus is derived from the land use around Lake Simcoe, which is predominately agricultural land, the forest, and wetland; in addition, the vegetation lands occupy 40% of the terrestrial area (Johnson and Nicholls, 1989). Only 5% of the catchments area is urban, where the population has doubled in the past twenty years (Kilgour et al., 2008; Eimers et al., 2005). Figure 3.1 shows the shape and land use of Lake Simcoe and it demonstrates that most of the land around this lake is covered by cultural meadow, thicket and the intensive agriculture land. These land cover types are summarized in the pie chart (see Figure 3.2). The landforms and land uses in the watershed significantly influence the water flowing to the lake, and thus the water quality of the lake.

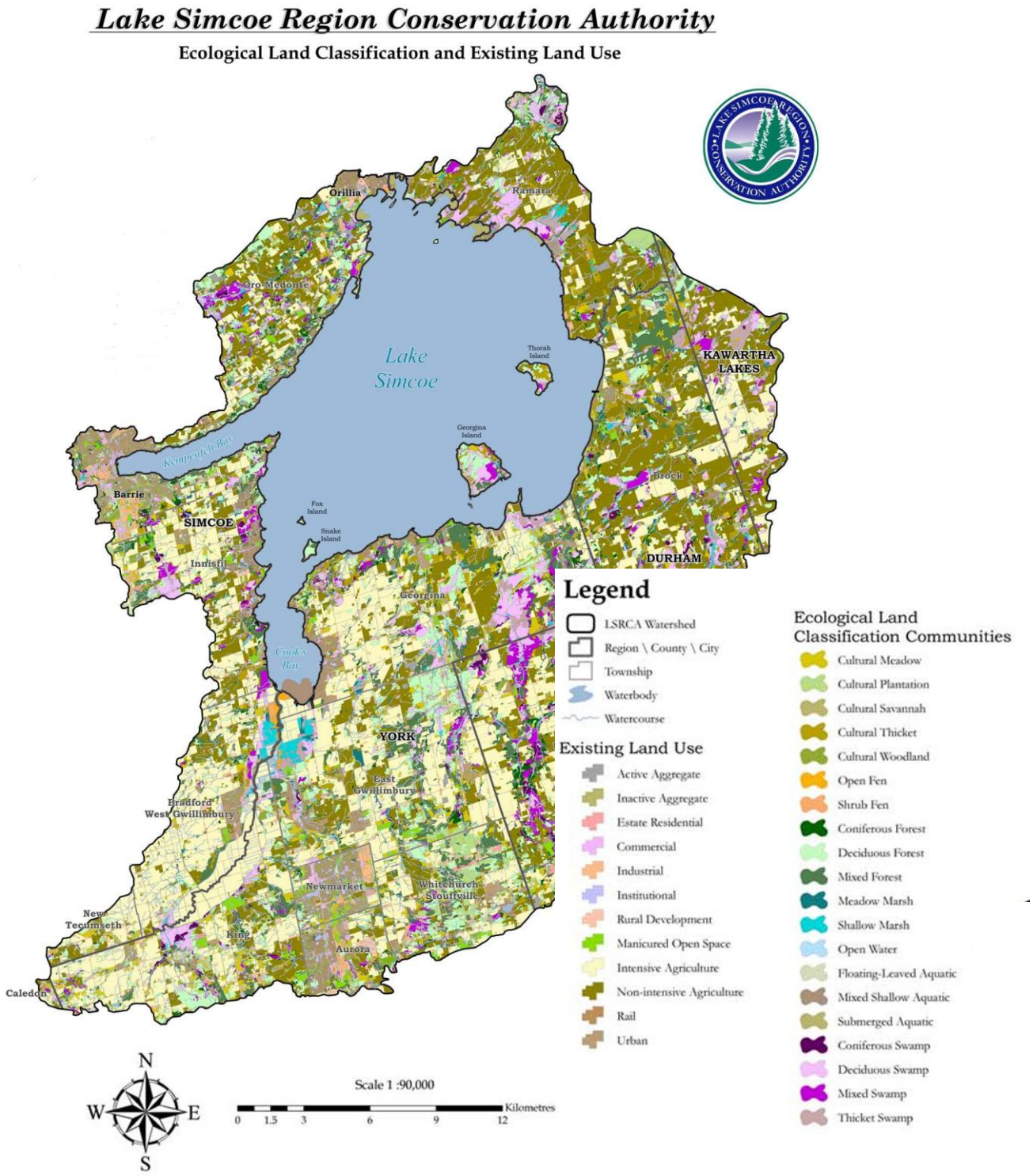


Figure 3.1 Ecological Land Use Map of Lake Simcoe
(Source: Lake Simcoe Region Conservation Authority)

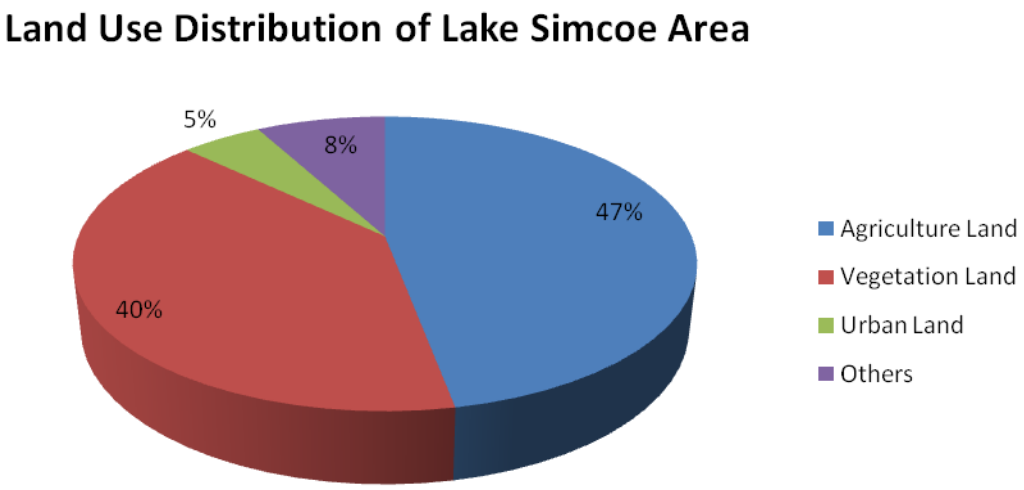


Figure 3.2 Land Use Distribution of the Lake Simcoe Area

3.1.2. Related Studies on Lake Simcoe

Since Lake Simcoe has a noticeable water-quality problem, in the past two decades, great attention has been paid to this watershed in order to improve its water quality. One example of this, the Lake Simcoe Environmental Management Strategy (LSEMS), which was initiated in the 1980s in response to a declining coldwater fishery in this lake, is to protect the health of the Lake Simcoe watershed ecosystem by restoring a self sustaining coldwater fishery, improving water quality, reducing phosphorus loads and protecting natural heritage features and functions (LSEMS, 2006). The creation of the Lake Simcoe Region Conservation Authority (LSRCA) played a role as the lead agency, as well as delivering cost effective programs and projects that will help restore a naturally reproducing coldwater fishery in Lake Simcoe (LSEMS, 2006). The LSEMS is to protect the local ecosystems and

contribute to quality of life in communities throughout the watershed. The Lake Simcoe Conservation Foundation (LSCF) was also founded to protect and restore the lake, in partnership with the LSRCA and it takes a watershed-based approach to cleaning up the lake (LSCF, 2006).

Besides these strategies made by the management departments, the scientists also paid significant attention to the problem of Lake Simcoe in the past few years. As phosphorus is the main problem of Lake Simcoe, related studies have been conducted in order to lower the phosphorus concentration in this lake. Winter et al. (2005) used an empirical modeling approach to derive a phosphorus loading objective, and this model was evaluated using the data from 1980 to 2004. The proposed total phosphorus (TP) loading target of 75 metric tons/year was predicted. Lu et al. (2006) described a 3-D hydrodynamic, thermal and water quality model, including wind-driven circulation, temperature variation, to stimulate eutrophication processes, such as the algal growth, in response to the loadings of phosphorus. Johnsong and Nicholls (1988) used a combination of sonar and core MSR measurements to deal with spatial variability in Lake Simcoe. In 1989, they did a temporal and spatial variability research of sediment and phosphorus loads in this lake. Nicholls (1997) did a limnological-based study for the phosphorus loading objective of Lake Simcoe.

The research above have contributed greatly to the management of Lake Simcoe, as the results of these studies reveal that the TP loads to Lake Simcoe are currently below the target 75 metric tons/year (Winter et al., 2007). However, algae and macrophytes continue to be problems in the near-shore zone of the lake. In addition, the dissolved oxygen concentrations remain below the level (7mg/L), which is considered to be protective of lake trout. Therefore, the lake water quality monitoring, as well as the efforts to reduce TP loads to this lake still need to be continued (Winter et al., 2007).

The previous studies imply that the water-quality monitoring of Lake Simcoe still relies on conventional approaches. However, there has been a reduction of ground stations on this lake since 1995. Now only 11 and 8 stations are left for the sampling of SDT and chl-a, respectively (Figure 3.3). This, coupled with limitations of the traditional water-quality monitoring programs, makes a supplementary or an alternative method necessary for the water-quality monitoring of Lake Simcoe. In this study, a remote sensing technique plays the role of a supplement to the old methodologies to estimate the water quality parameters from the TM images. Also, many of those studies focused on phosphorus concentration itself, rather than considering other water quality parameters, which can also indicate the problems that Lake Simcoe is facing. As the phosphorus does not had obvious optical characteristics, the SDT and the concentration of chl-a are utilized here as indicators of eutrophication effects.

This research can demonstrate the spatial distribution of the eutrophic areas, as well as simulate the trends of water quality change in Lake Simcoe. This study can build a bridge between water-quality parameters (e.g., SDT and chl-*a*) and TM imagery. In year 2008, the Canadian government opened a 30-million fund for the Lake Simcoe clean-up, to reduce the environmental stresses (ENS, 2008). Therefore, the results of this study are of great value and significance to the to the whole environmental management of Lake Simcoe.

3.2 Satellite Images and Lake Reference Data

This project has had data support from Environment Canada and Ontario Ministry of Environment (MOE). In this study, a total of 5 cloud-free and well-calibrated Landsat 5 TM images were provided by CCIW, and 17 original Landsat 5 TM images were obtained from the United States Geological Survey (USGS) website free of charge. The in-situ data, such as the SDT, chl-*a* concentration, and DOC concentration, were provided by MOE.

3.2.1 Landsat-5 TM Images

With Regard to the spatial coverage of Lake Simcoe, Landsat 5 TM imagery is considered as appropriate. Although Landsat TM data cannot be used reliably to monitor short-term trends in water clarity, this satellite system is well suited to provide broad spatial coverage and multi-year trends (Kloiber et al., 2002). The satellite images used in this study were from Landsat 5 TM, which has been operating over the entire period (1987-2008). The TM images have a spatial resolution of 30m and temporal resolution of 16 days. They consist of

seven bands and their parameters are shown in Table 3.1. Each of these bands or the combination of them can be selected to detect and monitor different types of Earth resources. For example, the first band of TM data, penetrating water for bathymetric mapping along coastal areas, is useful for soil-vegetation differentiation and for distinguishing forest types. TM2 detects green reflectance from healthy vegetation, and TM3 is useful for detecting chlorophyll absorption in vegetation. TM4 data is ideal for detecting near infrared (NIR) reflectance peaks in healthy green vegetation and for detecting water-land interfaces. The two mid-IR red bands on TM 5 and TM7 are useful for vegetation and soil moisture studies and for discriminating between rock and mineral types. The thermal infrared (TIR) band on TM6 is designed to assist in thermal mapping, and is used for soil moisture and vegetation studies (USGS, 2009). Therefore, the TM1 to TM 4 are combined and used to estimate the water quality parameters in this study.

A total of 22 Landsat 5 TM images with a time range of 22 years, from April 19, 1987 to September 19, 2008, are employed in this study. Table 3.2 shows the metadata of these images, which includes the dates, cloud cover, sun elevation and band numbers. All these data were downloaded from the website of the USGS free of charge. The selection of data is based on the principle that the cloud coverage is less than 10%, as well as the almost simultaneous measurements of in situ samplings, within +/- 10days. Table 3.3 lists the dates of each pair of the satellite data and in-situ data of this study.

Table 3.1 Parameters of Landsat 5 TM images

Bands	Spectrum	Wavelength (μm)	Spatial Resolution (m)
TM1	Blue	0.45 - 0.52	30
TM2	Green	0.52 - 0.60	30
TM3	Red	0.63 - 0.69	30
TM4	Near IR	0.76 - 0.90	30
TM5	SWIR	1.55 - 1.75	30
TM6	LWIR	10.40 - 12.50	120
TM7	SWIR	2.08 - 2.35	30

Table 3.2 Metadata of Landsat 5 TM images

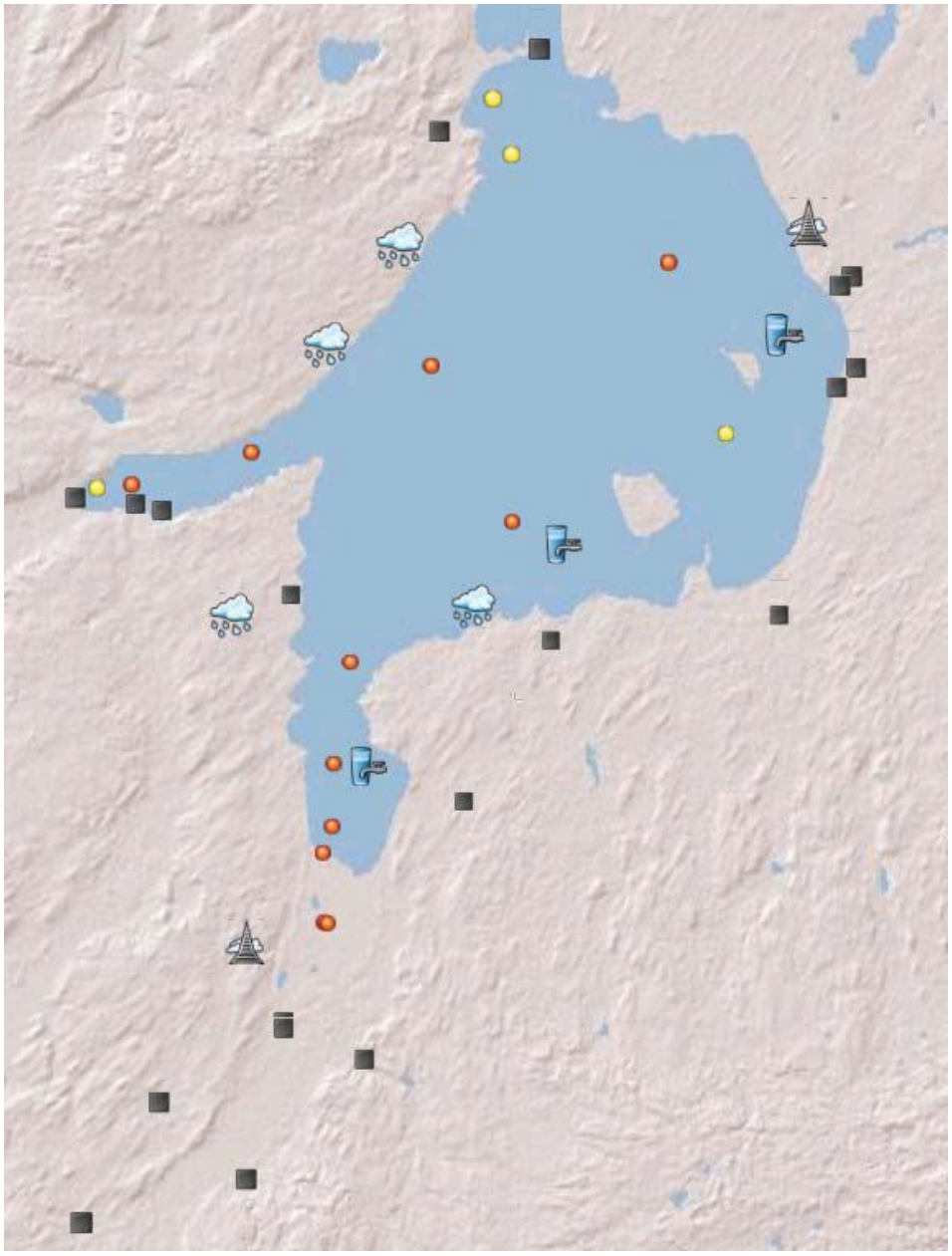
Date	Cloud Coverage	Sun Azimuth (°)	Sun Elevation (°)	No. of Bands
1987-04-19	0%	135.8392099	48.8680372	7
1987-05-05	0%	132.5496972	53.5759521	7
1988-05-07	0%	134.2458209	55.1267691	7
1989-06-11	0%	125.7547171	59.3775064	7
1990-08-17	0%	131.49719753	49.744003	7
1991-07-19	10%	124.59232065	55.98502985	7
1991-09-05	0%	139.27477421	45.2589758	7
1995-07-30	0%	121.13693138	51.27711498	7
1995-08-15	10%	125.73420458	47.95133289	7
1996-05-29	0%	122.74349888	56.2541567	7
2000-07-27	0%	131.1106679	56.4165954	7
2001-08-15	0%	137.67336349	52.71855079	7
2001-10-02	0%	152.7970659	38.04354591	7
2005-07-09	0%	131.90979757	60.66005591	7
2005-08-26	0%	144.07138577	50.63641318	7
2006-05-09	0%	142.04860848	58.10068656	7
2008-06-15	0%	132.5020829	62.13810488	7
2008-07-01	0.82%	131.34157724	61.39478308	7
2008-07-17	8.64%	132.60882833	59.39962323	7
2008-08-18	9.18%	140.70618293	52.51434213	7
2008-09-03	5.02%	145.90949289	47.91620946	7
2008-09-19	0.42%	150.83869111	42.75134619	7

3.2.2 Lake Reference Data

Ontario Ministry of the Environment (MOE) has been measuring all kinds of water quality parameters of Lake Simcoe for more than 20 years since 1982 and all these data are recorded in the documents.

It has been proven that the satellite data and the ground measurement data should be taken almost simultaneously, for the reason that water clarity usually does not exhibit large and rapid fluctuations in a given lake (Brezonik et al., 2007). Very fortunately, 22 of the measurements are found very close to the dates of the available Landsat 5 TM images of Lake Simcoe. There are a total of 15 ground sampling stations distributed on the lake (see Figure 3.3). Eleven of these stations, which are highlighted in red shown in Figure 3.3, are still in use while the other four yellow ones are the historical stations. Sampling is conducted routinely every two weeks, during the ice-free seasons through each of those years and the water would be analyzed in the laboratory to extract concentrations of different water quality indicators. The parameters include pH results, chl-*a* concentration, chl-*b* concentration, DOC and SDT. Both the chl-*a* concentration and the SDT are important indicators for lake water quality estimation. Table 3.3 shows the dates of both the image data and in-situ data used in this study, as well as the available quantity of measurement points. It reveals that the number of functional stations has been reduced since 1995, and there are only 11 stations left for the sampling of SDT, and 8 stations for

chl-*a*. Normally only the images within a 10-day overlap with the in situ data can be selected. However, there are three images with a longer period than 10 days compared with the ground measured data in Table 3.3. The reason for doing this is that these three images still have a good correlation with the in-situ measurements, and considering that the trends analysis is one of the main objectives of this study, they are kept in the list on the condition that they do not affect the accuracy of results.



Legend

Lake Water – Open Lake	Lake Water – Open Lake - Historical
Lake Water – Municipal Water Intakes	Meteorological + Deposition
Tributaries – Water Quality	Deposition

Figure 3.3 Location of the Sampling Stations on Lake Simcoe (Source: Ontario Ministry of the Environment)

Table 3.3 Available Stations for Each Pair of Data

Image Date	In-situ Date	No. of Stations		Image Date	In-situ Date	No. of Stations	
		SDT	Chl-*a*			SDT	Chl-*a*
1987-04-19	1987-05-21	15	15	2001-08-15	2001-08-18	11	8
1987-05-05	1987-05-21	12	12	2001-10-02	2001-09-28	11	8
1988-05-07	1988-05-08	12	12	2005-07-09	2005-07-11	11	8
1989-06-11	1989-06-12	12	12	2005-08-26	2005-08-28	11	8
1990-08-17	1990-08-15	15	15	2006-05-09	2006-06-05	11	8
1991-07-19	1991-07-17	15	15	2008-06-15	2008-06-21	11	8
1991-09-05	1991-09-04	15	15	2008-07-01	2008-07-05	11	8
1995-07-30	1995-07-25	12	12	2008-07-17	2008-07-19	11	8
1995-08-15	1995-08-08	12	12	2008-08-18	2008-08-16	11	8
1996-05-29	1996-05-20	12	15	2008-09-03	2008-08-30	11	8
2000-07-27	2000-07-30	11	8	2008-09-19	2008-09-14	11	8

3.3 Data Processing

Figure 3.4 presents an overall flowchart of the methodology used in this study. Both the satellite data and the in-situ data need to be preprocessed, which includes resampling, vector clipping, and the atmospheric correction for the satellite data; coupled with creating the point layers and the conversion to areas-of-interest (AOI) for the in-situ data. The DN values of the stations on the images are extracted through the AOIs and outputted as ASCII files, and then these values are compared with the in-situ data. Different models are used for the estimations of SDT and chl-*a* concentration, by utilizing linear regression analysis. Once the correlation coefficient R^2 is considered as high enough (more than 0.5 in this study specifically), a regression equation will be calculated and this equation can be used

for the whole image and generates a clarity or concentration map. Finally the result maps need to be smoothed, to remove the random or irregular pixels that seem to be discordant with the surrounding pixels. Therefore, a 5*5 median filter in ENVI software is used for the result display, which makes the maps more coherent and consistent. A more specific and detailed explanation for every single step in this flowchart will be given in the following sections. Sections 3.3.1 and 3.3.2 describe how the satellite images are pre-processed and how the atmospheric effects are removed, while Section 3.3.3 shows the steps of the in-situ data processing.

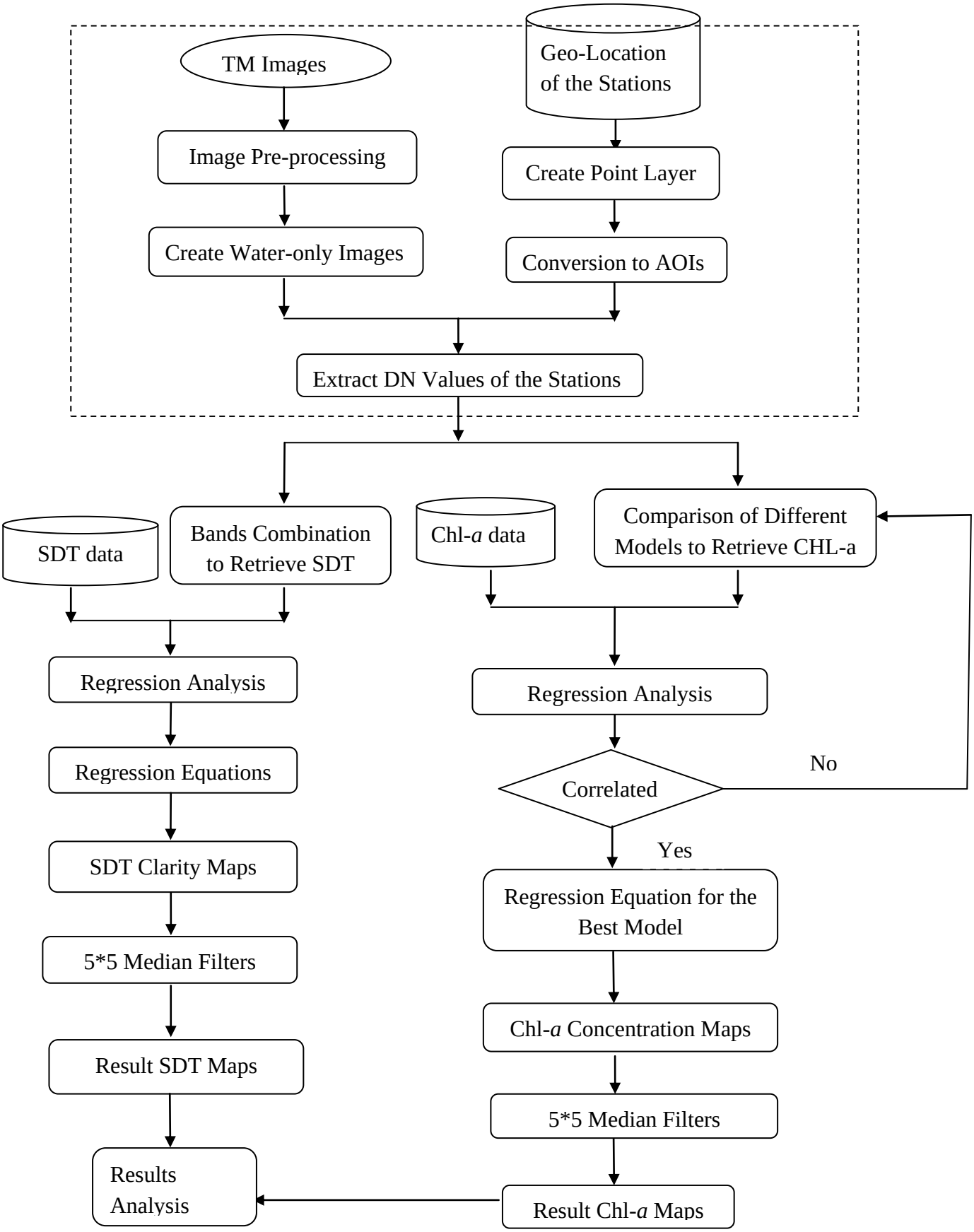


Figure 3.4 Flowchart of the Entire Image Processing

3.3.1 Satellite Image Pre-processing

Each TM image needs to be pre-processed and compared with the in-situ measured information. Both Figures 3.6 and 3.7 include the framework about how the TM images are pre-processed. The pre-processing procedure in this study includes three steps: resampling, registration, and the creation of water-only images.

Data Resampling

In remote sensing, as the given data are usually in one discrete space, the resampling methods are often utilized to represent the application data more accurately in the desired output discrete space (Trainer and Sun, 1991). Resampling is a process used to determine the pixel values to fill into the output matrix from the original image matrix (Lillesand et al., 2004). As most of the image data in this study are the raw data downloaded from the USGS website, firstly, all of these Landsat 5 TM images are resampled to 30m using the nearest neighbor resampling method available in ENVI 4.4. The nearest neighbor resampling uses the nearest pixel without any interpolation to create the warped image (ENVI, 2007).

Registration

The second step of the image pre-processing is registration, which is to reference the images to geographic coordinates and correct them to match the base image geometry. In this study, the image-to-image registration is available in ENVI 4.4. A total of 10 ground control points (GCPs) are selected, and all of the 17 raw images (the wrap images) are

geographically registered to the standard images (the base images) from CCIW. Figure 3.5 shows the distribution of GCPs in this study. In addition, these satellite images are uniformly projected to Zone 17N in the UTM (Universal Transverse Mercator Coordinate), with the datum of World Geodetic System 1984 (WGS84).

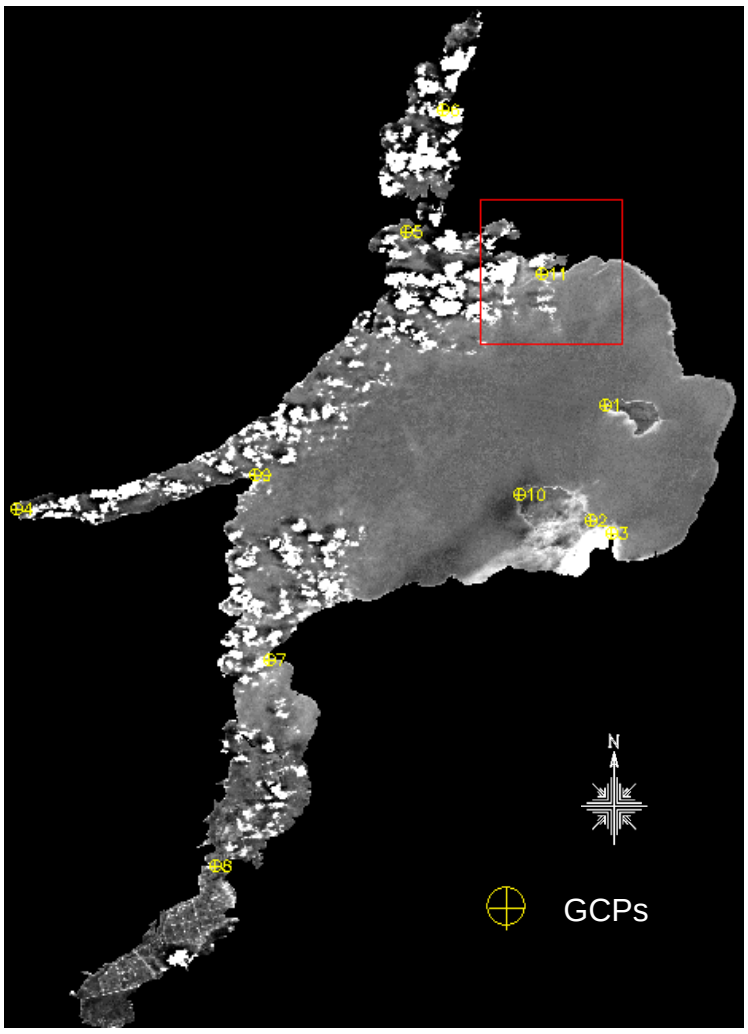


Figure 3.5 Distribution of Ground Control Points in Lake Simcoe

Creation of Water-Only Images

As Lake Simcoe only occupies a relatively small area of each of the satellite images, a vector boundary layer covering this lake and the southern part of Cook's Bay was created

using ArcGIS 9.2 software with the accordance of the well-calibrated satellite data. By using the same software package, each of the images is cut with this vector layer, and the water-only images are extracted. This step accelerates the speed of each calculation afterwards and gives a better colour display or distribution of the results.

3.3.2 Removal of Atmospheric Effects

The water quality information is usually so weak that some of the evitable influences, such as the errors caused by the atmosphere, need to be removed. Therefore, the atmospheric correction is performed to the original images after the pre-processing procedures.

The atmospheric correction is completed with the *ATCOR-2* module, which is an add-on module designed particularly for the atmospheric correction in flat terrains in the PCI Geomatica 10.1 package. The TM standard calibration file designed for Landsat 4 and 5 TM is selected for this correction. Calibrations of the images without a “header.dat” files are calculated by reference to the solar zenith and solar azimuth, according to the metadata of each image. In addition, two parameters need to be defined for this correction: the atmospheric condition of this study area is chosen as rural; while the conditions, as well as the thermal atmospheric definition, are selected as mid-latitude summer.

3.3.3 Areas of Interest from In-situ Data

The right upper part of Figure 3.4 shows how the water quality data in the text-file are transferred into visualized geographical information on the maps. As the in-situ data obtained from MOE include one geo-location file describing the 15 observation stations on Lake Simcoe, a point layer in accordance to the latitude and longitude of these 15 stations is created by ArcGIS 9.2 software. This vector layer can be transferred to an AOI by using ENVI 4.4, when it is overlaid with one of the satellite images. Then the DN values of the station locations on that image can be recorded to the ASCII files. Each of these ASCII files consists of the information of 7 bands, and those numbers can then be calculated or directly used to estimate the SDT and chl-*a* of Lake Simcoe.

3.4 Model Development and Validation

3.4.1 Models for SDT Estimation

In this section, the SDT is estimated from the satellite images, by utilizing both the general equation and the improved one in this study.

SDT Estimation using the General Equation

Figure 3.6 presents a flowchart showing more details about image pre-processing procedures, as well as the model development. After the data pre-processing and the creation of AOIs, the DN values of those specific stations are extracted from the images.

Then these DN values of different bands are combined and calculated based on Equation (1), which is the general predictive equation for water clarity of inland lakes, and then are compared with the ground measurements.

$$\ln(SDT) = a(TM1/TM3) + b(TM1) + c \quad (1)$$

Where, TM1 and TM3 are the brightness values measured by the Landsat sensor in the blue and red bands, respectively. *ln(SDT)* stands for the natural logarithm of the SDT. *a*, *b*, and *c* are the coefficients fitting to the in-situ data by the regression analysis. These coefficients can be used to estimate the SDT values for the entire lake in the same Landsat image, once they have been fitted to the equation. (Brezonik et al., 2002).

In this study specifically, however, the SDT value itself, rather than its natural logarithm, has a higher R^2 and therefore a better correlation with the satellite brightness values by correlation analysis in SPSS 16.0 software. So the in-situ SDT itself is compared with the estimated SDT from satellite images instead, and this relation can be described in Equation (2), where *a*, *b*, and *c* are the regression coefficients which will fit the ground truth data; TM1 and TM3 stand for the reflectance values of the TM1 and TM3 of the Landsat 5 TM images, and the SDT_{E1} refers to the estimated SDT from TM images.

$$SDT_{E1} = a(TM1/TM3) + b(TM1) + c \quad (2)$$

The SDT Estimation using the Improved Equation

A preliminary comparison is made between the reflectance values extracted from TM data and the in-situ measured SDT values. It shows that sampling points with low SDT are frequently accompanied with a high reflectance of TM3. Moreover, the reflectance values of TM3 alter following the change of SDT regularly, whereas the brightness data of TM1 do not show such a related trend. Therefore, TM3 is regarded as a better predictor than TM1 for the SDT estimation in this study. In Equation (3), TM3 reflectance values are subsequently employed to replace the TM1 reflectance in Equation (2), in order to improve the prediction accuracies of SDT.

$$SDT_{E2} = a\ (TM1/TM3) + b(TM3) + c \qquad (3)$$

where *a*, *b*, and *c* are the regression coefficients fitting the calibration data; TM1 and TM3 stand for the reflectance values of blue and red bands of the Landsat TM images, and the SDT_{E2} refers to the estimated SDT from TM images using Equation (3).

To summarize, based on the reflectance characteristics of the Landsat TM images in this study, as well as their similarities with the in-situ data, the generally used SDT equation is improved by replacing TM1 with TM3. Both of these two equations are used for the SDT estimations. The accuracies of these estimations are compared by utilizing the linear regression analysis which will be described in detail in Section 3.5.

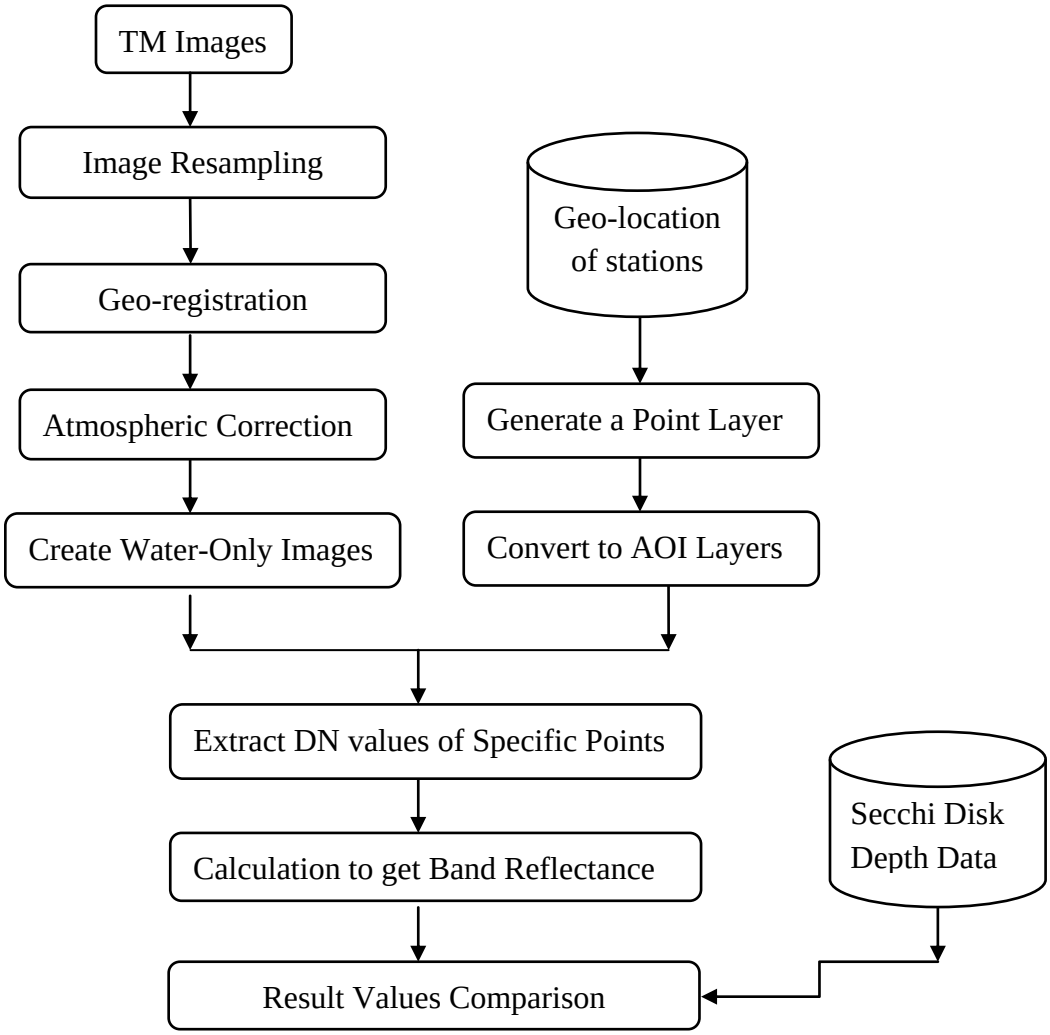


Figure 3.6 Flowcharts for SDT and Satellite Comparison

3.4.2 Models for Chl-*a* Estimation

Chlorophyll has a specific absorption and reflection spectrum. The absorption peak is near 440nm, the reflection peak near 550nm and a visible fluorescence near 685nm (Li et al., 2007). Therefore, these characteristics of absorption and reflection make TM1 to TM4 useful for the estimation of chl-*a* concentration. Figure 3.7 illustrates the work flow of estimating chl-*a* concentration from the Landsat 5 TM images. There have been a variety of

algorithms in previous studies to retrieve the chl-*a* concentration from the TM images, but it turned out that these models have different results in each image of this study. As no universal algorithm in this study is applicable to chl-*a* estimation, the predictive models for chl-*a* concentration are based on the empirical approach. In order to find at least one effective band combination for each of the images, different bands and their combinations are compared to get highest correlation coefficient R^2 for each pair of data. This best-fit model will then be chosen as the regression equation for that specific image and be considered as the one which can represent or estimate the chl-*a* concentration of that image. The predictive equation below can be used for the estimation of chl-*a* concentration,

$$\text{CHL-A}_E = kX+b \quad (4)$$

where k and b stand for the regression coefficients, CHL-A_E means the estimated chl-*a* concentration from the satellite reflectance data, and X refers to the bands combination with the highest correlation coefficient R^2 with the calibration data. A total of 9 bands combinations in total are compared to constitute the chl-*a* estimating equations of each image:

$$\text{CHL-A}_1=k\ (TM3*TM4) +b \quad (5)$$

$$\text{CHL-A}_2=k\ (TM3*TM4/\ln TM1) +b \quad (6)$$

$$\text{CHL-A}_3=k\ (TM3*TM4/\ln(TM1+TM2) +b \quad (7)$$

$$\text{CHL-A}_4=k\ (TM3/TM4) +b \quad (8)$$

$$CHL\text{-}A_5 = k\,(TM4 - TM3)/(TM4 + TM3) + b \quad (9)$$

$$CHL\text{-}A_6 = k\,(TM4/TM1) + b \quad (10)$$

$$CHL\text{-}A_7 = k\,(TM1/TM2) + b \quad (11)$$

$$CHL\text{-}A_8 = k\,(\ln(TM1/TM2)) + b \quad (12)$$

$$CHL\text{-}A_9 = k\,((TM1 - TM3)/(TM2 - TM3)) + b \quad (13)$$

where k and b are regression coefficients, TM1-TM4 are the reflectance value of TM band 1- band 4, respectively. CHL-A1 to CHL-A9 stand for the chl-*a* concentrations estimated by Equations (5) – (13).

Among these equations, Equation (5) is proposed by Chen et al. (1996), who indicated that TM3*TM4 was the most suitable variable for estimating chl-*a* concentration of seashore undersurface. Equations (6) and (7) are the improved equations used in the study of Li et al. (2007), which can possibly remove the effects caused by sediment scattering. Equation (9) utilizes the vegetation index as the independent variable. Equations (11) and (12) use TM1/TM2 as the predictor of chl-a concentration (Dwivedi and Narain, 1987; Li et al., 2007). In Equation (13), TM3 is subtracted from TM1 and TM2 before calculating the blue/green ratio to remove some of the scattering effects.

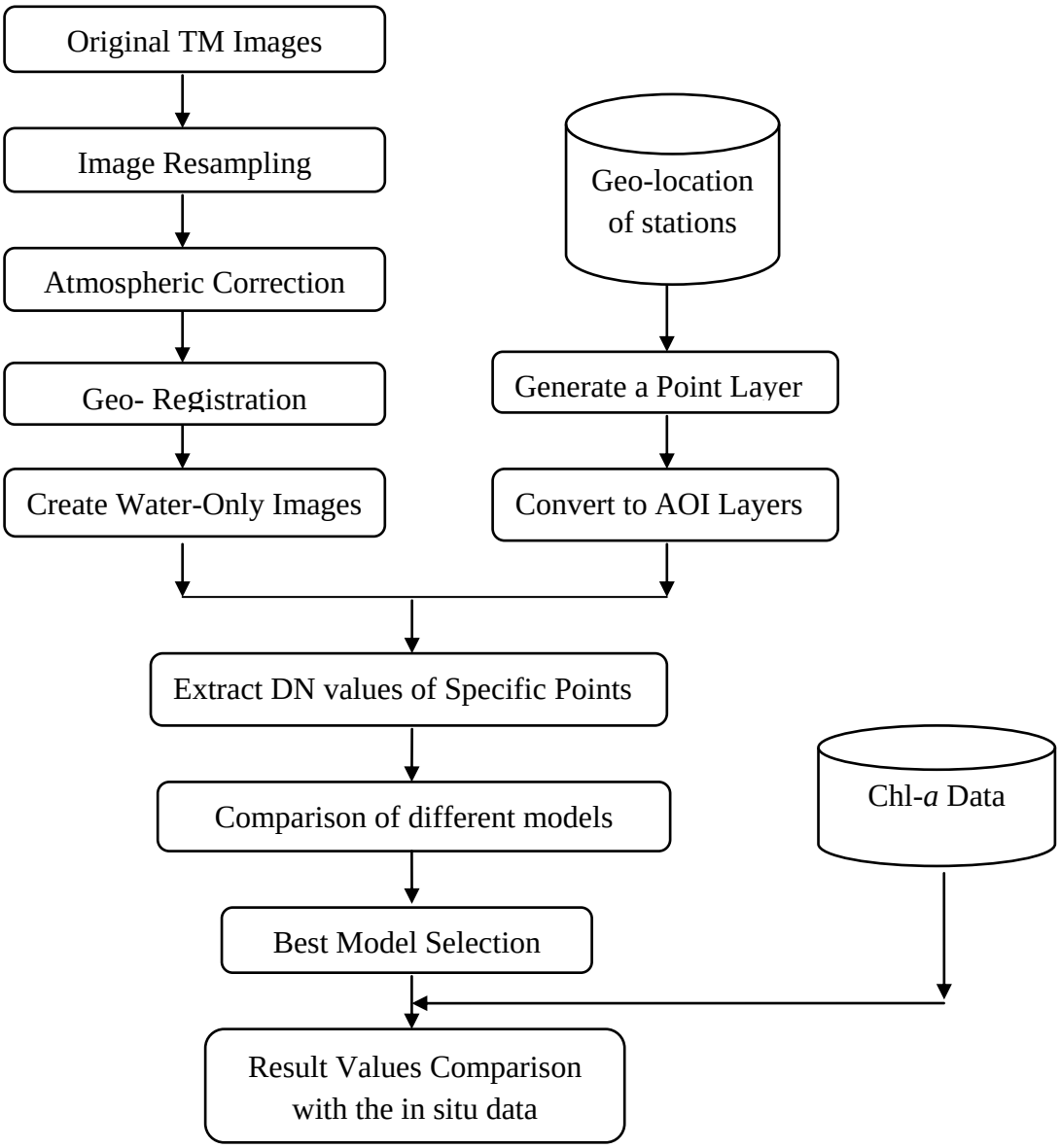


Figure 3.7 Flowchart of Chl-*a* Concentration Estimation

3.5 Regression Analysis

After utilizing different models to estimate SDT and chl-*a* concentration, the regression analysis is employed to find the best-fit equation to describe the relationship between each

of the estimated data and the corresponding in-situ data. The accuracies of these estimations are measured by the correlation coefficient R^2.

In statistics, the objective of regression analysis is to build a function of predictor (or independent) variables to express the response (or dependent) variable. Linear regression analysis, modeled by a least squares function, is to give information on the relationship between a dependent variable and one or more independent variables to the extent that information is contained in the data (Kaw and Kalu, 2008).

The correlation between two variables reflects the degree to which the variables are related and the Pearson Product Moment Correlation (called Pearson's correlation for short) is a commonly used measure of correlation. The correlation coefficient measures the degree to which two things vary together or oppositely. The maximum positive correlation is 1.00. If the two standardized variables covary positively and perfectly, or vary oppositely and perfectly, then the correlations will both equal to 1.00 (Rummel, 1976). When computed in a sample, it is designated by the letter "r" and is sometimes called "Pearson r". Pearson r reflects the degree of linear relationship between two variables and it ranges from +1 to -1. A correlation of "+1" means that there is a perfect positive linear relationship between variables. In a scatter plot depicting this relationship, the relationship is positive when the scores on the X-axis are high associated with high scores on the Y-axis.

Each of these 22 images is compared with the corresponding in-situ measured data. The multiple linear regression analysis is conducted under the environment of the statistical analysis software SPSS 16.0. Two couples of variables are compared by correlating to the calibration data, one couple is the reflectance value of TM1 and TM1/TM3, while the other is the reflectance value of TM3 and TM1/TM3. The couple with higher coefficients R^2 will be utilized as the independent variables for linear regression analysis, and the in-situ SDT will be employed as the dependent variable. The resulting regression equations will be applied to the entire lake area of those images in order to map and stimulate the SDT distribution of Lake Simcoe. Similar comparisons are conducted for the chl-*a* concentration estimation, while the combinations of reflectance value are more complex which include TM3*TM4 (Chen et al., 1996), TM3*TM4/lnTM1(Li et al., 2007), TM3*TM4/ln(TM1+TM2) (Li et al., 2007), TM3/TM4, (TM4-TM3)/(TM4+TM3), TM4/TM1, TM1/TM2 (Dwivedi and Narain, 1987), ln(TM1/TM2) (Li et al., 2007), and (TM1-TM3)/(TM2-TM3).

3.6 Spatial and Temporal Analysis

By applying the regression equations to each of the images in the database, the spatial patterns, as well as the temporal trends of SDT and chl-*a*, can be investigated.

3.6.1 Spatial Patterns of SDT and Chl-a

The spatial patterns of both SDT and chl-*a* are displayed in terms of the concentration maps. These maps are generated by employing each of the regression equations to calculate its corresponding image, on condition that these equations are reliable enough. A 5*5 medium filter is utilized for the resultant maps individually, in order to reduce the noise and make the output images smoother.

3.6.2 Temporal Trends of SDT and Chl-a

The seasonal patterns and temporal trends of water quality in Lake Simcoe are both analyzed in this study, as the available data for Lake Simcoe have a long time span of 22 years. These patterns and trends of SDT and chl-*a* can be determined in terms of both the line charts and the concentration maps. The line charts are calculated from the monthly averaged SDT and chl-*a* collected by the ground stations, using Microsoft Excel 2007. The concentration maps are based on the regression equations, see Section 3.6.1.

3.7 Chapter Summary

In this study, Lake Simcoe is selected as the study site because of its eutrophication problem, the reduction of ground stations, and the data availability, as well as the lack of remote sensing studies. A total of 22 Landsat 5 TM images ranging from 1987 to 2008, are chosen as the satellite data on condition that the cloud coverage is no more than 10%. The

calibration data, including SDT, chl-*a* and DOC concentrations, are selected when the in-situ measurements are approximately simultaneous (±10days) with the overpass time of the satellite.

All the images are pre-processed, including the data resampling, the water-only image creation, and the atmospheric correction. The empirical models based on different band combinations are developed or used to estimate the SDT and chl-*a* concentration from the satellite images. The accuracies of these estimations are measured by the correlation coefficient R^2.

Besides the general SDT prediction model, the model utilized in this study is improved by replacing TM1 with TM3. For the chl-*a* concentration, 9 current models are compared, so as to select one best model to estimate chl-*a* from each image. By performing the linear regression analysis, equations fitting the linear relationship between the reflectance values and the calibration data are developed and then applied to each image, for the purpose of estimating the SDT or chl-*a* concentration of Lake Simcoe.

Chapter 4

Results and Discussion

This chapter presents the results and explains the findings of this study. The beginning two sections describe and discuss the results of linear regression for the SDT and chl-*a* estimations. Section 4.3 graphs and analyzes the spatial distribution of eutrophic areas, as well as the trends of SDT and chl-*a* over these years. A lake water classification based on the trophic state criteria is depicted in Section 4.4. Section 4.5 discusses the influence of DOC on the water quality estimation. Finally this chapter is summarized in Section 4.6.

4.1 Evaluation of Lake Clarity Estimation

In this section, the accuracies of SDT estimation are evaluated. The SDT calculated from the Landsat images are compared to the in-situ measurements in three ways, which include: the individual comparison, the comparison for year 2008, and the comparison of the entire dataset.

4.1.1 SDT Evaluation for Individual Datasets

A Comparison between Two Models

Regression equations relating each TM image to the corresponding in-situ SDT are listed in Table 4.1, which includes the R^2, and *SEE* as well. The results indicate that the estimations using both models have good agreements with the in-situ SDT. The generally used SDT

model is related to the ground measurements with the R^2 ranging from 0.225 to 0.836, the average R^2 of 0.545 and the average standard error of estimation (*SEE)* of 1.545. On the other hand, the improved model in this study has the R^2 ranging from 0.232 to 0.861, the average R^2 of 0.622, and the average *SEE* of 1.374. In other words, the composed TM3 and TM1/TM3 is better correlated to the calibration data than combination of TM1 and TM1/TM3 in general. Additionally, 68.2% (15 out of 22 images) of these comparisons have higher R^2 by using the combination of TM3 and TM1/TM3, while only 7 images have higher R^2 by utilizing the general predictive variables for SDT estimation. Furthermore, it is impressive that a total of 5 comparisons of the commonly used combination, highlighted in Table 4.1, show poor correlations with the calibration data, while the improved one in this study has significantly better correlations instead. These comparisons are illustrated in the form of scatter plots in Figure 4.1, including the data of 1991-09-05, 2001-08-15, 2005-07-09, 2008-07-17 and 2008-09-03. In these scatter plots, 1991-09-05(a), 2001-08-15(a), 2005-07-09(a), 2008-07-17(a) and 2008-09-03(a), on the left side, are the estimations made by the general model, while 1991-09-05(b), 2001-08-15(b), 2005-07-09(b), 2008-07-17(b) and 2008-09-03(b) are the estimations made by the improved model. Each of these scatter plots contains a regression line, which is the equation made by linear regression. If the estimated SDT is well related to the ground-based SDT, the scatter points should be close to the reference line. Also, the regression line should be close to the line of 1:1, if the estimation is reliable. Thus, these

comparisons show that the improved model has a better agreement with the calibration data, which indicates that the TM3 and TM1/TM3 combination can greatly improve the accuracies of SDT estimations from Landsat 5 TM images in this study.

The results of multiple regression analysis for Landsat 5 TM imagery and the in situ SDT data also show that the reflectance value of TM1 has a much lower partial correlation than the ratio of TM1 and TM3 by comparison to the observed ground data. This disagreement is probably due to the high concentration of dissolved organic carbon in this lake, which will significantly influence the absorption of the blue light and therefore lower the reflectance of TM1. In this case, the partial correlation results would not have a good relationship with this band.

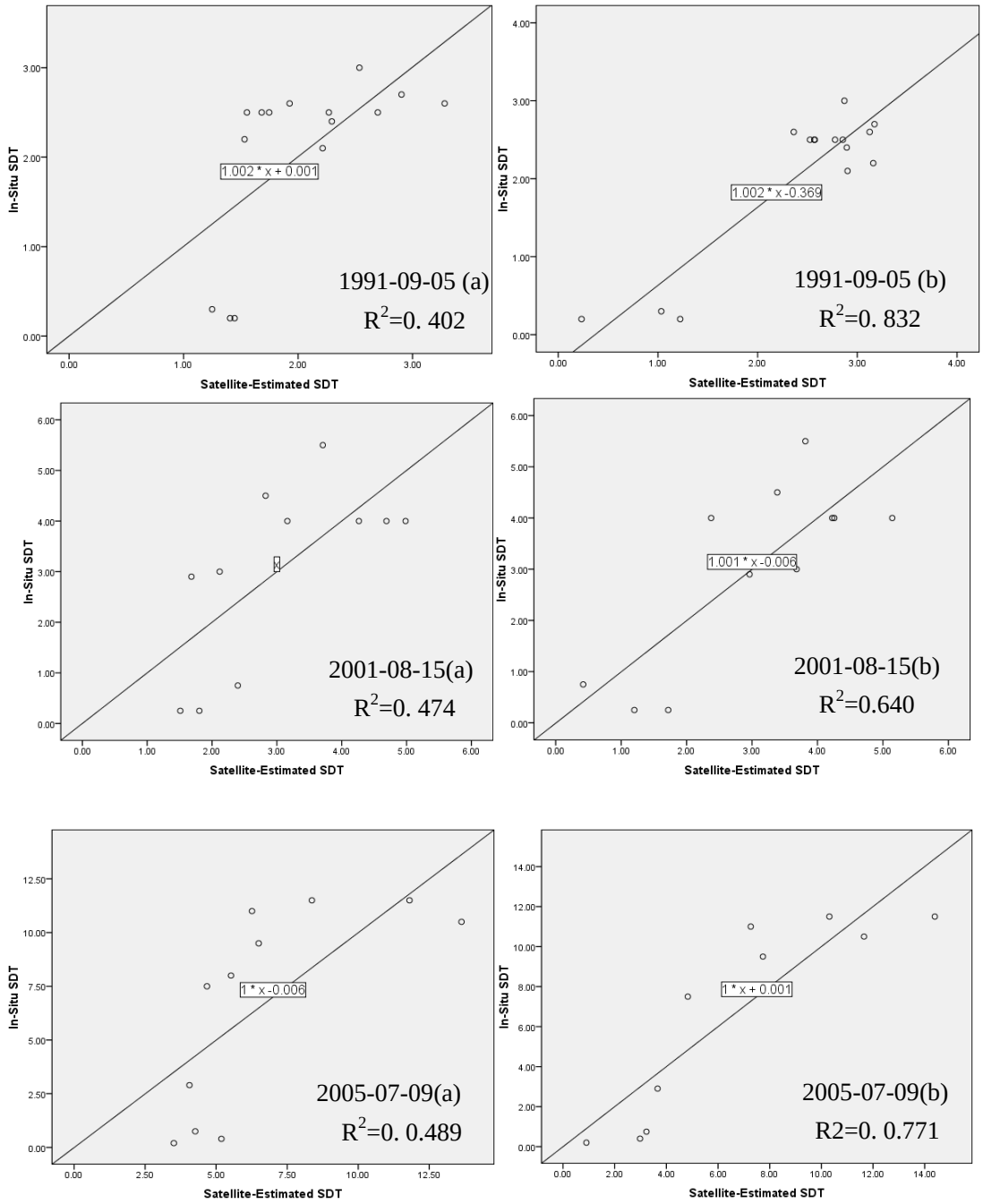


Figure 4.1 Scatter Plots Comparisons for SDT Estimations

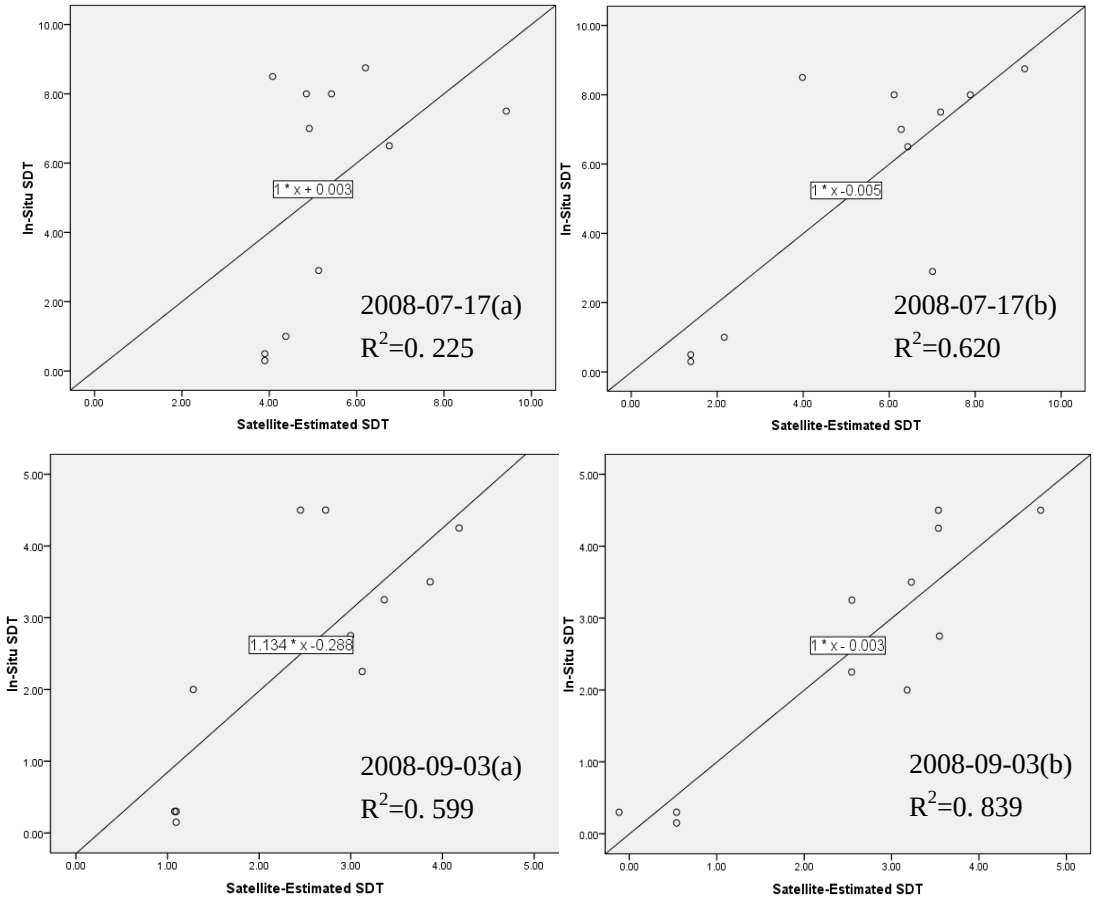


Figure 4.1(Cont.) Scatter Plots Comparisons for SDT Estimations

Table 4.1 Regression Equations for Prediction of SDT from TM Images

Date	R^2	SEE	R^2	SEE	Regression Equations
Variables	(TM1,TM1/TM3)		(TM3, TM1/TM3)		B1=TM1; B3=TM3
1987/04/19	0.502	1.34297	0.432	1.43450	2.086(B1/B3)+0.241B1-21.272
1987/05/05	0.609	1.18947	0.615	1.18075	0.507(B1/B3)+0.053B3+0.381
1988/05/07	0.524	0.59201	0.578	0.55720	-2.002(B1/B3)-0.652B3+22.027
1989/06/11	0.457	0.97484	0.453	0.97771	0.948(B1/B3)-0.140B1+11.221
1990/08/17	0.846	0.50091	0.839	0.51252	2.187(B1/B3)+0.096B1-6.109
1991/07/19	0.842	0.54136	0.714	0.73008	2.411(B1/B3)+0.025B1-3.433
1991/09/05	0.402	0.80485	0.832	0.42637	-0.032(B1/B3)-0.199 B3+3.842
1995/07/30	0.169	1.27560	0.248	1.21342	-0.011(B1/B3)-0.554B3+4.967
1995/08/15	0.718	0.41423	0.701	0.42689	0.215(B1/B3)+0.088B1+1.884
1996/05/29	0.323	1.11368	0.343	1.09725	-1.846(B1/B3)-0.587B3+10.488
2000/07/27	0.714	1.55438	0.834	1.18621	-5.159(B1/B3)-0.634B3+18.816
2001/08/15	0.474	1.46664	0.640	1.21379	-0.196(B1/B3)-0.254B3+6.141
2001/10/02	0.597	2.04098	0.600	2.15759	6.757(B1/B3)+0.103B3-1.298
2005/07/09	0.489	3.75593	0.771	2.51644	-2.120(B1/B3)-0.621B3+17.133
2005/08/26	0.598	1.90506	0.586	1.93198	2.373(B1/B3)+0.036B1+1.453
2006/05/09	0.707	2.08672	0.776	1.82628	1.956(B1/B3)-0.192B3+3.722
2008/06/15	0.445	2.50248	0.402	2.59570	1.643(B1/B3)+0.231B1+0.503
2008/07/01	0.225	3.18481	0.232	3.16913	0.009(B1/B3)-0.416B3+6.829
2008/07/17	0.225	3.38402	0.620	2.36985	-1.149(B1/B3)-0.919B3+13.861
2008/08/18	0.671	1.22578	0.770	1.02469	1.645(B1/B3)-0.477B3+5.381
2008/09/03	0.599	1.18845	0.839	0.75380	0.024(B1/B3)-0.328B3+4.794
2008/09/19	0.843	0.93669	0.861	0.93488	7.088(B1/B3)+0.218B3-2.231

Seasonal Influences to SDT Estimation

In Table 4.1, the estimations of three datasets (1987/04/19, 1989/06/11 and 2008/06/15) are not reliable enough that the R^2 of them are between 0.4 and 0.5. There are even two comparisons (1995/07/30 and 2008/07/01) with the R^2 lower than 0.250.

Since all the Landsat images have been corrected to remove the atmospheric affects, the possible reason to explain these poor agreements is the seasonal influence. Olmanson et al. (2007) and Stadelmann et al. (2001) pointed out in their studies of Minnesota Lakes that the late-summer period is the best index period for the remote sensing of water clarity. Two reasons are concluded for this perspective: one is that most lakes have their worst water clarity in this period; the other is that the lake water clarity has the minimum short-term variability in the late-summer season (Olmanson et al., 2007). Their point is verified in this study as well. The employed satellite data have a seasonal variability, which include the early spring, late spring, early summer, late summer and even fall. In this study, the late-summer is defined as August and September. The results in Table 4.1 show that the vast majority of comparisons in the last-summer period have very close relationships, whereas those badly related data are mostly collected in the spring or early summer. For Lake Simcoe, as the early spring is the period right after the snow melting, it is understandable that there may be an internal mixing in the lake. Also, locations in a lake can vary due to wind mixing fairly rapidly. All of these factors can contribute to the instability of lake water and the low agreements between the satellite-estimated SDT and the in-situ SDT.

4.1.2 SDT Evaluation for 2008

The available data in this study are more sufficient in 2008 than the other years. A total of 6 images and the corresponding in-situ data are collected in 2008, from the very early summer to fall. This section shows the comparison between these two datasets.

The scatter plot in Figure 4.2 shows that the satellite estimated SDT and the in-situ SDT have a fair agreement with the R^2 of 0.606. Figure 4.3 is the similar comparison, which only uses the late-summer data in 2008. The linear relationship is much closer than the previous one and the R^2 increases to 0.84. The late-summer comparison has a better correlation in that most of the samples are close to the reference line. Therefore, it can support the suggestion that the late-summer can be regarded as a better period for the lake water clarity prediction in 2008. However, compared with Figure 4.4, which is the scatter plot of in-situ SDT vs. the Estimated SDT for late-summer period in the entire database, the data in 2008 still have a better agreement. The gap between these two results indicates that the close dates can lead to similar water clarity conditions, as the lake water quality does not show big difference during a short period.

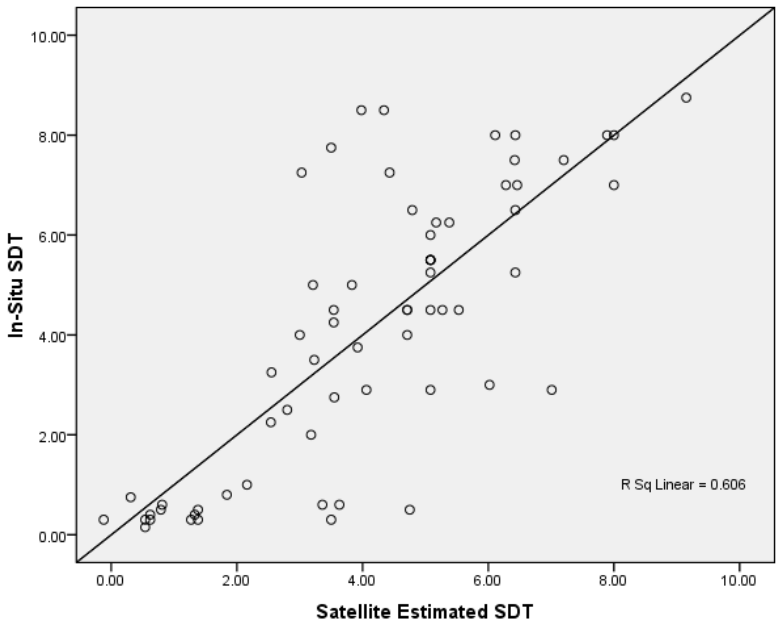


Figure 4.2 Comparison between In-situ SDT and Estimated SDT for 2008

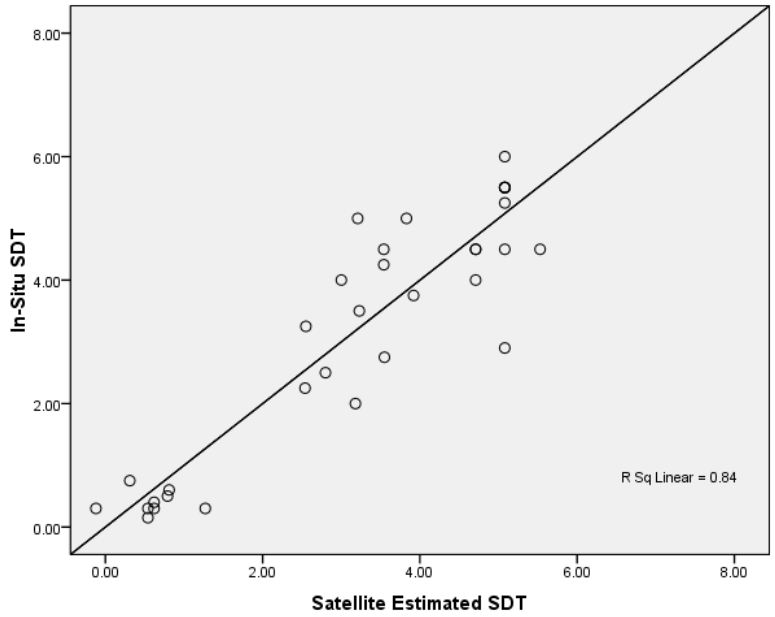


Figure 4.3 Comparison between In-situ and Estimated SDT for late-summer 2008

4.1.3 SDT Evaluation for Entire Database

This section shows the comparison results for the estimated SDT and in-situ SDT of the entire database. Firstly, all of the data of the late-summer period are employed in Figure 4.4, and then the dataset extended to the entire database in Figure 4.5. Both pairs of the datasets have good agreements, with the R^2 of 0.747 and 0.712, respectively. Therefore, it can be concluded that the Landsat 5 TM images are reliable enough to predict the water clarity in Lake Simcoe. Also, the advantage of sampling in the late-summer is demonstrated again, as the comparison in that period has a higher R^2 than the others. In the late-summer period, the lake usually has its worst water clarity (Olmanson et al., 2007), are there are more lights and signals reflected to the remote sensor. Therefore, the relationships between the reflectance values and in-situ data become stronger.

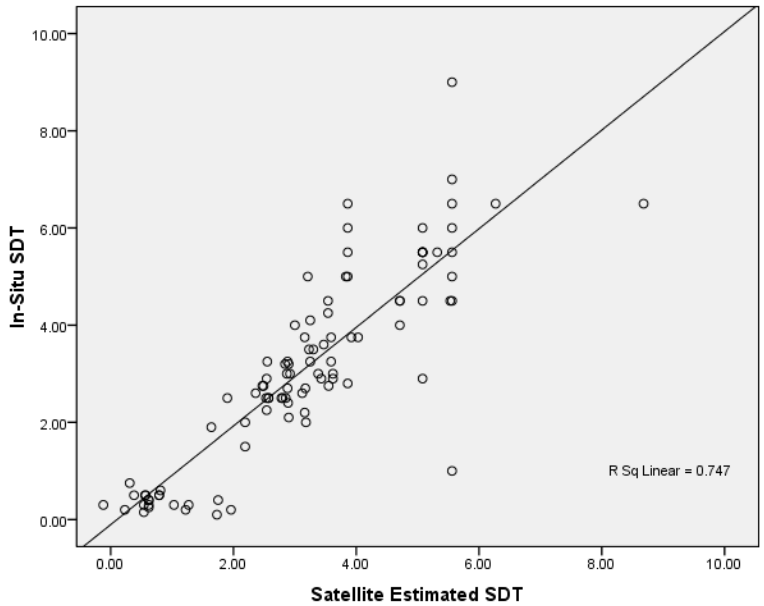


Figure 4.4 Comparison between In-situ SDT and Estimated SDT for late-summer period

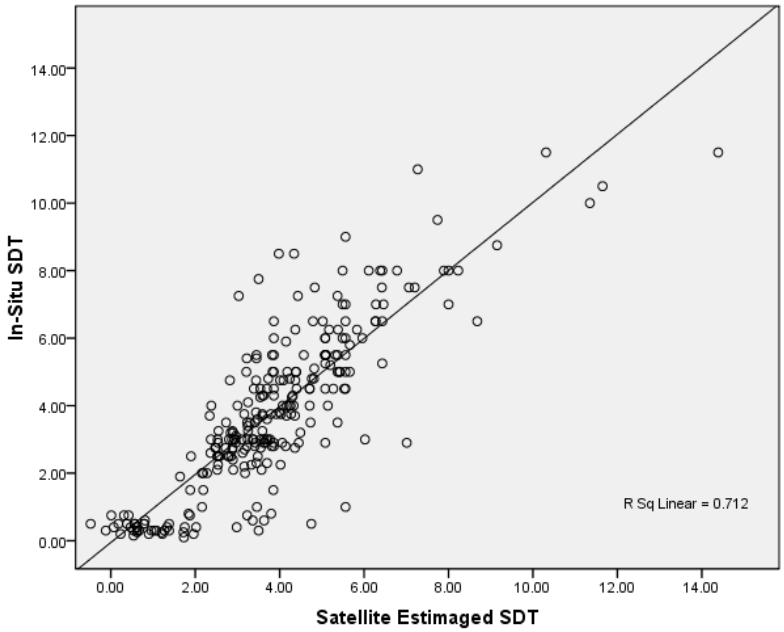


Figure 4.5 Comparison between In-situ and Estimated SDT for Entire Database

4.2 Evaluation of Estimates of Chl-a Concentration

4.2.1 Chl-*a* Evaluation for Individual Datasets

The chl-*a* concentration estimated by 9 models composed of different bands have been compared to the in-situ chl-*a* data, and none of them show an obvious superiority than the other eight. Therefore in this study no model is proven to be the best or most efficient model to estimate the chl-*a* concentration from the TM images. Table 4.2 lists the resultant parameters of these comparisons, which include the correlation coefficient *R*, R^2, *SEE*, best model and the regression equations for each pair of data, respectively. The results show that the best correlations between the Landsat TM images and the in-situ calibration data range from 0.142 to 0.941.

It is apparent that not every model suits all the imageries, however, each of those imageries have at least one best model with highest coefficient R^2 , which mean that model is possibly the most reliable one to predict the chl-*a* in that specific image. For example, Figure 4.6 shows four different models (B3*B4/lnB1, B3/B4, B3*B4 and Normalized Difference Vegetation Index (NDVI)) for the TM image of 1995-07-30, compared with the chemical data of July 23th, 1995. Those four comparisons have the correlation coefficients R^2 of 0.941, 0.486, 0.083, and 0.273, respectively. The prediction made by the model of B3*B4/lnB1 has a very good linear relationship with the chemical data, and the scatter points on the first plot seem to be close to the line, which means their relationship is probably linear. By contrast, in the other three scatter plots, the points are irregularly distributed and quite far-away from the line, so the estimated chl-*a* information extracted with those three models is in a very weak relationship with the in situ observed data.

This comparison above indicates that the B3*B4/lnB1model is the best model for retrieving chlorophyll-a concentration from the TM image of 1995-07-30. Similar studies have been conducted for the rest of 21 images, and the best models for each image are listed in Table 4.2.

Table 4.2 Results of Regression Analysis for Chl-a Estimations

Date	R	R^2	SEE	Best Model	Regression Equation (where X refers to the best model)
1987/04/19	0.594	0.353	15.17618	B3/B4	Y=-22.533X+45.088
1987/05/05	0.850	0.722	9.94329	B2/B1	Y=-12.737X+41.296
1988/05/07	0.642	0.412	0.90519	B2/B1	Y=-4.588X+17.094
1989/06/11	0.929	0.863	10.20124	B2/B1	Y=-156.791X+60.590
1990/08/17	0.779	0.607	13.69362	NDVI	Y=75.034X+11.139
1991/07/19	0.753	0.568	9.99204	NDVI	Y=43.284X+7.874
1991/09/05	0.795	0.632	43.95226	(B1-B3)/(B2-B3)	Y=-103.400X+91.043
1995/07/30	0.970	0.941	0.67165	B3*B4/LnB1	Y=0.774X+0.569
1995/08/15	0.733	0.538	0.76811	B1/B4	Y=0.875X+1.429
1996/05/29	0.783	0.614	15.52595	B1/B4	Y=21.215X-24.228
2000/07/27	0.710	0.505	5.57228	B3*B4	Y=0.065X-0.954
2001/08/15	0.464	0.216	1.05171	NDVI	Y=2.760X+3.101
2001/10/02	0.802	0.643	7.32872	B3*B4/LnB1	Y=0.575X-4.021
2005/07/09	0.377	0.142	0.35374	(B1-B3)/(B2-B3)	Y=0.175X+1.224
2005/08/26	0.557	0.310	0.34220	B3/B4	Y=-0.752X+3.421
2006/05/09	0.468	0.219	1.10766	B3*B4	Y=0.077X+2.313
2008/06/15	0.569	0.324	0.41241	B1/B4	Y=0.578X+1.404
2008/07/01	0.481	0.232	0.58793	B1/B4	Y=0.266X+2.082
2008/07/17	0.454	0.206	0.17637	B1/B2	Y=-0.794X+1.776
2008/08/18	0.748	0.559	0.61880	Ln(B1/B2)	Y=-3.986X+0.405
2008/09/03	0.838	0.702	1.11818	(B1-B3)/(B2-B3)	Y=-7.294X+5.758
2008/09/19	0.965	0.932	0.27899	B3*B4/LnB1	Y=0.043X+2.120

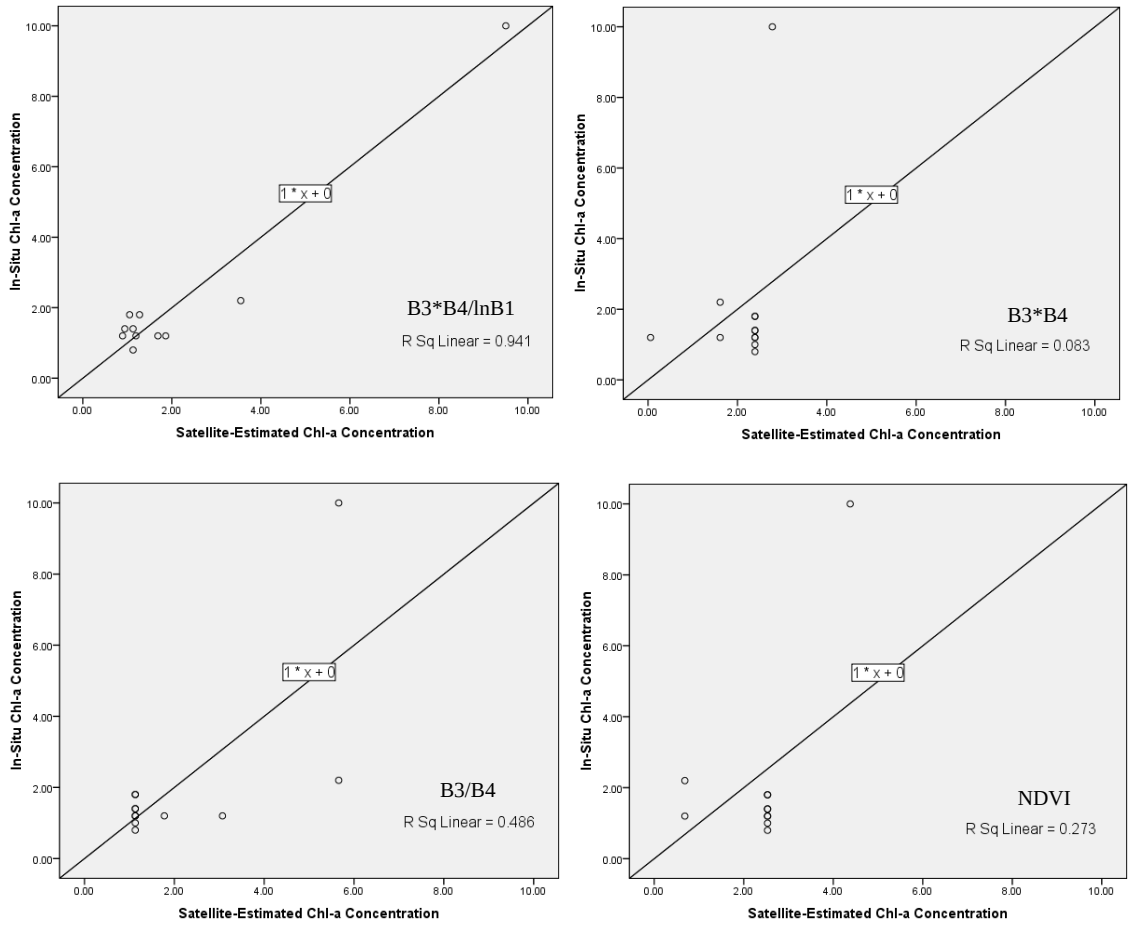


Figure 4.6 Comparison of In-situ and Estimated Chl-*a* from Four Models

4.2.2 Seasonal Influences for Chl-*a* Estimations

The estimation of chl-*a* concentration for Lake Simcoe also shows seasonal patterns. Calculated from Table 4.2, the chl-*a* comparisons have an average R^2 of 0.511. The late-summer comparisons have the average R^2 of 0.523, while the non-late-summer ones have the average R^2 of 0.492. The slight difference between these two sampling periods is enlarged when the data are limited within the same year. Take year 2008 with a total of 6 comparisons for example, the late-summer data pairs have an average R^2 of 0.731, whereas

the non-late-summer ones only have the mean R^2 of 0.254. The possible reasons for these seasonal lake-water variations include the lake geometry, air temperature and precipitation (Stadelmann et al., 2001). In the late-summer period, the precipitation is decreased, while the temperature of the lake is maximized. Therefore, more reflectance can be send back to the remote sensor. For the Lake Simcoe study, because there is a lack of these data, the relation between them and the water quality variations cannot be statistically analyzed. Nevertheless, the regression results show that chl-*a* has similar seasonal patterns with SDT.

4.2.3 Chl-*a* Evaluation for Entire Database

The estimation for chl-*a* concentration is not as good as the SDT prediction. Besides several poorly related individual comparisons, there are even 5 comparisons with the *SEE* larger than 10.00. The predictions from 9 models for 1991-09-05 have the *SEE* ranging from 43.952 to 70.247, respectively, which means this dataset is not suitable for the Chl-*a* estimation. Therefore, the data from this date are removed from the chl-*a* estimation for the entire database.

Figure 4.7 presents the scatter plot of all the satellite estimated chl-*a* concentration compared with the ground-sampled chl-*a*. The resultant R^2 is 0.632 and the *SEE* is 14.335. Highlighted in this figure, the regression fit line has a larger slope than the reference line

“y=x”. This offset indicates that the chl-*a* concentration is underestimated in this study, as the vertical coordinate of each point is greater than the horizontal one.

The colour in the lake water can increase the absorption of light and therefore decrease the signal back to the remote sensors (Moore, 1980). Therefore, this underestimation maybe due to the chl-*a* itself as its concentration is normally high in Lake Simcoe, not to mention the hyper-eutrophic water in the southern bay. Besides the phytoplankton, DOC in the lake water is another absorber of light (Kondratyev et al., 1998). The monthly averaged DOC concentration in Lake Simcoe was above 4mg/L in the past, and it began to gradually increase to 6mg/L since the late 1990s. The reflected light can be decreased by this matter, and the signals come back to the sensor are therefore weakened.

Brezonik et al. (2007) mentioned in their study that the reason for those low agreements can include the atmospheric conditions (e.g., haze and water vapor content), which will affect the light reflected by land and water surfaces as it travels back toward the satellite sensors. In this research, some of the images have slight cloud coverage between 0 and 10%. Moreover, the water vapor content condition is unknown. These can both be the possible reasons for the low agreements between estimated data and in-situ data.

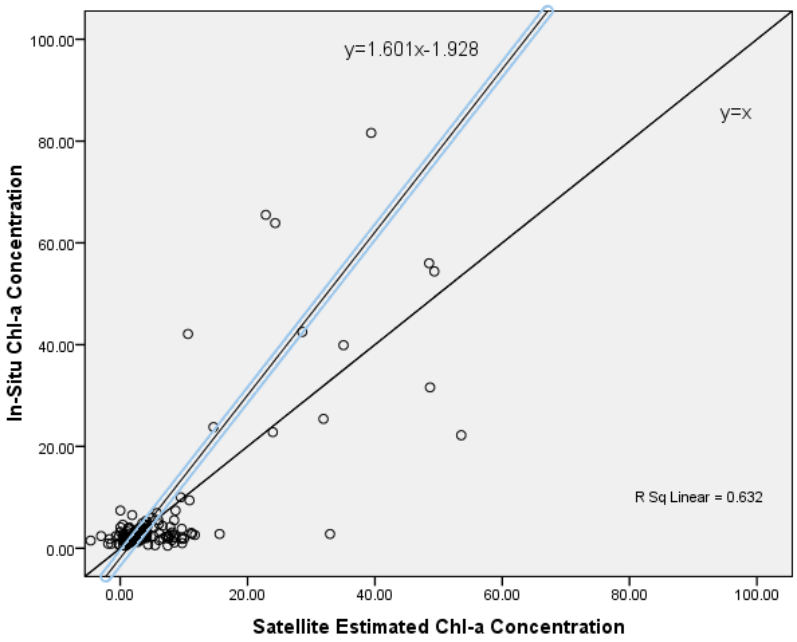


Figure 4.7 Comparison between In-situ and Estimated chl-*a* for Entire Database

4.3 Spatial and Temporal Analysis

4.3.1 Spatial Distribution of Polluted Areas

The concentration map estimated by the regression equations can indicate the spatial distribution of the eutrophic areas. Figures 4.8 and 4.9 are the blue/white and the green/white color maps displaying the SDT and chl-*a* concentration estimated from the regression equations on 1991-08-17. In the blue/white map, the lighter the colour is, the clearer the water is. In the other map, the lighter the colour is, the higher the chl-*a* concentration is. This means that the dark blue area has the lowest SDT and the dark green area has the lowest chl-*a* concentration. From the blue/white map, it is obvious that the middle part of the lake body is normally clearer than the near-shore area, and the turbid area is concentrated at the northeast shore and the southern near-shore area. Both of the

maps show that the SDT and chl-*a* concentration are related, because in the same area, the higher the chl-*a* concentration is, the lower the SDT clarity usually is. And the area of low chl-*a* concentration is normally with a high SDT value. This characteristic is more obvious when the water is away from the shore, as there are frequently confusions between the land-pixels and lake-pixels. In the scale of the whole Lake Simcoe, the northeast part of this lake has worse water quality than the southwest part. This indicates that the northeast part of Lake Simcoe is suffering from a more severe water quality problem than the southwest part. It can also be found that the water quality is turning bad gradually following the direction from northeast to southwest.

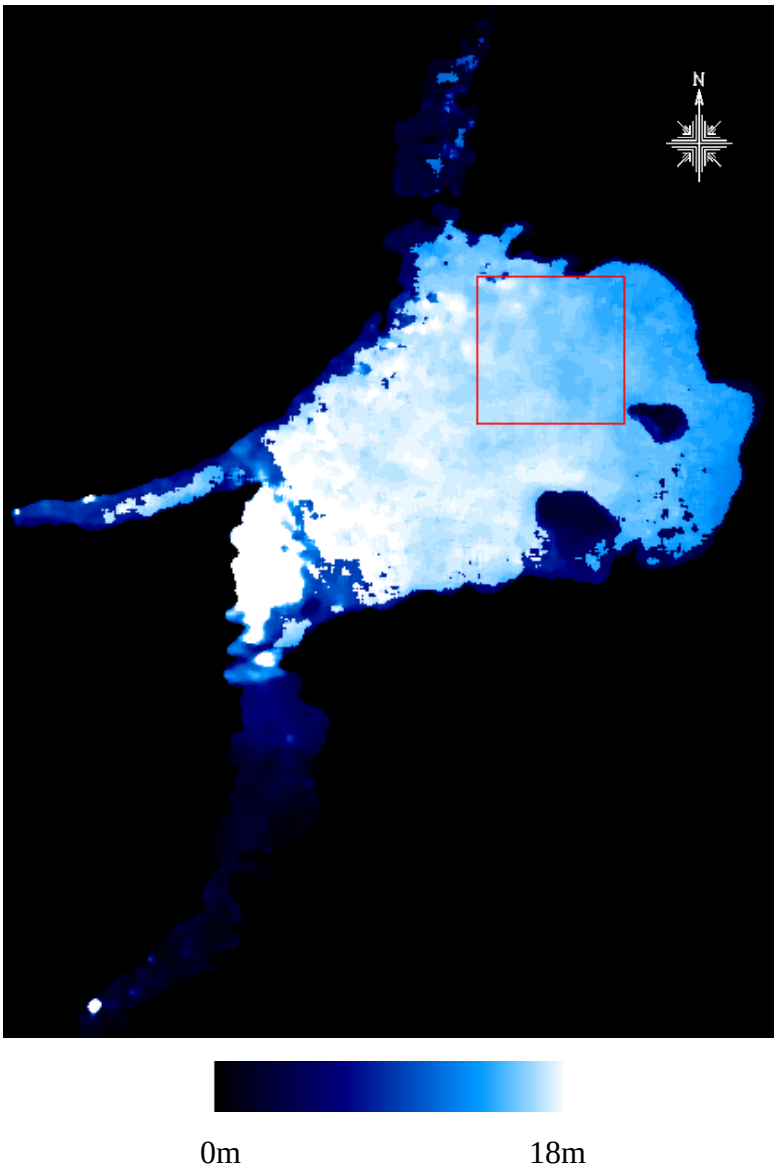


Figure 4.8 Blue/White Map of the Estimated SDT of 1991-08-17

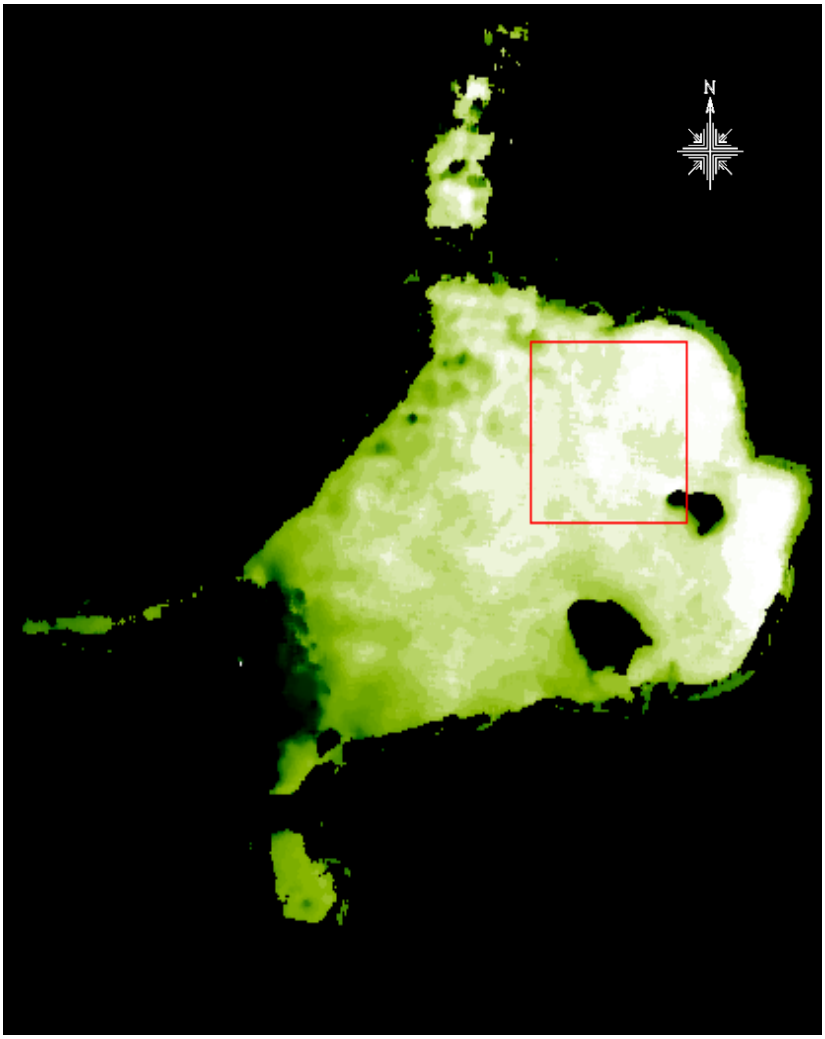


Figure 4.9 Estimated Chl-*a* Concentration Map of 1991-08-17

4.3.2 Trends of SDT

Figure 4.10 presents a line chart showing a trend of monthly average SDT, monthly maximum SDT and monthly minimum SDT dating from 1980 to 2008, based on the fix-point in-situ measurement data. During this time period, Lake Simcoe has an overall

monthly average SDT of 4.05m, a maximum monthly average SDT of 8.79m, in May 2005 and a minimum monthly average of 0.4m, in April 1984.

Cook Bay, which is located at the southernmost area of Lake Simcoe and about 60 km north of Toronto, is suffering from the most severe environmental problem among the Lake Simcoe areas. The contribution of TP is relatively high (2.1kg P/ha yr) in drainage water pumped from the intensively cultivated marsh (Nicholls and Maccrimmon, 1975). From the in-situ measured data, it is easy to find that the H3, H4 and H6 stations, which are the only three observation stations located within the Cook Bay area, always have a very low SDT and quite high chl-*a* concentration. It is normal for this bay to have the SDT less than 1 m, corresponding with a chl-*a* concentration higher than 80μg/L or even 140μg/L. According to Carlson's lake water quality criteria (Carlson and Simpson, 1996), the Cook Bay area should be classified as the eutrophic lake area, or hypereutrophic lake area. So the monthly minimum value of green line in Figure 4.10 can give an explicit explanation of the monthly SDT of the Cook Bay area, because it is also the most turbid part in the Lake Simcoe area. It can also be concluded that although the water quality of Lake Simcoe has been improved a lot since the 1990s, the water quality in Cook Bay has been continuously bad with little change.

According to the trend of the monthly averaged SDT, on an overall scale, there are three obvious periods of change: (1) the lake water clarity kept relatively stable from 1987 until fall 1992, (2) followed by a rising up until 2000, (3) lastly stayed consistent from that time to the summer of 2008. As the three ground stations (H3, H4 and H6) are historical stations, they joined the measurement since 1993 could no longer be used until 2000. It is possibly one explanation of the third stage, because these three areas were always turbid which would dramatically lower the average SDT of the whole area, considering there are only 15 fixed stations. Secondly, regarding to the maximum average SDT trend, the value also increased, which means the removal of three stations is not the only reason of stage 3, the strategies executed by the agencies also contributed to this clarity upgrading.

On the other hand, from an annual scale, it is not difficult to be aware that the water clarity has a similar trend every year. During each of the observation years, the SDT is minimized in the late summer period, normally in August or September time; the SDT is relatively higher before and after these two months, which means it will decrease to the lowest point and increase after then, during each of these 22 years.

This SDT trend of Lake Simcoe can also be performed by the concentration map. In Figure 4.11, 19 SDT maps, calculated by the regression equations from the TM images are compared to exhibit the change of water clarity over these years. The results of calculation are classified into 5 colors by using the density slice colour table. Dark red, light blue, cyan,

dark blue and black mean the SDT between 0-2m, 2-4m, 4-8m, 8-12m and 12-18m, respectively. The SDT trend presented in this way is coincident with the line chart. For instance, the averaged SDT in April 1987 was quite low on the line chart, so most of the water on Figure 4.11 (a) is classified as dark red. Figure 4.11 (b)-(h) show that there was a continuous increasing of water clarity from August 1990 to July 2000. There was a slight water clarity decreasing since August 2001, while it recovered soon in 2005. The consistency between the line chart and estimated water clarity maps indicates that the TM images can be a reliable source of estimated SDT for Lake Simcoe.

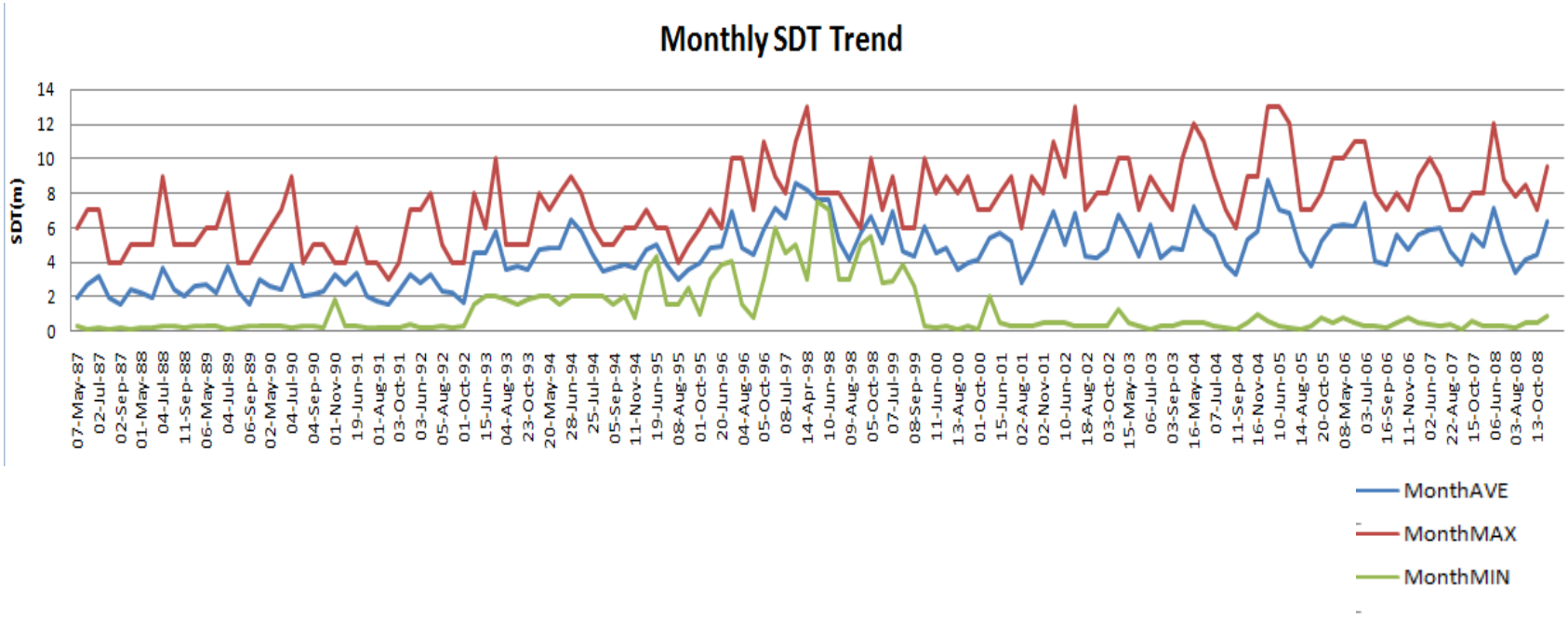


Figure 4.10 Monthly SDT trend from 1987 to 2008

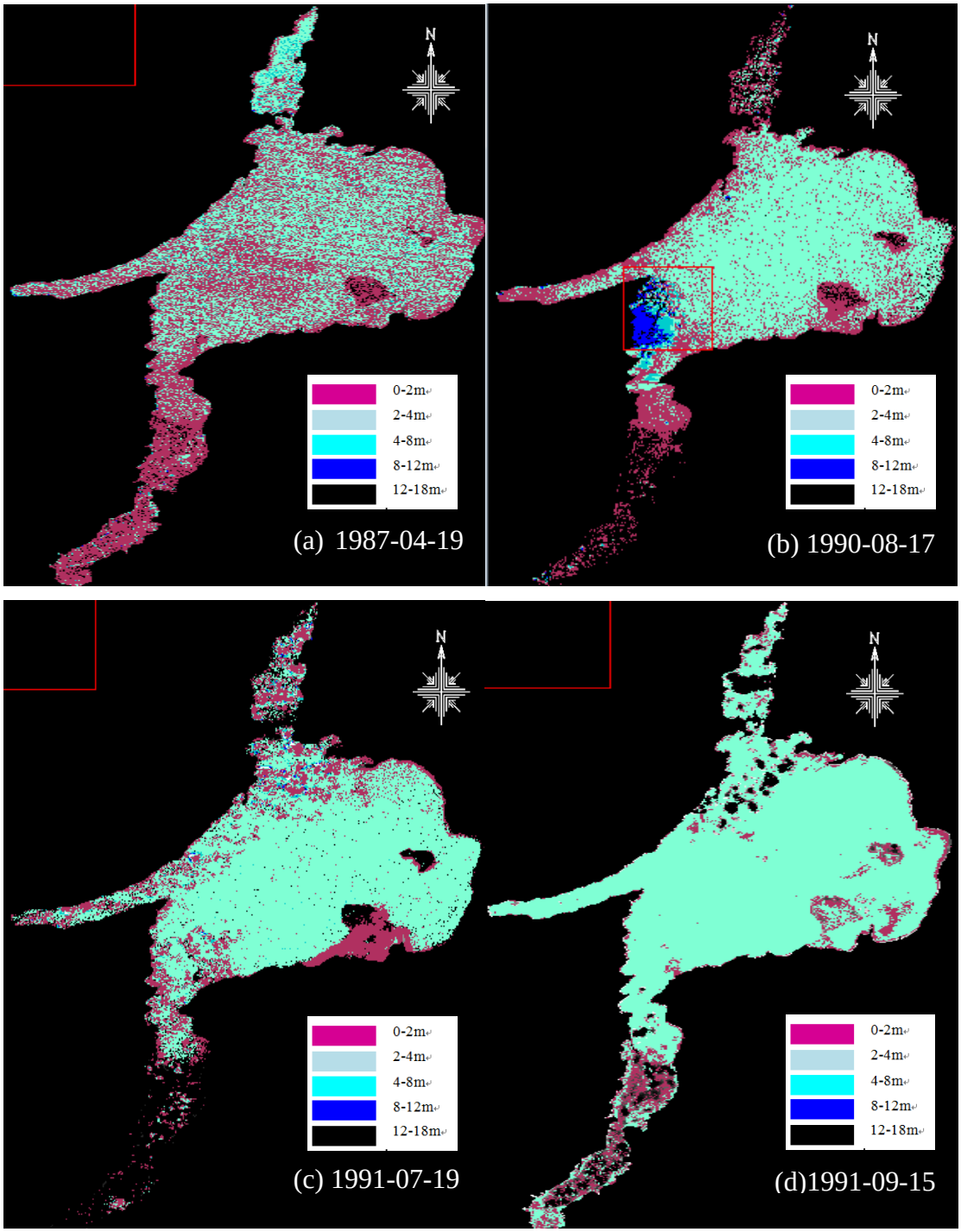


Figure 4.11 Estimated Water Clarity Maps of Lake Simcoe

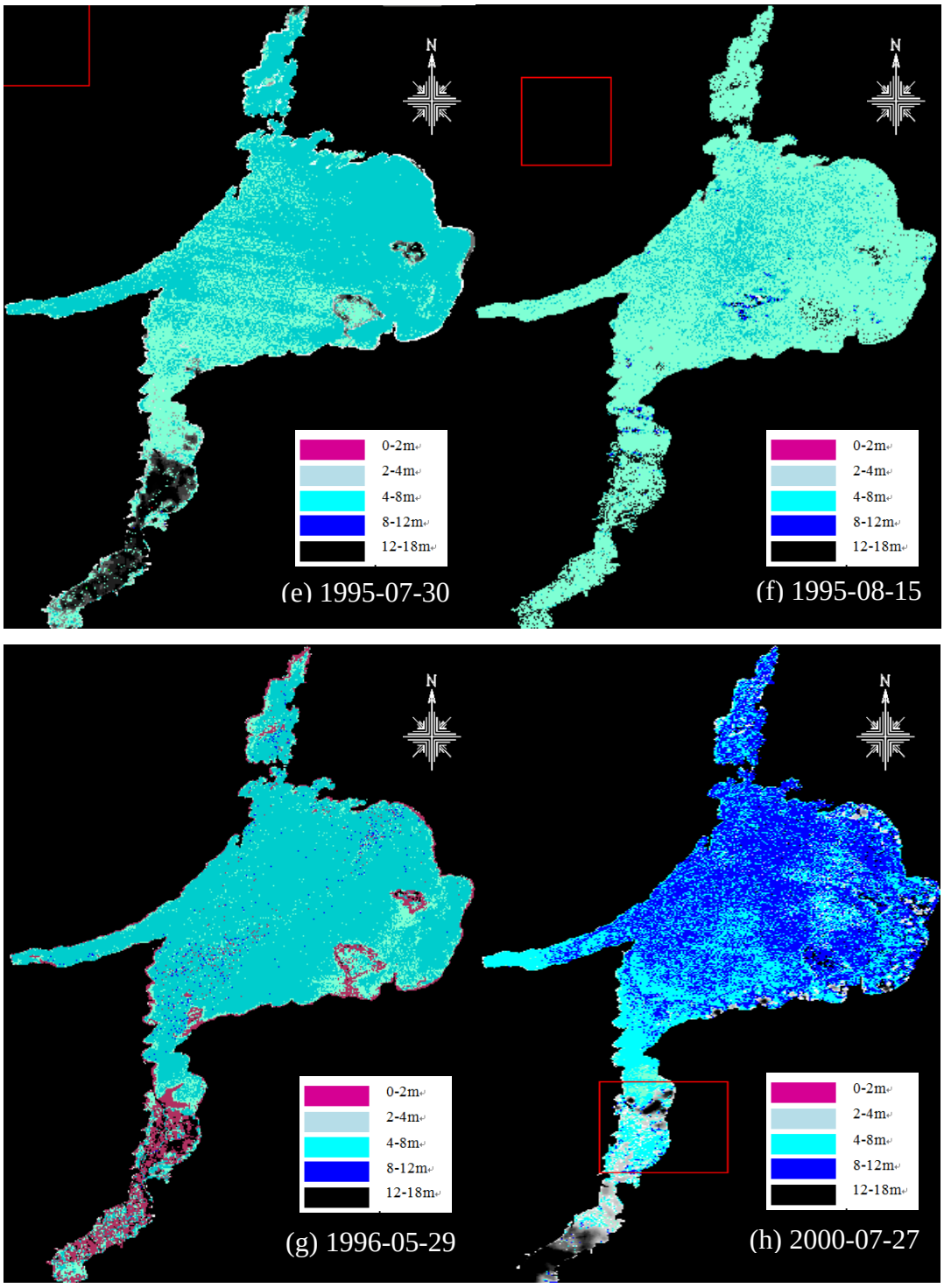


Figure 4.11(Cont.) Estimated Water Clarity Maps of Lake Simcoe

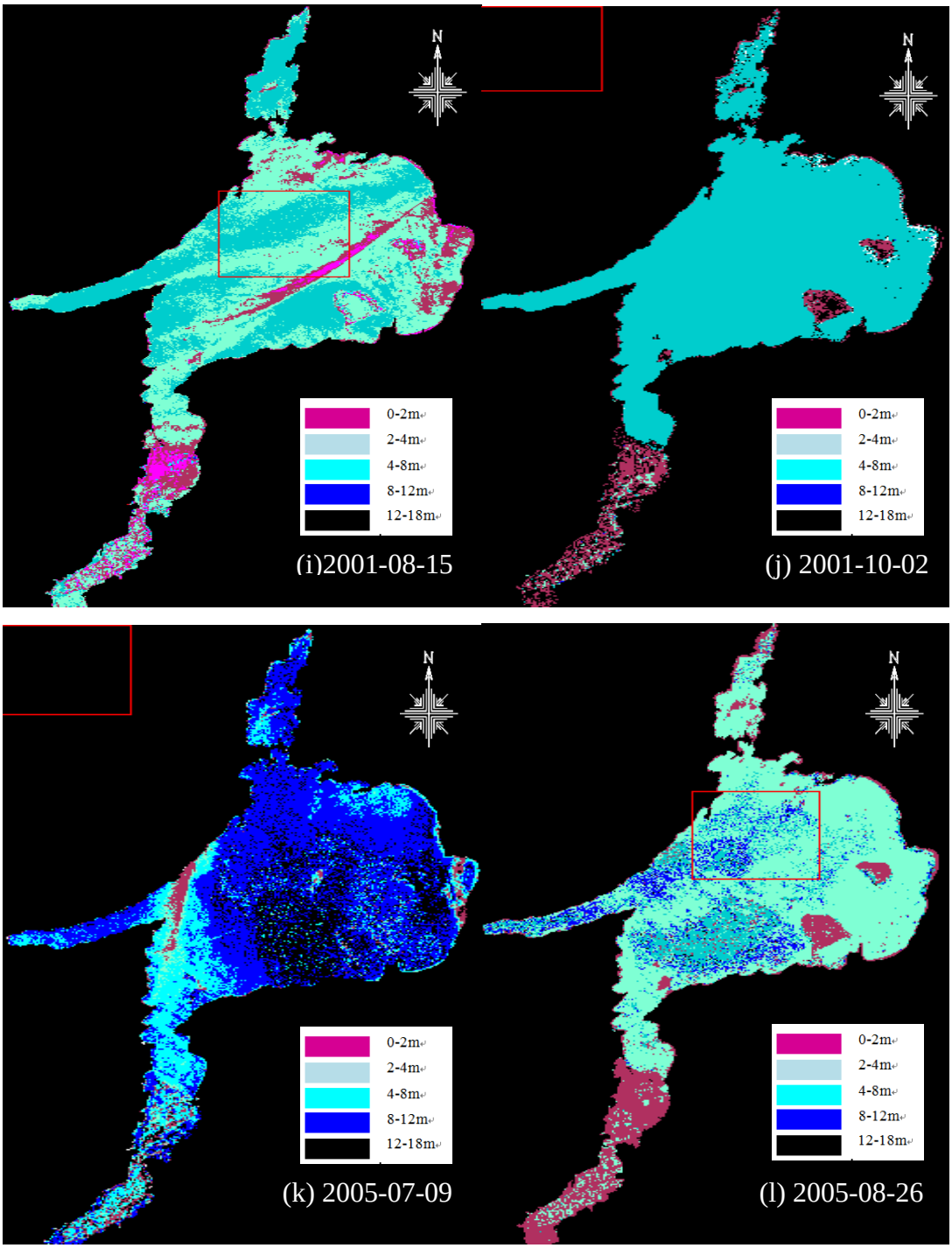


Figure 4.11(Cont.) Estimated Water Clarity Maps of Lake Simcoe

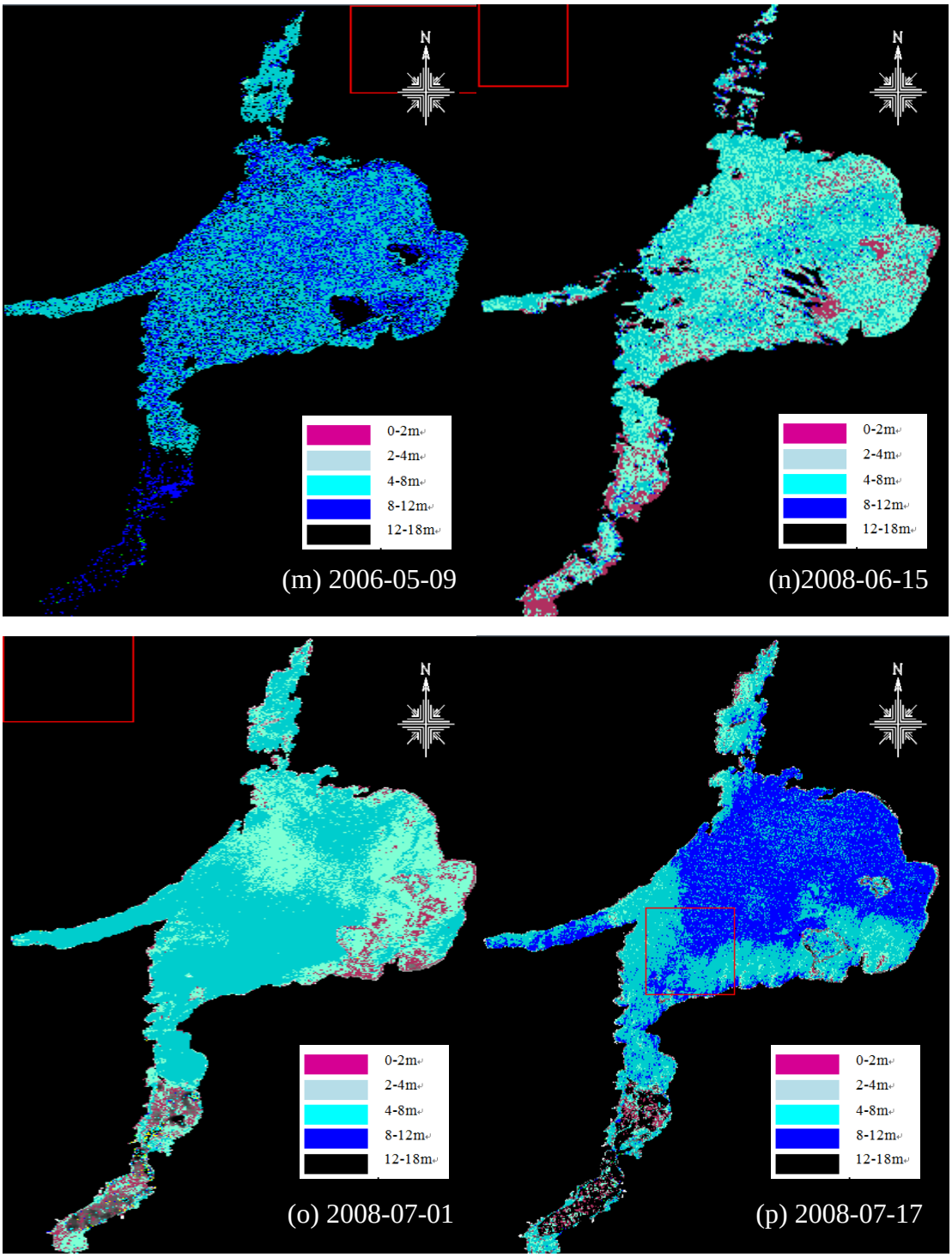
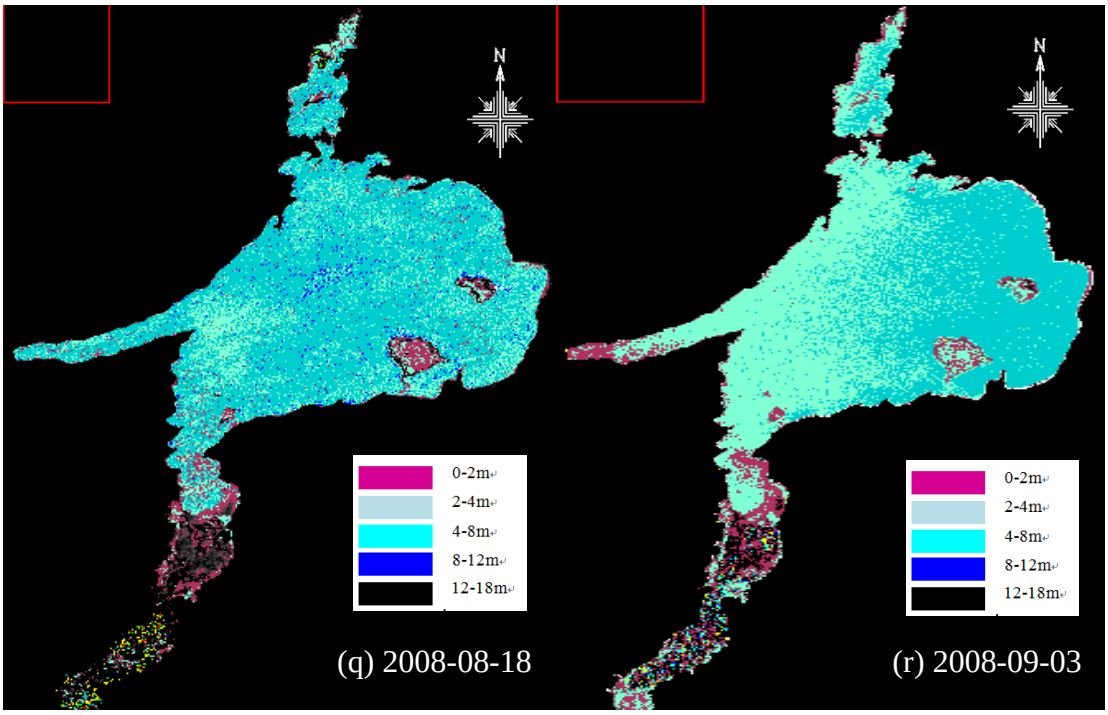


Figure 4.11(Cont.) Estimated Water Clarity Maps of Lake Simcoe



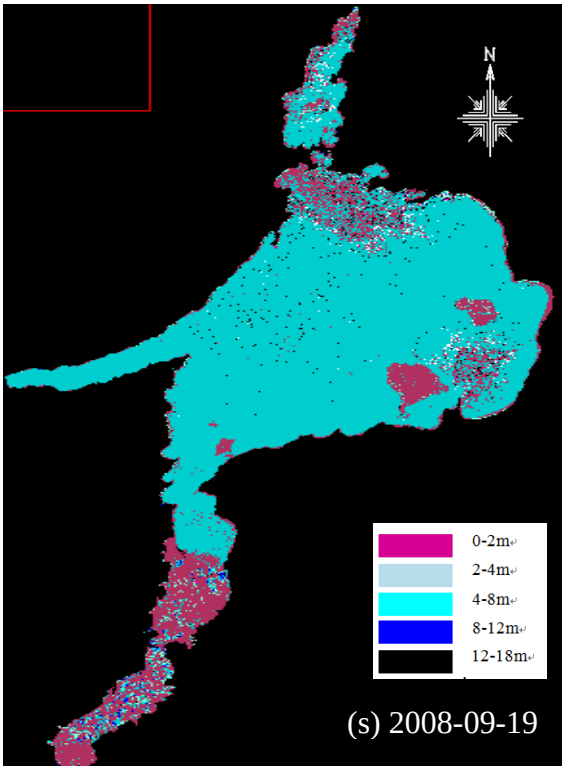


Figure 4.11(Cont.) Estimated Water Clarity Maps of Lake Simcoe

4.3.3 Trend of Chl-a Concentration

Figure 4.12 shows a trend of monthly average, monthly maximum and monthly minimum chl-*a* concentrations of Lake Simcoe, dating from 1987 to 2008, based on the fixed-point in situ measurement data. The blue line presents the monthly average chl-*a* concentration, while the red and green lines show the monthly maximum and minimum chl-*a* concentration, respectively. During this time period, Lake Simcoe has an overall monthly average chl-*a* concentration of 10.813µg/L, a maximum monthly average chl-*a* concentration of 56.4 µg/L, on August 20, 1997 and a minimum monthly average of 0.7 µg/L, on July 13, 2002. During this time period, Lake Simcoe has a monthly maximum chl-*a* concentration of 233µg/L, on September 4, 1991 and a monthly minimum concentration of 0.1 µg/L on June 3, 1992.

Because the chl-*a* concentration sampled from H3, H4 and H6 stations always has a big difference from the other stations, we need to consider the availability of these three stations during each of these years. According to the record of the in-situ data, these three stations were in use during 1982 to 1993, 1996 to 1998, and 2001 to 2002 periods, respectively. It is the possible reason why the chl-*a* concentration during those time periods are higher than the periods around them. The blue line chart shows that in these time period, the monthly average chl-*a* concentration is higher than the other years without those stations. However, it is also obvious that the chl-*a* concentration in this lake has an overall

trend of decreasing, because since 1993, not only the monthly average chl-*a* concentration decreased, the monthly maximum also decreased a lot, even though the H3,H4 and H6 stations were added in during 1995 to 1997, and 2001 to 2001.

The concentration of chl-a shows a similar trend as the SDT during a yearly period. The chl-*a* concentration has a maximum peak in August or September of each year, and then decreases until November. Before this period of every year, the concentration of chl-*a* keeps lowering until the late summer. It is also evident that the water quality of Cook' Bay has a very serious problem during all these years as its chl-*a* concentration usually is above 100µg/L and can even be higher than 200µg/L sometimes. So according to Carlson's criteria (1996), the water quality of this place can be considered as hypereutrophic most of the time in a year. Even though some programs were developed to improve the water quality after the 1990s, the concentration of chl-*a* in this bay was still above 60µg/L, which means hypereutrophic again. The trends of both SDT and chl-*a* imply that although the water quality of the entire Lake Simcoe has improved, the water quality in the southern area, Cook's Bay, is still hypereutrophic, and more attention needs to be paid to this area.

Figure 4.13 shows two estimated chl-*a* maps from bands combination models using colours in blue, green, red and yellow. Two imageries in July of different years are employed here to demonstrate the change of chl-*a* concentration in Lake Simcoe. The darker the color is, the lower the chl-*a* concentration the lake has. On the contrary, yellow means the chl-*a*

concentration is very high in water. The comparison of those two estimated maps shows that Lake Simcoe has an extremely high concentration of chl-*a* in the summer of 1991, however this situation has been changed a lot until 1995 summer. This change of chl-*a* concentration is coincident with the line chart generated from the in-situ monthly averaged chl-*a* concentration.

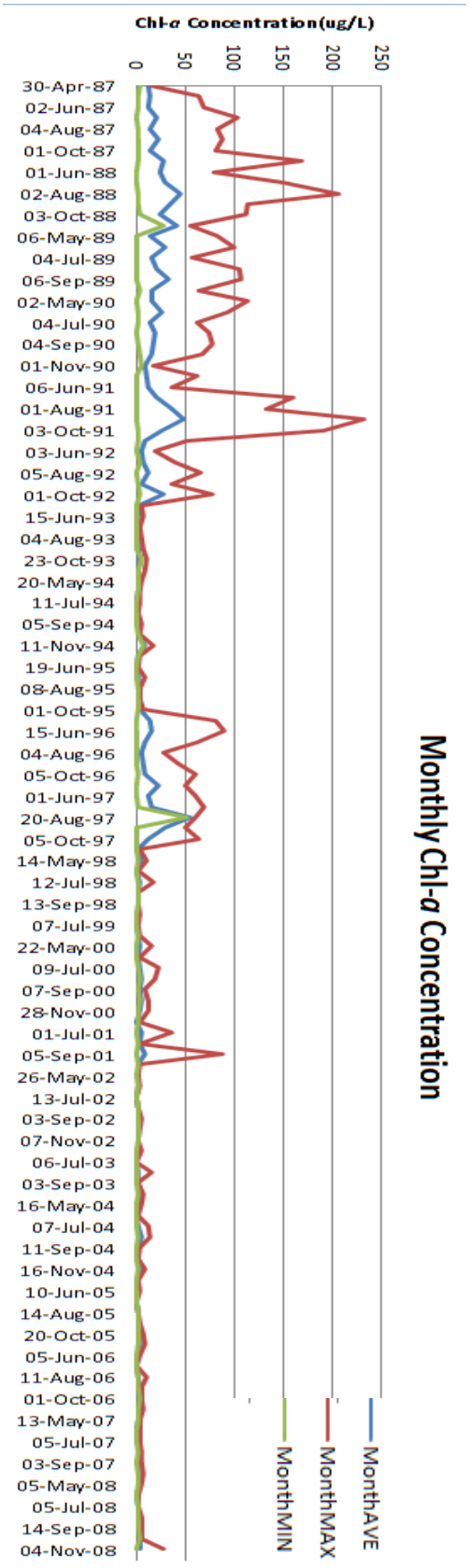


Figure 4.12 the Monthly Trends of Chl-*a* Concentration from 1987 to 2008

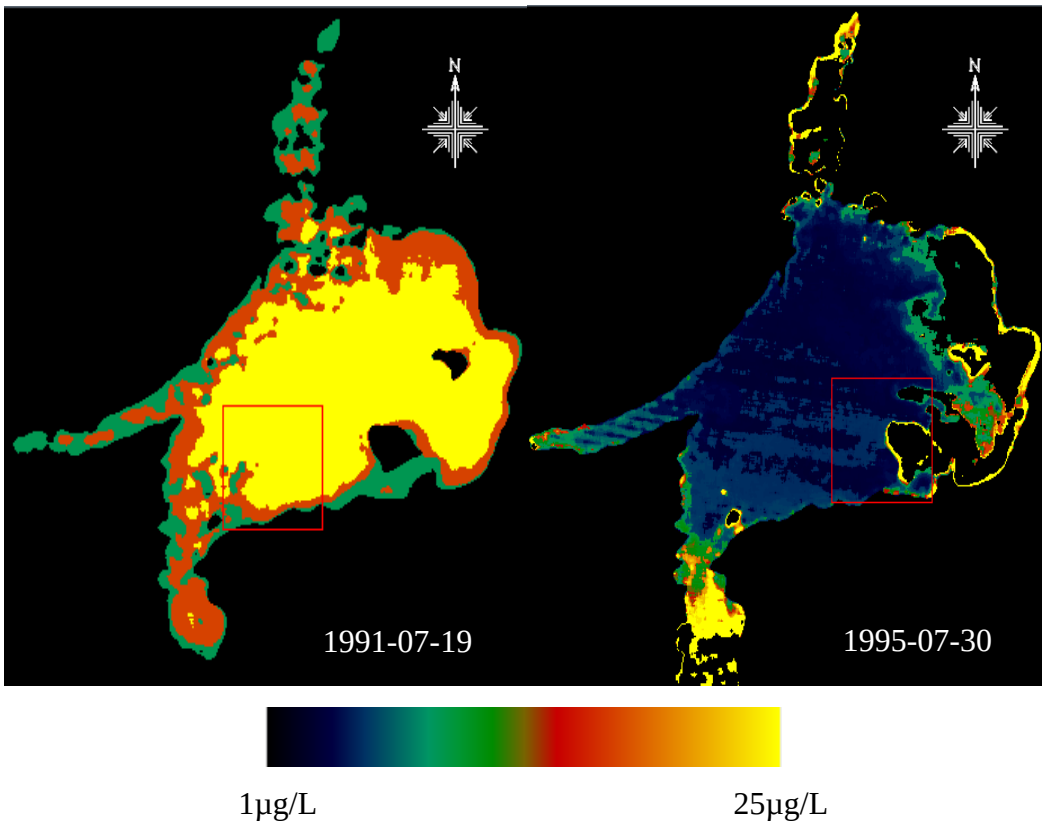


Figure 4.13 A Comparison of the Estimated Chl-*a* Concentration

4.4 Water Quality Classification

4.4.1 Trophic States Criteria

A lake is usually classified as being in one of three possible classes: oligotrophic, mesotrophic and eutrophic. Lakes with extreme trophic indices may also be considered hyperoligotrophic or hypereutrophic. In Carlson and Simpson's study (1996), they also summarized a table demonstrating how the TSI values translate into trophic classes and Table 4.3 shows a classification of different parameters including Trophic Index, Chlorophyll, Phosphorus, Secchi depth, and Trophic Class (Carlson and Simpton, 1996). In

this table, oligotrphic lakes generally host very little or no aquatic vegetation and are relatively clear, while eutrophic lakes tend to have large quantities of organisms, including algal blooms. If the algal biomass in a lake or other water body reaches too high a concentration, massive fish die-offs may occur as decomposing biomass deoxygenates the water.

Table 4.3 Classification of Lake Water Trophic State Index (Carlson and Simpson, 1996)

TI	CHL	P	SD	Trophic Class
<30-40	0-2.6	0-12	>8-4	Oligotrophic
40-50	2.6-7.3	12-24	4-2	Mesotrophic
50-70	7.3-56	24-96	2-0.5	Eutrophic
70-100+	56-155+	96-384+	0.5-<0.25	Hypereutrophic

4.4.2 Water Quality Classification of Lake Simcoe

As the eutrophication effect is the main problem of Lake Simcoe, the chl-*a* concentration is considered to be the most important parameter here to indicate the water quality of this lake. The Carlson's modified tropic state index TSI_m (see Table 4.3) is applied in this study, as it only uses the chl-*a* concentration to represent the trophic state index of inland lakes (Aizaki et al., 1981; Wang et al., 2008). In this study, the lake water is classified into four categories: Oligotrophic, lower-mesotrophic, higher-mesotrophic and hypereutrophic. The modified trophic state index is calculated by

$$TSI_M\,(\text{Chl-a}) = 10*(2.46 + (\text{Log}_{10}\,(\text{Chl-a})/\text{Log}_{10}2.5)) \qquad (14)$$

The estimated chl-*a* concentration maps are calculated using the equations above and the result is displayed using density slice to show the different classes. When the resulting value is between 0 and 38, it would be classified as oligotrophic; for the values between 38 and 45, it would be considered as lower-mesotrophic lake water; for the values between 45 and 53, it should be classified as higher-mesotrophic water; for the values higher than 53, it is called hypereutrophic lake waters.

Figure 4.14 shows the comparison of Lake Simcoe's TSI_M of 1995 and 2008, and the oligotrophic, lower-mesotrophic, higher-mesotrophic and hypereutrophic waters are represented by blue, green, cyan and red, respectively. In the left figure which represents the TSI_M in 1995, most of the lake water is classified as the hypereutrophic category, and only the north part and east part are shown in green which is lower-mesotrophic. By contrast in the right figure of 2008, apparently the water quality has improved a lot, as the hypereutrophic areas are now in blue, oligotrophic.

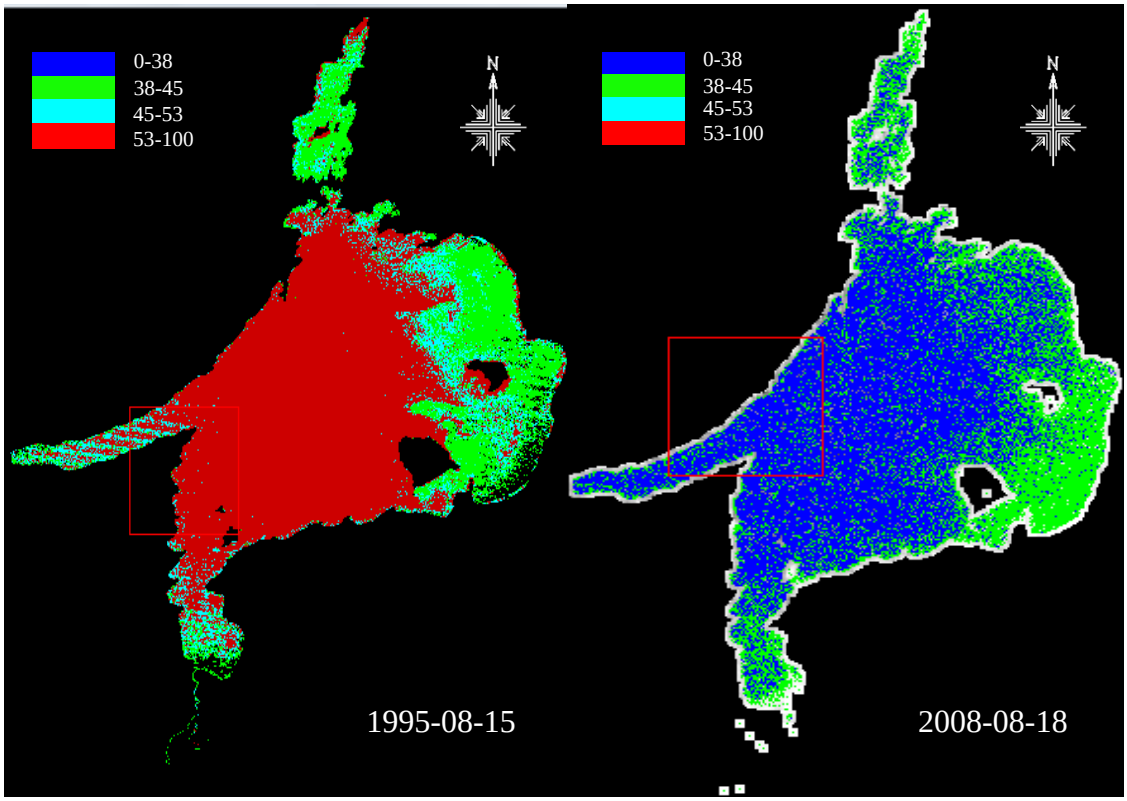


Figure 4.14 TSI comparison of 1995 and 2008 for Lake Simcoe

4.5 The Influence of Dissolved Organic Carbon Concentration

It might be noticed that the estimated SDT and chl-*a* concentration of certain dates do not have a good correlation relationship with the in-situ data. Besides the errors caused by the sensor and the atmosphere, there are two reasons that may contribute to those results with relatively low correlation.

The first reason is that most of the observation stations are located very close the shore of Lake Simcoe. This means that we cannot guarantee it is the water quality problem that

contributes to the turbidity or the water itself is not deep enough. Also, the near-shore lake water is much more complex than the water far away from the shore, as there might be some other components in the water such as the garbage left by the residents. The eroded soil will make the near-shore water turbid as well. On the other hand, there are usually a confusion of the pixels in satellite imagery between the water and soil if the spatial resolution of the satellite images is not high enough. Sometimes even if the stations are located perfectly on the imagery, it still may turn out that the DN values of soil rather than water are extracted. All of these reasons may more or less affect the results of estimation.

Another reason is the effect of the DOC concentration in the lake water. This parameter influences not only the water quality itself, but also the absorption of blue light, which means that TM1 can not accurately represent the real reflectance of the water body. However, both the SDT and chl-*a* concentration estimation need the blue light band to be considered and calculated. Figure 4.15 shows the overall trend of the monthly average DOC from 1987 to 2008. It is obvious to find that the DOC concentration stays quite stable around 4mg/L until 1995, and it starts to increase gradually since 1996 and becomes unstable. There are two concentration peaks which are the 1997 summer and 1998 summer, when the DOC concentration has achieved 8mg/L and almost 10mg/L respectively. The average concentrations are about 2mg/L higher during the period after 1996 than before 1996. Correspondingly, the correlation coefficients before 1996 are relatively high, which

means the estimated water quality parameters are well-correlated to the ground measured water quality parameters; on the contrary, not all the correlation coefficients after 1996 are high enough to prove the relationship between the estimated data and in-situ data, such as the estimated chl-*a* of July 2007, August 2005, May 2006 and June 2008, they only have the highest R^2 of 0.505, 0.310, 0.219 and 0.324 respectively. Although there is no statistical or quantitative evidence in this study to prove that the reasons are exactly the increase of DOC concentration, based on the literature, as well as the overlap of decreased correlation and increased DOC concentration, it is believed here, to some extent, that the rising DOC concentration is one reason for these poor relationships.

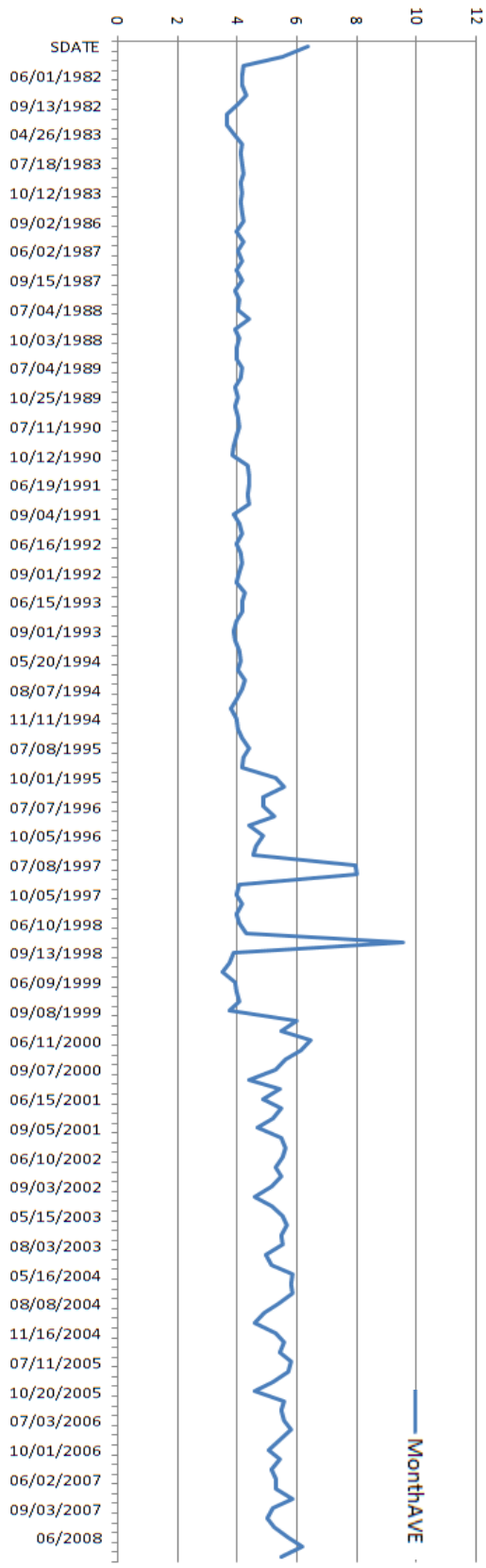


Figure 4.15 Trend of Monthly Average DOC Concentration from 1987 to 2008

4.6 Chapter Summary

This chapter presents and discusses the results of this study. The linear regression shows that the vast majority of image-estimated SDT are closely correlated with the in-situ ones. Meanwhile, the improved model exhibits an advantage over the general one in this study. On the other hand, the prediction for chl-*a* concentration only reveals fair agreements with the in-situ data. Moreover, none of the nine models show obvious superiorities. The reasons for this are possibly due to the absorption of light caused by chl-*a* and DOC in the water, as well as the location of stations. Both the SDT and chl-*a* reveal noticeable seasonal variations, and the estimations during the late-summer period have higher accuracies.

The results of the estimated maps indicate that the spatial distribution of eutrophic area is normally concentrated at the near-shore areas and the northeast part of Lake Simcoe. Meanwhile, the water quality of the southwest part of Lake Simcoe is much better. Also, the southern Cook's Bay, is suffering from an extremely serious water quality problem.

The trends analysis indicates that the water quality of Lake Simcoe is worst in August and September while it is much better during the other months annually. According to the overall trend of the monthly averaged SDT, the SDT kept relatively stable until the fall of 1992, followed by a rising up until 2000, and lastly stayed consistent from that time to the

summer of 2008. The chl-*a* concentration reveals an inverse trend to SDT, which means the higher the concentration of chl-*a* is, the more turbid the water is.

Chapter 5

Conclusions and Recommendations

5.1 Conclusions

This study focused on estimating the water quality change of Lake Simcoe from 1987 to 2008, via the approach of comparing Landsat 5 TM images and the in-situ data, including SDT and chl-*a* concentration. Specifically, this research was able to:

1) Build the empirical relationships between the TM images and the in-situ water quality data, so as to estimate SDT and chl-*a* concentration of Lake Simcoe during a period of 22 years, from 1987 to 2008.
2) Identify the turbid and eutrophic regions based on the concentration maps calculated from the regression equations.
3) Simulate the water quality trends of the past 22 years by the ways of both the line chart and the concentration maps.
4) Classify the lake water into four eutrophic levels based on the inland lake classification criteria.

The estimated results from TM images reveal that the predicted SDT has a close agreement with the in-situ data. A total of 77.3% comparisons have good correlation coefficients. Moreover, TM3 is found to be a better predictor than TM1 for the SDT estimation. 68.2%

of the comparisons show higher predictive accuracies by utilizing the improved model, composed of TM3 and the TM1/TM3 ratio. On the other hand, after comparing 9 chl-*a* prediction models, only 59.1% estimations for chl-*a* concentration have acceptable agreements with the in-situ measurements. The possible reasons for the low correlation are the chl-*a* itself, as well as the DOC in Lake Simcoe, which can strongly absorb the blue and green light and therefore weaken the signals back to the satellites.

It is indicated that both the SDT and chl-*a* concentration estimations have seasonal variations, i.e., the late-summer comparisons frequently lead to high agreements. The concentration maps show that the eutrophic area is normally centralized at the near-shore areas and the northeast part of Lake Simcoe. The overall trend of SDT begins with decrease values during 1980 to 1982, followed by a relatively stable stage until the fall of 1992, then rises up till 2000, and lastly stays consistent to the summer of 2008. The chl-*a* has an inverse trend with the SDT.

5.2 Suggestions for Future Research

As the in-situ stations on Lake Simcoe have been continuously reduced since 1995, the estimated water quality parameters can play a role to supplement the traditional measurements. Nevertheless, some limitations still exist in this study and they can possibly be improved in the future.

The first limitation of this study is the seasonal variability. As the late-summer period is the best season to estimate the lake water quality, more accurate results would be given during this period. However, almost 50% of the data in this study are collected in spring and fall season. As consequence, some of these images have poor agreements with the in-situ data. The reasons for this unsatisfied data collection include: (1) the 16-day temporal resolution of Landsat, (2) the demand for cloud-free images during the ice-free period, (3) the coincidence with the in-situ data, and (4) the limited free data availability on the USGS website. Consequently, the odds of qualified data are actually low. Not surprisingly, only 22 imageries are acquired during the whole 22-year period. In order to stimulate the long-period trend, the unification of seasons has to be sacrificed. In spite of this, the seasonal variation can be overcome by combining MODIS and Landsat images as the datasets. The high temporal resolution of MODIS images can guarantee an adequate late-summer data selection, while the fair spatial resolution of Landsat images can exhibit the spatial patterns of water quality parameters.

A second point is that the reasons for the spatial patterns of water quality parameters have not been explained. As the MOE only collects the water quality data of Lake Simcoe, insufficient land-use data have been obtained to conduct quantitative research. Therefore, the causes for those spatial patterns are still unknown. These causes can be analyzed and therefore interpreted once the detailed land-use data become available.

Moreover, the spatial resolution of TM images is not high enough for some sampling stations, such as the stations H_3, H_4, and H_6, located at the Cook's Bay, where the shape of the lake becomes quite narrow. Therefore, this region can easily get confused with the agriculture lands surrounding by it on the images. As a result, when estimated by the regression equations, the pixels within this region cannot be calculated correctly. Since the Cook's Bay is a key area that is worth monitoring, owing to its consistently poor water quality, images with higher spatial resolution are required in the future studies.

A fourth point includes the errors during both the data sampling and the processing procedures. Firstly, it is not guaranteed that the water is sampled from the same depth every time. The water in different depths can result in various concentrations of chl-*a*. Therefore, the water quality data from the same depth should be compared to the reflectance values if the depth information is available. On the other hand, there are various methods to remove the atmospheric effects from the Landsat-5 TM images. However, in this study, only one method in PCI Geomatica is utilized to correct this effect. Thus, it is valuable to use or compare other softwares and methodologies for the atmospheric correction in the future.

Last but not least, the improved equation for SDT estimation needs to be applied to other study areas. Although TM3 and the TM1/TM3 ratio have impressive estimation accuracies in this study, the universality of this prediction model is still doubtable. One feasible approach is to employ this method to other research sites of the existing studies using the

general SDT prediction equation, and to compare the accuracies with those studies afterwards. On the other hand, the exploration for a general prediction equation of chl-*a* concentration should not be given up either.

To conclude, this research can be improved by acquiring satellite data with higher spatial and temporal resolutions, obtaining quantitative land-use data of the Lake Simcoe region, and testing the improved SDT equation in other study areas.

References

Aizaki, M., T. Iwakuma and N. Takamura, 1981. Application of modified Carlson's trophic state index to Japanese lakes and its relationship to other parameters related to trophic state, *Research Report on National Institute of Environmental Studies,* 23: 13-31.

Albright, M.F., 1996. Hydrological and nutrient budget for otsego Lake, N.Y. and relationships between land form/use and export rates of its sub-basins. Occas. Pap. #29. SUNY Oneonta Bio. Fld. SUNY Oneonta.

Allee, R.J., and J.E. Johnson, 1999. Use of satellite imagery to estimate surface chlorophyll a and Secchi disc depth of Bull Shoals Reservoir, Arkansas, USA, *International Journal of Remote Sensing,* 20(6): 1057-1072.

Alsdorf, D., D. Lettenmaier, C. Vorosmarty and the NASA Surface Water Working Group, 2003. The need for global, satellite-based observations of terrestrial surface waters, *EOS*, 84(29): 269-280.

Amtstaetter, F., and C.C. Willox, 2004. Survival and growth of lake whitefish from two strategies in Lake Simcoe, Ontario, *North American Journal of Fisheries Management,* 24(4): 1214-1220.

Binding, C.E., J.H. Jerome, R.P. Bukata, and W.G. Booty, 2007. Trends in water clarity of the lower Great Lakes from remotely sensed aquatic color, *Journal of Great Lakes Research,* 33(4):828-841.

Brando, V.E., and A.G. Dekker, 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality, *IEEE Transactions on Geoscience and Remote Sensing*, 41(6): 1378-1387.

Brivio, P.A., C. Giardino, and E. Zilioli, 2001. Validation of satellite data for quality assurance in lake monitoring applications, *The Science of Total Environment*, 268(1-3): 3-18.

Brown, D., R. Warwick, and R. Skaggs, 1977. Reconnaissance analysis of lake condition in east– central Minnesota. Report No. 5022. Minnesota Land Management Information System, Center for Urban and Regional Affairs, University of Minnesota, Minneapolis, MN, 19 pp.

Bukata, R.P., 2005. *Satellite Monitoring of Inland and Coastal Water Quality*, CRC Press, Boca Raton, FL, 246p.

Canfield, D.E., Jr., C.D. Brown, R.W. Bachmann and M.V. Hoyer, 2002. Volunteer lake monitoring: testing the reliability of data collected by the Florida LAKEWATCH program, *Journal of Lake and Reservoir Management*, 18(1): 1-9.

Carlson, R.E., 1977. A trophic state index for lakes, *Limnology and Oceanography*, 22(2):361-369.

Carlson, R.E. and J. Simpson, 1996. A Coordinator's Guide to Volunteer Lake Monitoring Methods. North American Lake Management Society. 96 pp. URL: http://en.wikipedia.org/wiki/Trophic_state_index (Last date accessed: July 30, 2008)

Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications,* 8:559–568.

Chambers, P.A., M. Guy, E.S. Roberts, M.N. Charlton, R. Kent, C. Gagnon, G. Grove and N. Foster, 2001. *Nutrients and their impact on the Canadian environment,* Agriculture and Agri-Food Canada, Environment Canada, Fisheries and Oceans Canada, Health Canada and Natural Resource Canada. 241 p.

Chen, C.Q., P. Shi, A and Q.P. Mao, 1996. Study on modeling chlorophyll concentration of surface coastal water using TM data, *J. Remote Sensing*, 11(3): 168-175.

Chen, Q., and Y. Zhang, 2007. Water quality monitoring using remote sensing in support of the EU water framework directive (WFD): A case study in the Gulf of Finland, *Environment Monitoring and Assessment*, 124(1-3): 157-166.

Chen, Z., C. Hu, and F. Muller-Karger, 2007a. Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery, *Remote Sensing of Environment*, 109(2): 207-220.

Chen, Z., F.E. Muller-Karger, and C. Hu, 2007b. Remote sensing of water clarity in Tampa Bay, *Remote Sensing of Environment*, 109(2):249-259.

Cox, R. M., R. D. Forsythe, G. E. Vaughan, and L. L. Olmsted, 1998. Assessing water quality in the Catawba River reservoirs using Landsat Thematic Mapper satellite data, *Lake and Reservoir Management,* 14: 405– 416.

Dall'Olmo, G., A.A. Gitelson, D.C. Rundquist, B. Leavitt, T. Barrow, and J.C. Holz, 2005. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands, *Remote Sensing of Environment,* 96(2): 176-187.

Dekker, A.G., R.J. Vos and S.W.M. Peters, 2001. Comparison of remote sensing data, model results and in situ data for total suspended matter (TSM) in southern Frisian lakes, *The Science of the Total Environment,* 268(1-3): 197-214.

Dekker, A.G. and S.W.M. Peters, 1993. The use of the Thematic Mapper for the analysis of eutrophic lakes: a case study in the Netherlands, *International Journal of Remote Sensing,* 14(5):799-821.

Dekker, A.G., R.J. Vos and S.W.M. Peters, 2002. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data, *International Journal of Remote Sensing,* 23(1): 15-35.

Dewider, K. and A. Khedr, 2001. Water quality assessment with simultaneous Landsat-5 TM at Manzala Lagoon, Egypt, *Hydrobiologia,* 457(1-3):49-58.

Dwivedi, R. M. and A. Narain, 1987. Remote sensing of phytoplankton, an attempt from the Landsat Thematic Mapper, *International Journal of Remote Sensing,* 8(10):1563-1569.

Eimers, M.C., and J.G. Winter, 2005. *Lake Simcoe Water Quality Update 2000-2003.* LSEMS Implementations Tech. Rep. No. Imp. B.20. Lake Simcoe Region Conservation Authority, Newmarket, ON.

Environment Canada. Inland Waters Branch. 1973. *Inventory of Freshwater Lakes*. Ottawa.

Environment Canada, 2009. Cleaning up Lake Simcoe: Part of the Government of Canada's Action Plan for Clean Water, URL: http://www.ec.gc.ca/paae-apcw/Default.asp?lang=En&n=63494C3C-1 (Last date accessed: 10 June 2009)

Environment News Service (ENS), 2008. Canada opens CAN$30m fund to clean popular vacation lake, URL: http://www.environmental-expert.com/resultEachPressRelease.aspx?cid=4797&codi=34211 (Last date accessed: 12 June 2009).

Evans, D.O., K.H. Nicholls, Y.C. Allen, and M.J. Mcmurty, 1996. Historical land use, phosphorus loading, and loss of fish habitat in Lake Simcoe, Canada, *Canadian Journal of Fisheries and Aquatic Sciences*, 52(Supplement 1):194-218.

Forget, P., P. Broche, and J. Naudin, 2001. Reflectance sensitivity to solid suspended sediment stratification in coastal water and inversion: a case study, *Remote Sensing of Environment*, 77(1): 92-103.

Fraser, R.N., 1998. Multispectral remote sensing of turbidity among Nebraska Sand Hills Lakes, *International Journal of Remote Sensing*, 19(15): 3011-3016.

Fuller, L.M., S.S. Aichele, and R.J. Minnerick, 2004, Predicting water quality by relating secchi-disk transparency and chlorophyll a measurements to satellite imagery for Michigan inland lakes, August 2002: U.S. Geological Survey Scientific Investigations Report 2004-5086, 25 p.

Gan, T.Y., O.A. Kalinga, K. Ohgushi, and H. Araki, 2004. Retrieving seawater turbidity from Landsat TM data by regressions and an artificial neural network, *International Journal of Remote Sensing,* 25(21):4593-4615.

Giardino, C., V.E. Brando, A.G. Dekker, N. Strmbeck and G. Candiani, 2007. Assessment of water quality in Lake Garda (Italy) using Hyperion, *Remote Sensing of Environment,* 109(2): 183-195.

Harma, P., J. Vepsalainen, T. Hannonen, T. Pyhalahti, J. Kamari, K. Kallio, K. Eloheimo and S. Koponen, 2001. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland, *The Science of the Total Environment,* 268: 107-121.

Harrington, J.A., F.N. Joe, Jr. and F.R. Schiebe, 1992. Remote sensing of Lake Chicot, Arkansas: Monitoring suspended sediments, turbidity, and Secchi depth with Landsat MSS data, *Remote Sensing of Environment,* 39(1): 15-27.

Heiskary, S., J. Lindbloom and C.B. Wilson, 1994. Detecting water quality trends with citizen volunteer data, *Journal of Lake and Reservoir Management,* 9(1): 4-9.

Hellweger, F.L., P. Schlosser, U. Lall, and J.K. Weissel, 2004. Use of satellite imagery for water quality studies in New York Harbor, *Estuarine, Coastal and Shelf Science*, 61(3):437-448.

Henderson-Sellers, B. and H.R. Markland, 1987. *Decaying Lakes* (Chichester: John Wiley).

Hirthle., H. and A. Rencz, 2003. The relation between spectral reflectance and dissolved organic carbon in lake water: Kejimkujik National Park, Nova Scotia, Canada, *International Journal of Remote Sensing*, 24(5): 953-967.

Howard, R.G., 1978. Removal of atmospheric effects from satellite imagery of the oceans, *Journal of Applied Optics*, 17(10): 1631-1636.

Islam, M.M., and K. Sado, 2005. Water quality monitoring of case 2 water using field spectroradiometer and remote sensing data, *Goettinger Geographische Abhandlungen*, 113:159-166.

Johnson, M.G., and K.H. Nicholls, 1989. Temporal and spatial variability in sediment and phosphorus loads to Lake Simcoe, Ontario, *Journal of Great Lakes Research*, 15(2): 265-282.

Kallio, K., J. Attila, P. Harma, S. Koponon, J. Pulliainen, U.M. Hyytiainen, and T. Pyhalahti, 2008. Landsat ETM+ Image in the estimation of seasonal lake water quality in Boreal River Basins, *Environmental Management*, 42(3): 511-522.

Kaw, A. and E. Kalu, 2008. *Numerical methods with applications (1st ed.)*, URL: http://numericalmethods.eng.usf.edu/topics/textbook_index.html (Last date accessed: Apr 15, 2009)

Kaufman, Y.J. and C. Sendra, 1988. Algorithm for automatic atmospheric correction to visible and near-ir satellite imagery, *International Journal of Remote Sensing*, 9(8): 1357-1381.

Keiner, L.E., and X. Yan, 1998. A neural network model for estimating sea surface chlorophyll and sediments from Thematic Mapper imagery, *Remote Sensing of Environment*, 66(2): 153-165.

Kerr, M., E. Ely, V. Lee, and A. Mayio, 1994. A profile of volunteer environmental monitoring: National survey results, *Journal of Lake and Reservoir Management*, 9: 1-4.

Khorram, S., H. Cheshire, A.L. Geraci and G. La Rosa, 1989. Water quality mapping of Augusta Bay, Italy from Landsat-tm data, *International Journal of Remote Sensing*, 12(4):803-808.

Khorram, S. and H.M. Cheshire, 1985. Remote sensing of water quality in the Neuse River Estuary, North Carolina, *Photogrammetric Engineering and Remote Sensing*, 51(3): 329-341.

Kilgour, B., C. Clarkin, W. Morton, and R. Baldwin, 2008. Influence of nutrients in water and sediments on the spatial distributions of Benthos in Lake Simcoe, *Journal of Great Lakes Research,* 34(2):365-376.

Kishino, M., A. Tanaka, and J. Ishizaka, 2005. Retrieval of Chlorophyll a, suspended solids, and colored dissolved organic matter in Tokyo Bay using ASTER data, *Remote Sensing of Environment*, 99(1-2): 66-74.

Kloiber, S.M., P.L. Brezonik, L.G. Olmanson, and M.E. Bauer, 2002. A procedure for regional lake water clarity assessment using Landsat multispectral data, *Remote Sensing of Environment,* 82(1):38-47.

Kloiber, S.M., P.L. Brezonik, and M.E. Bauer, 2002. Application of Landsat imagery to regional-scale assessments of lake clarity, *Water Research,* 36(17): 4330-4340.

Kloiber, S.M., T.H. Anderle, P.L. Brezonik, L. Olmanson, M.E. Bauer and D.A. Brown, 2000. Trophic state assessment of lakes in the Twin Cities (Minnesota, USA) region by satellite imagery, *Archive Hydrobiologie Special Issues Advances in Limnology,* 55: 137-151.

Kondratyev, K.Y., D.V. Pozdnyakov, and L.H. Pettersson, 1998. Water quality remote sensing in the visible spectrum, *International Journal of Remote Sensing,* 19(5): 957-979.

Koponen, S., J. Pulliainen, K. Kallio, and M. Hallikainen, 2002. Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data, *Remote Sensing of Environment,* 79(1): 51-59.

Koponen, S., J. Pulliainen, H. Servomaa, Y. Zhang, M. Hallikainen, K. Kallio, J. Vepsalainen, T. Phhalahti, and T.Hannonen, 2001. Analysis on the feasibility of multi-source remote sensing observations for chl-*a* monitoring in Finnish Lakes, *The Science of Total Environment,* 268(1): 95-106.

Koponen, S., K. Kallio, J. Pulliainen, J. Vepsalainen, T. Pyhalahti, and M. Hallikainen, 2004. Water quality classification of Lakes Using 250-m MODIS data, *IEEE Geoscience and Remote Sensing Letters,* 1(4): 287-291.

Koponen, S., 2006. Remote sensing of water quality for Finnish lakes and coastal areas, Ph.D dissertation, Helsinki University of Technology Laboratory of Space Technology Publications, Espoo, Finland, 74p.

Lake Simcoe Environment Management Strategy (LSEMS), 2003. *State of the Lake Simcoe Watershed,* Lake Simcoe Region Conservation Authority, Newmarket, ON.

Lathrop, R.G. and T.M. Lillesand, 1986. Utility of Thematic Mapper data to assess water quality, *Photogrammetric Engineering and Remote Sensing,* 52: 671-680.

Lathrop, R.G., 1992. Landsat Thematic Mapper monitoring of turbid inland water quality, *Photogrammetric Engineering and Remote Sensing,* 58: 465-470.

LSEMS, 2006. Lake Simcoe Environmental Management Strategy. URL:

http://www.lsrca.on.ca/AboutUs/LakeSimcoeEnviroManagement.html (Last date accessed: Oct 21, 2008)

LSCF, 2006. The Lake Simcoe Conservation Foundation. URL: http://www.lsrca.on.ca/Foundation/About-Us.html (Last date accessed: Oct 21, 2008).

Lillesand, T. M., W. L. Johnson, R. L. Deuell, O. M. Lindstrom, and D. E. Meisner, 1983. Use of Landsat data to predict the trophic state of Minnesota lakes, *Photogrammetric Engineering and Remote Sensing*, 49: 219– 229.

Lillesand, T.M., 2002. Combining satellite remote sensing and volunteer Secchi disk measurement for lake transparency monitoring, *Proceedings of National Monitoring Conference 2002*, 19-23 May, Madison, WI.

Lillesand, T.M., R.W. Kiefer, and J.W. Chipman, 2004. *Remote Sensing and Image Processing 5th Edition*, New York, John Wiley, 763p.

Li, Y.,W. Lu, and H. Wang, 2007. Inversing chlorophyll-a concentration by multi-temporal models using TM images, *Proceedings of the 27th International Congress on High-Speed Photography and Photonics*, 17-22 September, Xian, China.

Lu, Q., F. Duckett, R. Nairn and A. Brunton, 2006. 3-D Eutrophication modeling for Lake Simcoe, Canada, *Proceeding of the American Geophysical Union 2006 Fall Meeting*, 11-15 December, San Francisco, CA.

Morel, A. and L. Prieur, 1977. Analysis of variation in ocean color, *Limnology and Oceanography*, 22(4): 709-722.

NASA, 1986. Coastal zone imagery for selected coastal regions. Level II: Photographic Product. Park Ridge II.: The Water. NASA Publication/A. Bohan Co., Washington, DC.

Nelson, S.A.C., P.A. Soranno, K.S. Cheruvelil, S.A. Batzli, and D.L. Skole, 2003. Regional assessment of lake water clarity using satellite remote sensing, *Journal of Limnology*, 62(Suppl.1): 27-32.

Nicholls, K.H., 1992. Water quality trends in Lake Simcoe 1972-1990, Implications for basin planning and Limnological research needs, Lake Simcoe Region Conservation Authority, 23p.

Nicholls, K.H., and H.R. Maccrimmon, 1975. Nutrient loading to cook bay of lake simcoe from the Holland river watershed, *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 60(2): 159-193.

Nicholls, K.H., 1997. A limnogolical basis for a Lake Simcoe phosphorus loading objective, *Lake and Reservoir Managemen*t, 13(3):189-198.

Obrecht, D.V., M. Milanick, B.D. Perkins, D. Ready and J.R. Jones, 1998. Evaluation of data generated from lake samples collected by volunteers, *Journal of Lake and Reservoir Management*, 14(1): 21-27.

Olmanson, L.G., M.E. Bauer, and P.L. Brezonik, 2002. Use of Landsat imagery to develop a water quality atlas of Minnesota's 10,000 Lakes, *Proceedings of ISPRS Comission I*

Mid-Term Symposium in conjunction with Pecora 15/Land Satellite Information IV Conference, 11-15 November, Denver, CO.

Olmanson, L.G., M.E. Bauer and P.L. Brezonik, 2008. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes, *Remote Sensing of Environment,* 112(11): 4086-4097.

Onderka, M., and P. Pekaroua, 2008. Retrieval of suspended particulate matter concentrations in the Danube River from Landsat ETM data, *Science of the Total Environment,* 397(1-3):238-243.

Ostlund, C., D. Pierson, T. Flink and N. Strombeck, 2001. Mapping of the water quality of Lake Erken, Sweden, from imaging spectrometry and Landsat Thematic Mapper, *Science of the Total Environment,* 268: 139-153.

Oyama, Y., B. Matsushita, T. Fukushima, K. Matsushige and A. Imai, 2009. Application of spectral decomposition algorithm for mapping water quality in a turbid lake (Lake Kasumigaura, Japan) from Landsat TM data, *ISPRS Journal of Photogrammetry and Remote Sensing,* 64(1):73-85.

O'Reilly, J.E., S. Maritorena, B.G. Mitchell, D.A.Siegel, K.L.Carder, S.A.Garver, M.Kahru, and C.McClain, 1998. Ocean color chlorophyll algorithms for SeaWiFS, *Journal of Geophysical Research,* 103(C11): 24,937-24,954.

Pinkerton, M.H., K.M. Richardson, P.M. Boyd, M.P. Gall, J. Zeldis, M.D. Olover, and R.J. Murphy, 2005. Intercomparison of ocean colour band-ratio algorithms for chlorophyll

concentration in the subtropical front east of New Zealand, *Remote Sensing of Environment*, 97(3): 382-402.

Pearson's Correlation (1 of 3), 2009. HyperStat Online Contents, URL: http://davidmlane.com/hyperstat/A34739.html (Last date accessed: Jan 15, 2009)

Prangsma, G.J., and J.N. Roozekrans, 1989. Using NOAA AVHRR imagery in assessing water quality parameters, *International Journal of Remote Sensing*, 10(4-5): 811-818.

Ritchie, J.C., C.M. Cooper and F.R. Schiebe, 1990. The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon Lake, Mississippi , *Remote Sensing of Environment*, 33: 137-148.

Richard, R.A., 1996. A spatially adaptive fast atmospheric correction algorithm, *International Journal of Remote Sensing*, 17(6): 1201-1214.

Ruhl, C.A., D.H. Schoellhamer, R.P. Stumpf, and C.L. Lindsay, 2001. Combined use of remote sensing and continuous monitoring to analyze the variability of suspended-sediment concentrations in San Francisco Bay, California, *Estuarine, Coastal and Shelf Science*, 53(6): 801-812.

Rummel, R.J., 1976. *Understanding Correlation*, URL: http://www.hawaii.edu/powerkills/UC.HTM (Last date accessed: Feb 10, 2009)

Sawaya, K.E., L.G. Olmanson, N.J. Heinert, P.L. Brezonik, and M.E. Bauer, 2003. Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery, *Remote Sensing of Environment*, 88(1):144-156.

Schott, J.R., C. Salvaggio and W.J. Volchok, 1988. Radio atmospheric metric scene normalization using pseudo-invariant features, *Remote Sensing of Environment*, 26: 1-16.

Smith, V. H. 1990. *Introduction to Applied Phycology,* SPB Academic Publishing.

Smith, V. H. 1998. *Limitations, and Frontiers in Ecosystem Ecology.* New York: Springer-Verlag.

Sipelgas, L., U. Raudsepp, and T. Kouts, 2006. Operational monitoring of suspended matter distribution using MODIS images and numerical modeling, *Advances in Space Research,* 38(10): 2182-2188.

Stadelmann, T.H., P.L. Brezonik, and S. Kloiber, 2001. Seasonal patterns of Chlorophyll *a* and Secchi Disk Transparency in lakes of East-Central Minnesota: implications for design of ground- and satellite-based monitoring programs, *Journal of Lake and Reservoir Management,* 17(4): 299-314.

Sudheer, K.P., I. Chaubey and V. Garg, 2006. Lake water quality assessment from Landsat Thematic Mapper data using neural network: an approach to optimal band combination selection, *Journal of the American Water Resources Association,* 42(6):1683-1695.

Tanaka, A., M. Kishino, R. Doerffer, H. Schiller, T. Oishi, and T. Kubota, 2004. Development of a neural network algorithm for retrieving concentrations of chlorophyll, suspended matter and yellow substance from radiance data of the ocean color and temperature scanner, *Journal of Oceanography,* 60(3): 519-530.

Thiemann, S., and H. Kaufmann, 2000. Determination of Chlorophyll content and trophic state of lakes using field spectrometer and IRS-1C satellite data in the Mecklenburg Lake District, Germany, *Remote Sensing of Environment*, 73(2): 227-235.

Trainer, T.J. and F.K. Sun, 1991. Image resampling in remote sensing and image visualization applications, *Proceeding of the SPIE's 1991 International Symposium on Optical Applied Science and Engineering*, San Diego, CA (USA), 21-26 July 1991.

Ucuncuoglu, E., O. Arli, and A.H. Eronat, 2006. Evaluating the impact of coastal land uses on water-clarity conditions from Landsat TM/ETM+ imagery: Candarli Bay, Aegean Sea, *International Journal of Remote Sensing*, 27(17):3627-3643.

USGS, 2009. Landsat Thematic Mapper Data (TM), URL: http://eros.usgs.gov/guides/landsat_tm.html#tm7 (Last date accessed: Feb 18, 2009)

United States Environmental Protection Agency, 2007. Carlson's Trophic State Index. *Aquatic Biodiversity*. URL: http://www.epa.gov/bioindicators/aquatic/carlson.html (Last date accessed: Dec 11, 2008)

Wang, F., L. Han, H.T. Kung and R.B. Vanarsdale, 2006. Application of Landsat-5 TM imagery in assessing and mapping water quality in Reelfoot Lake, Tennessee, *International Journal of Remote Sensing*, 27(23): 5269-5283.

Wang, Y., H. Xia, J. Fu, and G. Sheng, 2004. Water quality change in reservoirs of Shenzhen, China: detection using Landsat TM data, *Science of the Total Environment*, 328(1-3): 195-206.

Wang, Z., J. Hong and G. Du, 2008. Use of satellite imagery to assess the trophic state of Miyun Resevoir, Beijing, China, *Environmental Pollution*, 155(1): 13-19.

Werdell, J., 2007. OceanColor Empirical Chlorophyll Analyses, OceanColor web, URL: http://seabass.gsfc.nasa.gov/eval/oc.cgi, (Last date accessed: 11 Aug 2008).

Winter, J.G., M.C. Eimers, P.J. Dillon, L.D. Scott, W.A. Scheider, and C.C. Willox, 2007. Phosphorus inputs to Lake Simcoe from 1990 to 2003: Declines in tributary loads and observations on lake water quality, *Journal of Great Lakes Research*, 33(2): 381-396.

Winter, J.G., M. Walters and C. Willox, 2005. Scientifically derived phosphorus loading objective and adaptive watershed management for Lake Simcoe, Canada, *Proceedings of 53rd Joint Meeting of American Geophysical Union*, 23-27 May, Louisiana, NO.

Witte, R.S., 1989. Describing relationships: *Correlation, Statistics, Third Edition*, (T.Buchholz, S.Arellano, L.Pearson, and J.N.Johnson, editors), J&R Services Inc., New York, N.Y. pp113-133.

Wolter, P.T., C.A. Johnston and G.J. Niemi, 2005. Mapping submergent aquatic vegetation in the US Great Lakes using Quickbird satellite data, *International Journal of Remote Sensing*, 26(23): 5255-5274.

Wu, G.F., J.D. Leeuw, A.K. Skidmore, H.H.T. Prins and Y.L. Liu, 2008. Comparison of MODIS and Landsat TM5 images for mapping tempo-spatial dynamics of Secchi disk depth in Poyang Lake National Nature Reserve, China, *International Journal of Remote Sensing*, 29 (8): 2183-2198.