Monitoring the Underwater Light Climate for Aquatic Vegetation in Poyang Lake by Remote Sensing

FEI TENG
FEBRUARY, 2006
Monitoring Underwater Light Climate for Aquatic Vegetation in Poyang Lake by Remote Sensing

Thesis

by

Fei Teng

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: Planning and Coordination in Natural Resources Management

Thesis Assessment Board

Prof. Dr. Ir. A. de Gier
Chairman, ITC
Dr. Liu Xuehua
External Examiner, Tsinghua University
Prof. Dr. Liu Yaolin
Second Supervisor, Wuhan University
Dr. M.J.C. Weir
Internal Examiner, ITC
Prof. Dr. Liu Yanfang
Internal Examiner, Wuhan University

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION
ENSCHEDE, THE NETHERLANDS
Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.
Abstract

In this thesis, we present a methodology for monitoring the underwater light climate for aquatic vegetation based on spectral reflectance above the water surface. This method is composed by three consequences steps:

First, based on the semi-physical bio-optical model, we built a water reflectance model. This model relates the inherent optical properties (IOP) of water to the apparent optical properties of the lake water and thus can simulate different spectral remote sensing reflectances above the water surface by different combinations of water constituents. This step is called “forward process”. In this study, the water reflectance model gives satisfactory simulation on most cases.

Then in the second step, an artificial neural network is trained by those simulations, and thus can retrieve concentrations of water constituents by inverse the water reflectance model. This step is called “inverse process”. By this method, the ANN-based algorithm can retrieve the concentrations of SPM, CHL and CDOM at the same time. And the accuracies are acceptable (R²=0.758, 0.741 and 0.389 for SPM, CHL and CDOM concentrations respectively).

In the third step, an underwater light climate model was built to predict the light climate at the bottom of the lake, where aquatic vegetation grows, according to the concentrations of water constituents retrieved in step two and the water depth.

The advantage of this method, comparing with empirical methods, is we don’t need to collect a lot of water samples to build the regression formula for retrieving the concentrations of water constituents. And is more site independent.

Key words: IOP, AOP, light climate, lake, underwater, water reflectance model, ANN retrieving, Poyang Lake
Acknowledgements

This thesis is a summary of my postgraduate study on monitoring natural resources through remote sensing. As an exchanging student, I would like to give my hearty thanks to ITC (International Institute for Geo-Information Science and Earth Observation) and Wuhan University, who have given me the precious study chance. I am especially grateful to the sponsor -- Netherlands Fellowship Program (NFP) of my scholarship for its generosity.

I am very much indebted to my supervisors: Dr. C. G. Atzeberger, Dr. J. D. Leeuw and Prof. Dr. liu Yaolin for their invaluable guidance. Through the discussions with you, I got a more and more deeper understanding of the scientific research matter.

I would like to give my special thanks to Prof. Dr. Liu Yanfang, Dr. F. Corsi , Prof. Ai Tinghua Dr. Zhan Qinming and Prof. Dr. A.K. Skidmore who feared no hardship in long international travels in order to give me important advices.

I also want to express my deep appreciation to my dear friends Wu Guofeng, Wang Tiejun and Zheng Ding. Their substantial support made my study possible, and made my life in the Netherlands colorful. My thanks also go to staff of the Poyang Lake Natural Reserve who kindly offered great helps to assist me for the field data collection.

Thanks to my classmates and colleagues in ITC. Our friendship is forever.

Last but not the least, I’d like to thank my family—my father and my mother. With their strong support, I always have the courage to grasp the nettle. By the way, my lovely puppy dog “Do-do” is also thanked, for his quiet company during my busy days.
# Table of contents

1. **INTRODUCTION** ........................................................................................................... 1
   1.1. **BACKGROUND** ................................................................................................. 1
   1.2. **OBJECTIVE AND OUTLINE** ................................................................. 2
       1.2.1. General Objective ...................................................................................... 2
       1.2.2. Specific Objectives ................................................................................. 2
       1.2.3. Research Questions ................................................................................. 3
   1.3. **PRINCIPLE OF THE LIGHT TRANSFER IN THE WATER** ...................... 3
   1.4. **RESEARCH APPROACH** ............................................................................ 4

2. **MATERIAL AND METHODS** .................................................................................. 5
   2.1. **INTRODUCTION** ............................................................................................. 5
   2.2. **CONCEPTS AND DEFINITIONS** .................................................................... 5
       2.2.1. Radiometric Concepts ................................................................................ 5
       2.2.2. Inherent Optical Properties (IOP) ............................................................... 5
       2.2.3. Apparent Optical Properties ......................................................................... 6
       2.2.4. Water Classes ............................................................................................. 6
   2.3. **STUDY AREA** ................................................................................................ 6
   2.4. **DATA AND FIELD DATA COLLECTION** ...................................................... 7
   2.5. **SOFTWARE BEEN USED** ............................................................................... 8
   2.6. **DATA PROCESSING AND ANALYSIS** ........................................................... 9
       2.6.1. The Forward Process ................................................................................. 9
           2.6.1.1. Modelling the Absorption of Natural Water .................................. 9
           2.6.1.2. Modelling the Backscattering of Natural Water ............................... 10
           2.6.1.3. Reflectance above Water Surface Simulation .................................... 12
       2.6.2. The Inverse Process ....................................................................................... 13
           2.6.2.1. Artificial Neural Network ................................................................. 14
           2.6.2.2. Retrieval of the Water Constituents with ANN ............................... 16
       2.6.3. Determining ANN Architecture and Noise Adding for Optimal Water Constituent Retrieval ................................................................................................ 19
       2.6.4. Underwater Light Intensity Model ................................................................. 20

3. **RESULT** ..................................................................................................................... 22
   3.1. **FORWARD PROCESS** .................................................................................. 22
       3.1.1. Model the IOPS of Nature Water ................................................................. 22
       3.1.2. Validation for Forward Process ................................................................. 25
       3.1.3. Conclusion for the Forward Process ......................................................... 27
   3.2. **THE INVERSE PROCESS** .............................................................................. 27
       3.2.1. Determining ANN Architecture for Optimal Water Constituent Retrieval ............................................................................................................. 27
3.2.2. Evaluation the Performance of ANN-Based Inversion Algorithm...... 28
3.2.3. Conclusions for the Inverse Process ........................................ 34
3.3. UNDERWATER LIGHT INTENSITY MODEL ................................ 34
4. APPLICATION TO THE TM IMAGERY (AN ATTEMPT) .............. 37
   4.1. INTRODUCTION ........................................................................ 37
   4.2. IMAGERY DATA ...................................................................... 37
   4.3. ATMOSPHERIC CORRECTION ............................................... 37
   4.4. METHOD ................................................................................. 37
   4.5. UNDERWATER LIGHT CLIMATE............................................ 38
5. DISCUSSION .................................................................................. 42
   5.1. GENERALLY DISCUSSION ...................................................... 42
   5.2. ASSUMPTIONS AND LIMITATIONS IN THE MODEL .............. 44
       5.2.1. Water reflectance model .................................................. 44
       5.2.2. ANN retrieval model ....................................................... 44
       5.2.3. Light intensity model ...................................................... 45
6. CONCLUSION ................................................................................. 46
   6.1. CONCLUSION ......................................................................... 46
   6.2. FUTURE WORK ...................................................................... 46
REFERENCES ...................................................................................... 48
APPENDIX .......................................................................................... 51
List of figures

Figure 1-1 Main conceptual diagram ................................................................. 2
Figure 1-3 Framework and structure of the study ........................................... 4
Figure 2-2 Sampling design ............................................................................ 8
Figure 2-3 Phase function of pure water ........................................................... 11
Figure 2-4 How artificial neural network works .............................................. 14
Figure 2-5 A Multi-Layer-Perception ............................................................... 16
Figure 2-6 400 spectral reflectance simulations .............................................. 17
Figure 2-7 2% noise imported to this model .................................................... 17
Figure 2-8 Over-fitting: the ANN started to fit the noises in training data (blue line), while the error on the validation data become larger (red line) .......... 18
Figure 2-9 Errors of 5 times spectral reflectance sampling ............................. 19
Figure 2-10 Relationship between SPM concentration & underwater light intensity 20
Figure 2-11 Relation between water depth and light intensity ....................... 21
Figure 3-1 Absorption coefficient of pure water .............................................. 22
Figure 3-2 Absorption coefficient of CHL ........................................................ 22
Figure 3-3 Absorption coefficient of SPM ....................................................... 23
Figure 3-4 Absorption coefficient for CDOM .................................................... 23
Figure 3-5 Backscattering coefficients of pure water, SPM and CHL ............ 24
Figure 3-6 Simulation 1: Changing the concentration of SPM ...................... 24
Figure 3-7 Simulation 2: Changing the concentration of CHL ....................... 25
Figure 3-8 Simulation 3: Changing the concentration of CDOM................... 25
Figure 3-9 Simulation test 1 ........................................................................... 26
Figure 3-10 Simulation test 2 ........................................................................... 26
Figure 3-11 Simulation test 3 ........................................................................... 26
Figure 3-12 Scatter plot showing the performance of the ANN based SPM retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set) ................................................................. 31
Figure 3-13 Scatter plot showing the performance of the ANN based CHL retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set) ........................................................................ 32
Figure 3-14 Scatter plot showing the performance of the ANN based CDOM retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set) ................................................................. 33
Figure 3-15 Underwater light climate model .................................................... 35
Figure 3-16 Accuracy of predictions and measurements ............................... 36
Figure 4-1 Water depth map of Dahuchi ......................................................... 39
Figure 4-2 SPM concentration map of Dahuchi .............................................. 40
Figure 4-3 Light intensity map at the bottom of the Dahuchi lake ................. 41
List of tables

Table 1  Data and data collection ................................................................. 7
Table 2  Software been used ...................................................................... 8
Table 3  Water absorption coefficients as function of wavelength ............. 9
Table 4  Ranges of the three water components in simulations ................. 16
Table 5  Measurements dataset collected in-situ .................................... 19
Table 6  The validation of the water reflectance model ......................... 27
Table 7  Evaluation the performance of ANN candidate 1 .................... 29
Table 8  Evaluation the performance of ANN candidate 2 ...................... 29
Table 9  Constant L for atmospheric correction ...................................... 37
Table 10 Wavelengths have been selected to simulate TM image .......... 38
Table 11 Range of retrieved concentrations ............................................ 45
Table 12 Determining ANN Architecture for Optimal Water Constituent Retrieval ..................................................................................................... 55
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>AOP</td>
<td>Apparent Optical Properties</td>
</tr>
<tr>
<td>BP</td>
<td>Back-Propagation</td>
</tr>
<tr>
<td>CDOM</td>
<td>Colored Dissolved Organic Matter</td>
</tr>
<tr>
<td>CHL</td>
<td>Chlorophyll</td>
</tr>
<tr>
<td>DOM</td>
<td>Dissolved Organic Matter</td>
</tr>
<tr>
<td>IOP</td>
<td>Inherent Optical Properties</td>
</tr>
<tr>
<td>MLP</td>
<td>Multi-Layer Perception</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RT</td>
<td>Radiative transfer (RT)</td>
</tr>
<tr>
<td>SPM</td>
<td>Suspended Particulate Matter</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1. BACKGROUND

Many birds feed on tubers of submerged aquatic vegetation during migratory stopover or at their overwintering grounds (Senar and Borras 2004; Taylor and Smith 2005). Therefore, the distribution and biomass of submerged aquatic vegetation is a crucial factor for birdlife’s habitat. The information on the distribution of tuber’s biomass is traditionally collected by field surveys which are time and labor consuming. (Santamaria and Garcia 2004) Remote sensing techniques could have potential ability to produce such maps more efficiently and cheaply. So far, however, few studies has been undertaken to map the distribution of aquatic plant species using this techniques, because it seems impossible to find such submerged vegetations from the RS image directly. Still, techniques have been used to map submerged vegetation (Costa 2004; Steward, Virmstein et al. 2005; Wolter, Johnston et al. 2005). Armstrong reported that almost 80% of the variability in canopy biomass of sea grass could be explained by Landsat TM band 2 and 3, but the accuracy is not satisfied (Armstrong 1993).

Some studies and field observation indicate that there are relationships between the light-climate and the biomass of distribution of submerged aquatic vegetations, because the light intensity is one of the most important elements that affect the process of photosynthesis of these vegetations (Best, Buzzelli et al. 2001; Gulati and van Donk 2002). Take Vallisneria spiralis, the main food of crane, for example, the growth of Vallisneria is strongly influenced by light intensity and the optimum water depth for the growth of this species is 2-3 meters (Chambers and Kalff 1987). As the strong relationship exist, it’s possible for us to map the distribution and biomass of these submerged vegetations indirectly. That means mapping the light climate for these vegetations first, then predicting the growing status of them in different light climates. This study only focuses on the former part which is mapping the light climate for these vegetations using RS image as its title.

In general, water color remote sensing is one of the passive remote techniques. The sensor, mounted on a remote platform, detects the radiometric flux at a range of wavelengths in the visible and near-infrared domains. The signal that received by the sensor is determined by the:

1. Back scattering of sunlight by the water
2. Reflection of sunlight at the water surface (glitter)
3. Bottom reflectance

Only part 1 contains information about the water itself and its inherent optical properties (IOP). However, this part of signature is always quite weak in remotely sensed imagery (Karpouzli, Malthus et al. 2003).
The object system is described below: Water transparency determines the depth of light penetration. It varies as the water quality changes. And the water depth changes through the terrain of the bottom of the lake. The light intensity at the bottom of a lake is depended by both water transparency and water depth. Therefore, one has to map water transparency and water depth respectively to calculate the light intensity distribution in the bottom of a lake by using indirect mapping method. The main conceptual model of this study is exhibited as follow (Figure 1-1).

![Main conceptual diagram](image)

Figure 1-1  Main conceptual diagram

First, from the spectral reflectance above the water surface, we retrieve the key concentrations of water constituents. Then we build a model to simulate the light intensity underwater with a certain concentrations of constituents. Last, we use real value to generate a light intensity map on our study area.

**1.2. OBJECTIVE AND OUTLINE**

**1.2.1. General Objective**

Monitoring the underwater light climate for aquatic vegetation by predicting the underwater light intensity based on spectral reflectance above the water surface.

**1.2.2. Specific Objectives**

In order to achieve the general objective above, several specific questions will be addressed:

- Investigation of the relationship between concentrations of water constituents and spectral reflectance above water surface.
- Find or build a suitable physical based model (water reflectance model) that considers necessarily water constituents parameters as inputs, and gives outputs on spectral reflectance above the water surface.
- To retrieve important water parameters from a spectral reflectance by artificial neural network (ANN) based on the model built in forward process.
- To calculate the light intensity underwater using these retrieved concentrations as parameters.
1.2.3. Research Questions

- Is it feasible to predict the underwater light climate by using RS technique?
- Can we retrieve water constituents from RS image by a physical based model?
- What is the accuracy of this kind of prediction?
- Can we find relationship and how to find it?
  - Between RS image and water constituents and their concentrations
  - Between concentration(s) of the key water constituent(s) and water transparency
  - Underwater light intensity and water depth
- How is the accuracy of the water quality parameters that retrieved from the model?

1.3. PRINCIPLE OF THE LIGHT TRANSFER IN THE WATER

Radiative Transfer through the Earth’s atmosphere is described by the radiative transfer equation (Lenoble 1993). Total radiance, \( L_t \) recorded by a remote sensing system over water is a function of the electromagnetic energy received from:

\[
L_p = \text{atmospheric path radiance} \\
L_s = \text{water surface layer reflectance} \\
L_v = \text{subsurface particle reflectance} \\
L_b = \text{bottom reflectance}
\]

In general, \( L_t = L_p + L_s + L_v + L_b \) (as Figure 1-2)

A radiative transfer model describes how light is absorbed and scattered by components of the water column in a physical way. Give it all the necessary parameters such as water quality parameters, water depth, and bottom reflectance; it will work out the total reflected radiance (\( L_v \)).

After the model been calibrated, it’s possible to reverse it and retrieve all the input parameters from the former output: total radiance (\( L_t \)). That is, we can retrieve the \( L_v \) (subsurface particle reflectance) from the RS imagery.

Some researches had been done by using the reversible physical based model. Mapping the concentration of organic and inorganic materials has been implemented in lake water bodies. Some software was developed to recover hydro-biological parameters from multi- and hyperspectral remote sensing data. Mapping of substrate cover types has been carried out successfully in optically clear, shallow coastal waters. With limited exceptions, successful remote sensing applications have mapped either water constituents or substrate cover types, but not both.
1.4. RESEARCH APPROACH

**FORWARD PROCESS**

- Pure Water
- Pigment (CHL)
- Colored Dissolved Organic Matter (CDOM)
- Suspended Particular Matter (SPM)

**INVERSE PROCESS**

- Simulations for Training
  - Randomly Water Constituents Concentrations Simulation
  - Corresponding Reflectance Above the Water Surface Simulations

**STEP 1**

- Untrained Artificial Neural Network (ANN)

**STEP 2**

- Real Reflectance Measurements
- Trained Artificial Neural Network (ANN)

**UNDERWATER LIGHT CLIMATE**

- Bottom DEM
- Water depth
- Suspended Particular Matter (SPM)
- Underwater Light Intensity Model
- Underwater Light Climate

**FORWARD PROCESS**

- Simulate the spectral remote sensing reflectance just above the water surface by given concentrations of water constituents: SPM, CHL, CDOM

**INVERSE PROCESS**

- Training an ANN with simulations generated in forward process, then use the trained ANN to retrieve water constituents and their concentrations

**UNDERWATER LIGHT CLIMATE**

- Based on SPM concentration retrieved, predicting the light intensity at the bottom of the lake that can received by aquatic vegetation growing at the bottom

Figure 1-3 Framework and structure of the study
2. MATERIAL AND METHODS

2.1. INTRODUCTION

For calculating the underwater light climate, the key step is to retrieve concentrations for water constituents correctly, especially the concentrations of suspended particulate matters (SPM) that dominates the lake water. Then, we can build an underwater intensity model which can relate concentration and water depth to the light intensity at the bottom of the lake. In this study, we divide the method to three parts: forward process, inverse process, and light intensity modelling.

The forward process tries to build a physical based model to simulate how solar radiation is absorbed in the water and scattered back to the air. Giving the model concentrations of water constituents which are optically significant, the model should computed the IOPs of the water—total absorption coefficient and total scattering coefficient. Then, based on the IOPs, the model predicts spectral reflectance above the water surface as a simulation. This procedure is called forward process, because it tries to simulate what nature does.

Then, after a carefully built artificial neuron network (ANN) “learnt” hundred simulations generated by the forward process, it stepped in the inverse process, in which the ANN is trying to give prediction of concentrations of these three optically significant water constituents: SPM, CHL (for pigment) and CDOM, based on the given reflectance above the water surface. This is called inverse process, because it tries to seek for the reasons (concentrations) based on the result (reflectance).

At last, the concentration of water constituents retrieved from the part 2 are inputted to the underwater light intensity model, and compute the light climate at the bottom of the lake, where aquatic vegetation grows.

2.2. CONCEPTS AND DEFINITIONS

2.2.1. Radiometric Concepts

- **Radiance** $L(\theta, \varphi, \lambda)$: The radiant power in a beam per unit solid angle per unit area perpendicular to the beam per unit wavelength interval.
- **Irradiance** $E(\lambda)$: The radiant power per unit area per unit wavelength interval. where $\theta$ is the zenith angle, and $\phi$ is the azimuth angle.

2.2.2. Inherent Optical Properties (IOP)

The fundamental IOPs are the absorption coefficient and the scattering coefficient.

- **The spectral absorption coefficient** $a(\lambda)$ is defined as the spectral absorptance per unit distance of photon travel in a dielectric medium.
- **The spectral scattering coefficient** $b(\lambda)$ is defined as the spectral scatterance per unit distance of photon travel in a dielectric medium.
2.2.3. Apparent Optical Properties

Two apparent optical properties are widely used to describe the apparent color of the water: hemispherical reflectance \( R(\lambda) \) and remote sensing reflectance \( RRS(\lambda) \).

- **The spectral hemispherical reflectance** is defined as the ratio of spectral upward to downward irradiance below the water surface.
- **The spectral remote sensing reflectance** is defined as the ratio of the water leaving radiance to the downward irradiance just above the water surface.

2.2.4. Water Classes

- **Case I water** is the waterbody that its optical properties are determined solely by phytoplankton and their degradation products and water itself.
- **Case II water** is the waterbody that its optical properties are determined additionally by suspended particulate matter (SPM) and colored dissolved organic matter (CDOM).

2.3. STUDY AREA

Po Yang Lake in Jiangxi Province is a huge lake surrounded by a number of smaller lakes (Figure 2-1). In the summer the lakes are filled by the flooding of the Yangtze River (Zhou, Lin et al. 2002), however in the fall as the water subside, vast areas of shallow water and extensive muddy banks are created, producing ideal habitat for cranes, ducks and geese (Zhang and Lu 1999).

As the largest freshwater lake in China and wintering area for nearly the entire world’s endangered Siberian Cranes as well as a major wintering ground for White-naped and Hooded Cranes, Poyang Lake in Jiangxi Province is a critical area for the future of crane conservation in China, with 98% percent of the global population of Siberian Cranes, 60% of White-naped Cranes, 50% of Swan Geese, and tens of thousands of egrets, spoonbills, swans, geese, ducks, and shorebirds wintering at the site (ICF 1998; Kanai, Ueta et al. 2002; Fasola, Galeotti et al. 2004; Higuchi, Pierre et al. 2004).

The Poyang flood plain is subject to massive changes in water level. In the dry season the size of the lake shrinks to less than 1,000 km², and in the rainy season (late summer) the size of the lake grows up to 4,000 km². The average depth is 8 meter; the maximum depth is 23 m.

The Chinese government has taken measurements to protect wildlife. In 1983, the local Government of Jiangxi Province established the Poyang Lake Nature Reserve mainly for the protection of Siberian crane. The Chinese crane Conservation Union, established in 1984 by the former Ministry of Forestry, includes 19 crane range provinces and autonomous regions, for the conservation, management, and research of crane in China. The regulation, constituted by Jiangxi Province, has been implemented from January 1st 2004 forbidding
hunting and fishing at Poyang Lake from October to March for the conservation of over-wintering waterfowl. Although the Nature Reserve protects some of the most important wintering sites in this area, but the birds also use adjacent sites outside the reserve.

This study is focused on Dahuchi reserve, the core area of Poyang Lake Nature Reserve.

2.4. DATA AND FIELD DATA COLLECTION

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Data Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigate the relationship between concentrations of water constituents and above surface reflectance.</td>
<td>SPM concentration; chlorophyll concentration; Yellow substance concentration Spectral reflectance above water surface</td>
</tr>
<tr>
<td>Remove atmospheric impact on spectral reflectance</td>
<td>White plate reflectance</td>
</tr>
<tr>
<td>Train the multiple layers artificial neuron network to retrieve the concentrations</td>
<td>800 simulations generated from the forward process</td>
</tr>
<tr>
<td>Investigate the relationship between water quality and light intensity under water.</td>
<td>Light intensity under water at different water quality</td>
</tr>
<tr>
<td>Investigate the relationship between water depth and light intensity under water.</td>
<td>Light intensity under water at different water depth</td>
</tr>
<tr>
<td>Modelling the underwater light intensity in Poyang lake</td>
<td>RS imagery (TM5)</td>
</tr>
<tr>
<td>Testing the accuracy of the light intensity model</td>
<td>Independent field dataset of light intensity underwater</td>
</tr>
</tbody>
</table>

The water samples and field data was collected in these plots (Figure 2-2):
In this study, random sampling method is not economical feasible for us, however, in order to have a global impression of the water quality in the study lake, the sample line we design started from the south bank, go across the lake to the north bank, than turn right, to the east bank of the lake (as Figure 2-2 shown above).

In each plot, we stop the boat and measure the spectral reflectance by a radiometer in the sunny side of the boat just above the water surface. And we recorded 5 spectral reflectances at the same plot for error analyses. A white plate was used to record the downwelling solar radiance before sampling each plot. By the white plate data, we could erase the variance of atmospheric conditions.

2.5. SOFTWARE BEEN USED

Table 2  Software been used

<table>
<thead>
<tr>
<th>Function</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record Hyperspectral reflectance</td>
<td>Spectral Wizard</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>MS Excel</td>
</tr>
<tr>
<td>Statistical Programming</td>
<td>SPSS, Matlab 6.1, SYSTAT 11</td>
</tr>
<tr>
<td>Programming</td>
<td>Mathwork Matlab 6.1</td>
</tr>
<tr>
<td>Image processing</td>
<td>Erdas 8.7, ArcMAP</td>
</tr>
<tr>
<td>Word processing</td>
<td>MS Word</td>
</tr>
</tbody>
</table>
2.6. DATA PROCESSING AND ANALYSIS

2.6.1. The Forward Process

It has been introduced in chapter 1, this study consists of three main parts, the first part is to simulate how radiance transferring through the water and goes back to the air. For this sake, we need to build the Water reflectance model. That is called “forward process”.

2.6.1.1. Modelling the Absorption of Natural Water

We knew that

\[ A = a_w + C_s a_s + C_c a_c + C_y a_y \]

Where \( a_w \) refers to the absorption coefficient for the pure water. \( a_s, a_c, a_y \) is the absorption coefficients for suspended particulate matter (SPM), pigment (CHL), and yellow substance (CDOM) (Chomko and Gordon 2001). For the purpose of building the water reflectance model, we have to fix the absorption coefficients for each constituent in water as well as for water itself.

(1) Pure Water

The absorption coefficient of pure water \( a_w (\lambda) \) has been measured with high accuracy for most applications and often taken as known (Mobley 1994). Smith and Baker (Smith and Baker 1981) gave a set of data of water absorption coefficient at each wavelength as the Table 3 shown below, from the wavelength 400 to 800 (nm). The columns in grey are the wavelengths, columns in white are the water absorption coefficients.

<table>
<thead>
<tr>
<th>wavelength</th>
<th>Water abs</th>
<th>460</th>
<th>470</th>
<th>480</th>
<th>490</th>
<th>500</th>
<th>510</th>
<th>520</th>
<th>530</th>
<th>540</th>
<th>550</th>
<th>560</th>
<th>570</th>
<th>580</th>
<th>590</th>
<th>600</th>
<th>610</th>
<th>620</th>
<th>630</th>
<th>640</th>
<th>650</th>
<th>660</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>0.0171</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>0.0153</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>430</td>
<td>0.0144</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>440</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>460</td>
<td>0.0171</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>470</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>490</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>510</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>520</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>530</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>540</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>560</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>570</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>580</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>590</td>
<td>0.0156</td>
<td>0.0156</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>600</td>
<td>0.0162</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>610</td>
<td>0.0153</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>620</td>
<td>0.0144</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>630</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>640</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>650</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>660</td>
<td>0.0145</td>
<td>0.0176</td>
<td>0.0196</td>
<td>0.0257</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0507</td>
<td>0.0558</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Redraw from Smith and Baker (1981)

(2) Pigment Concentration

The pigment concentration is mostly affected by chlorophyll a (CHL) in the lake. So we take CHL concentration as pigment concentration in this study. Unfortunately, we couldn’t find any accurate
CHL absorption coefficient model, because the coefficient varies through locations and species of phytoplankton (Prieur and Sathyendranath 1981; Morel 1988; Carder, Hawes et al. 1991; Hoepffer and Sathyendranath 1993; Bricaud, Morel et al. 1998).

What we do is that we obtained data of absorption coefficient of CHL from "EUMELI" – a French project leaded by A. MOREL (Morel 1992). We averaged the absorption coefficient of sampling data near the sea surface (depth<30), and use the result in this study as the absorption coefficient of CHL in lake water.

(3) Suspended particulate matter (SPM)

The absorption coefficient of SPM $a_s$ can be derived from formula given by (Roesler and Perry 1995), which is shown below:

$$ a_{spm}(\lambda) = a_{spm}(400)\exp[-0.011(\lambda - 400)], $$

Where and $a_{spm}(400)$ depends on the average diameter of the particles, normally in a lake, $a_{spm}(400)$ is around 0.1 (Chomko, Gordon et al. 2003) (Bricaud, Morel et al. 1998). This number will be taken in this study to derive the absorption coefficient of SPM. And we will see in the later parts of the thesis, it’s a nice approximation.

(4) Yellow Matter (CDOM)

The absorption coefficient for yellow mater is formulized by Bricaud (Bricaud, Morel et al. 1981) in this format:

$$ a_y(\lambda) = a_y(\lambda_0)\exp[-0.014(\lambda - \lambda_0)], $$

Where $a_y(\lambda_0)$ is the absorption coefficient at wavelength $\lambda_0$. At 440nm we have (Mobley 1994) $a_y(440) = 0.30$.

By now, we have modelled the absorption coefficients of natural water, which has SPM, CHL and CDOM in it. Next, we are ready to model to scattering coefficient of case II water.

2.6.1.2. Modelling the Backscattering of Natural Water

Let’s review the formula:

$$ B = b_{bw} + C_s b_{bs} + C_w b_{bc} \quad \text{(Chomko and Gordon 2001)} $$

Similar from formula of absorption coefficient, the $b_{bw}, b_{bs}, b_{bc}$ are the backscattering coefficient for pure water, SPM and CHL, respectively. The yellow substance (CDOM) is considered only absorbing but not scattering the radiation (Dekker, Pasterkamp et al. 1999).

We know that:

$$ \tilde{\beta}(\psi, \lambda) = \frac{\beta(\psi, \lambda)}{b(\lambda)} $$
Where $\beta(\psi, \lambda)$ is the volume scattering function, $b(\lambda)$ is the scattering coefficient (notice: NOT backscattering coefficient). And the term $\beta(\psi, \lambda)$, is called phase function. Writing the volume scattering function $\beta(\psi, \lambda)$ as the product of the scattering coefficient $b(\lambda)$ and the phase function $\beta(\psi, \lambda)$ divide $\beta(\psi, \lambda)$ into two parts: a factor giving the strength of the scattering, $b(\lambda)$, and the other factor giving the angular distribution of the scattered photons, $\beta(\psi, \lambda)$. (Mobley 1994) (Fournier and Forand 1994)

Then the Backscattering coefficient can be solved by:

$$b_b(\lambda) = b(\lambda) \times \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \beta(\psi, \lambda),$$

Or

$$b_b(\lambda) = b(\lambda) \times \text{Backscattering ratio;}$$

Where Backscattering ratio is defined as $\frac{b_b(\lambda)}{b(\lambda)}$. (Kempeneers, Sterckx et al. 2005)

(1) **Pure Water**

The scattering coefficient of pure water is often taken as known. The total scattering coefficient $b_{bw}(\lambda)$ is given by Mobley (Mobley 1994)

$$b_{bw}(\lambda) = 16.06 \left( \frac{\lambda_0}{\lambda} \right)^{4.32} \beta_w(90^\circ, \lambda_0)$$

Where $\beta_w(90^\circ, \lambda_0) = 0.93 \times 10^{-4}$, when $\lambda_0 = 550$nm.

The phase function of pure water is (Mobley 1994)

$$\beta_n(\psi) = 0.06225(1 + 0.835 \cos^2 \psi).$$

---

**Figure 2-3** Phase function of pure water

So

$$b_{bw}(\lambda) = b_w(\lambda) \times \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \beta_w(\psi, \lambda)$$
From it we can derive the backscattering coefficient for pure water, which is plotted in Figure 3-5

(2) Suspended particulate matter

For SPM, it is assumed that \( b_s(\lambda) \) is proportional to SPM concentration with (Morel, Sicre et al. 2003)

\[
\frac{b_s(\lambda)}{SPM} = 0.54
\]

And the backscattering ratio \( \frac{b_{bs}}{b_c} = 0.0096 \) (Kempeneers, Sterckx et al. 2005)

By \( b_b(\lambda) = b(\lambda) \times \) Backscattering ratio, we could have the backscattering coefficient for SPM from wavelength of 400 to 800. (Figure 3-5)

(3) Chlorophyll

We know the scattering coefficient due to phytoplankton and covarying particles \( b_c(\lambda) \) is given as a function of CHL: \( b_c(\lambda) = b_c(660) \times \left( \frac{660}{\lambda} \right)^{0.7} \), with \( b_c(660) = 0.407 \times CHL \). And the backscattering ratio is 0.01833. (Kempeneers, Sterckx et al. 2005) Then we could get the backscattering coefficient for CHL (Figure 3-5) as well as SPM.

2.6.1.3. Reflectance above Water Surface Simulation

Models that predict water spectral reflectance as a function of the total absorption coefficient \( A \) and the total backscattering coefficient \( B \) were used for analysis and processing of ocean color data. (Gordon, Brown et al. 1988; Morel 1988; Garver and Siegel 1997; Gons, Rijkeboer et al. 2002)

Park (Park and Ruddick 2005) developed an expression to derive the subsurface reflectance from absorption coefficient \( A \) and the backscattering coefficient \( B \), as below,

\[
r_s(0^-) = l_1 \frac{b_b}{a + b_b} + l_2 \left( \frac{b_b}{a + b_b} \right)^2
\]

Where \( r_s(0^-) \) is the subsurface reflectance, which is defined by the ratio of upwelling radiance to downwelling irradiance just below the water surface. And \( l_1 = 0.0949, l_2 = 0.0794 \) (Gordon, Brown et al. 1988)

To get the reflectance just above the surface (RS reflectance) from the subsurface reflectance, we need to know the percentage of radiation that could penetrate the water-air surface. The Fresnel reflectance \( r(\theta) \) lays in the interval 0 to 1 and gives the fraction of photons incident in a beam along (from air or water) that is reflected by the surface. Therefore, \( 1 - r(\theta) \) is the percentage we are seeking. The function of Fresnel reflectance (Xu and Liu 2004) when \( \theta = 0 \) is in the form of:
\[ r(0) = \left( \frac{n_w - 1}{n_w + 1} \right)^2 \]

Where \( n_w \) is the refractive index of the water, \( n_w \approx 1.33 \). So, \( r(0) \approx 0.2 \)

Because the behaviour of the Fresnel reflectance function, For both air and water-incident rays, the reflectance is 0.02 to 0.03 for rays with incident angles of less than 30° (Mobley 1994). That means: the ratio of radiance that could penetrate the water-air surface to the total upwelling radiance is not very sensitive to the incident angle, when the angle is less than 30°. In this study, \( r(0) = 0.2 \) is taken.

Finally, the reflectance just above the water surface can be calculated by:

\[
rs(0^+) = l_1 \frac{b_b}{a + b_b} + l_2 \left( \frac{b_b}{a + b_b} \right)^2 \times (1 - r_0) = 0.98 rs(o^-) \\
= 0.98[l_1 \frac{b_b}{a + b_b} + l_2 \left( \frac{b_b}{a + b_b} \right)^2]
\]

Here, \( a = A, b_b = B \). We could solve it by putting \( a, b_b, l_1 \) and \( l_2 \) which are already known into the formula.

### 2.6.2. The Inverse Process

Through the content above in this chapter, we have simulated the water leaving radiance (spectral reflectance) just above the water surface successfully. That means we have finished the so-called “forward process”. We now turn our attention to the “inverse process”, which converts the received spectral information (here we mean hyperspectral reflectance above the water surface) into the prediction of the concentrations of the three elements—SPM, Pigment, and yellow matter (CDOM).

The inverse process is challenged by both conceptual and practical limits (Atzberger 2004). The first limit is called “uniqueness of the solution” -- if we can retrieve same constituents and their concentrations from two or more different spectral reflectances, the retrieval solution is not “unique”. Put it in another way, the question is: Does it even in principal have the possibility to inverse the water reflectance model and have a unique solution? The second limit is “sensitivity” problem -- consider the following case: We could change the water constituents in its concentration a litter, say, add 5%. It is possible that the spectral remote sensing reflectance (or other apparent optical property) varies very little. That means, in the inverse process, a small error in measuring the spectral remote sensing reflectance (or other AOPs) will lead to a large error or even unphysical solution in the concentration retrieved. We can see it happens in the latter Chapter.

Theoretical, since we developed a physical based model, we can reverse the process by mathematical approach. However, this method requires a fairly sophisticated understanding of radiative transfer equation, plus the algorithm is so complicated that it takes quite a long time to retrieve constituents at just one point even for a PC with a 1.2MHZ CPU. Besides these mentioned above, the solvability of such mathematical algorithm is uncertain.

For those reasons people starting to figure out other relatively simple ways to solve the problem, including the Monte Carlo ray tracing methods(Ishimoto and Masuda 2002), invariant imbedding solution methods (Mobley 1994), and Numeric matrix-operator inversion (Dekker, Pasterkamp et al.)
1999; Fell and Fischer 2001), most of which are trying to give approximating solutions to the mathematic equations.

Algorithms for retrieving pigment concentration in Case I water from hyperspectral image or from just above the surface are available (O'Reilly, Maritorena et al. 1998; Cipollini, Corsini et al. 2001; Dransfeld, Tatnall et al. 2004; Chauhan, Nagamani et al. 2005). Unfortunately, these algorithms mostly failed in Case II water such as inland lakes or shallow water (Hetscher, Krawczyk et al. 2004). The reasons that make the retrieval more difficult in Case II water than Case I water are listed below:

The main reason is the extremely high concentration of SPM in Case II water by definition. When the suspended particulate matter dominates the water body, the spectral reflectance is also dominated by the characteristic spectrum of SPM, and it may overcast the contribution of other constituents to the measured signal.

The second reason is that we have to retrieve three constituents from Case II water instead of two constituents in Case I water, because SPM is not considered in Case I water by definition.

In the following, a method is proposed to retrieve the concentrations of constituents in Poyang Lake based on artificial neural network techniques. Inputs to the model are the spectral hemispherical reflectance at all available wavelengths that can be either collected just above the water surface or be retrieved from the satellite images. A synthetic dataset that generated from our water reflectance model which has been built in this chapter will be used to train the ANN. And we expect the ANN gives accurate predictions on concentrations of constituents in the water of the corresponding position.

2.6.2.1. Artificial Neural Network

Neural networks are composed of simple elements operating in parallel. (Demuth, Beale et al. 2005) In other words, a neural network is composed of a number of “neurons”, which are arranged in different layers of the network and are connected by weighted links. By adjusting the weights between neurons, we can train an ANN to execute a certain function. The training process is like this, as Figure 2-4 shown below: The network is adjusted automatically, based on comparing the output and the target, until the network output matches the target within a certain range (threshold).

![Figure 2-4 How artificial neural network works](Redraw from (Demuth, Beale et al. 2005))
In a neural network, two or more neurons can be combined in one layer, and a particular network could contain one or more such layers.

In this study, the Multi-Layer-Perception (MLP) is used to find the relationship between the water color (broad sense) and the concentrations of the lacustrine constituents—SPM, pigment (CHL), and Yellow Matter. A MLP consists of three or more layers: one input layer, several hidden layers and an output layer. (Figure 2-5)

It had been proved that a neural network with one hidden layer can approximate any continuous function to any given accuracy if a sufficient number of neurons are set (Cybenko 1989; Masters 1993). A bias parameter is then added to all the layers (both input layer and hidden layers). The neurons in the hidden layers are connected with each others and with every neuron in the input and output layers. The way of transferring information through the network is described by the following equation:

\[ O = s_c(W^{HO} \times s_c(W^{HH} \times I)) , \]

Where \( I \) is the input vector to the MLP in which consists of spectral reflectance values at a certain position just above the water surface. The weight matrix \( W^{HH} \) contains the weights of all connections between input and hidden layers, and the weight matrix \( W^{HO} \) contains the weights of all connections between hidden and output layer. At each neuron of hidden and output layer, a non-linear sigmoid function is applied to them to transfer the information. The sigmoid function is defined by:

\[ s_c(x) = \frac{1}{1 + \exp(-c_s x)} , \]

Where \( c_s \) is called the “temperature constant”. The ANN import the sigmoid function in order to confine the output of the network to the interval of \([0, 1]\).

Training of the ANN is done by back-propagation (BP) method. Back-propagation is a gradient descent algorithm, in which the weights are moved along the negative of the gradient of the performance function. The term “back-propagation” refers to the way in which the gradient is computed for multilayer networks.
2.6.2.2. Retrieval of the Water Constituents with ANN

It has been shown that Artificial Neural Network (ANN) techniques have a potential to derive the water constituents (Gross, Thiria et al. 1999; Keiner and Brown 1999). Two kinds of data are used in this study: In-situ measurement and water reflectance simulations. They both relate the spectral remote sensing reflectance to the concentrations of pigment, suspended particulate matter, and CDOM.

- Training Data

The training data used to train the ANN for lake constituents retrieval was generated from computer code developed in this chapter, based on the water reflectance model. The code can generate hemispherical spectral reflectance curve by randomly given concentrations of water constituents.

Because the MLPs use sigmoid function as links between layers, it is not good at extrapolation. If the actual measurement goes out of the ranges of training data, the ANN will give a “saturated” output—which has the same value to the largest (or smallest) output of the training data. Therefore in this case, it must be made sure that the ranges of the parameters in training data cover the possible ranges of actual measurements. In this study, the ranges of the three components are listed in Table 4 below:

<table>
<thead>
<tr>
<th>Components</th>
<th>Unit of Concentration</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPM</td>
<td>Mg/m³</td>
<td>1</td>
<td>1000</td>
</tr>
<tr>
<td>CHL</td>
<td>g/m³</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>CDOM</td>
<td>Mg/m³</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

We can see from the table, the range of training data is limited. By this approach we could reduce the potential ambiguous solutions for the retrieval of three constituents.
It is assumed, in this study, the concentration of each of components is independent from the others (Park and Ruddick 2005). So it’s not necessary to find the correlation between constituents.

Based on the ranges of concentrations of constituents above and the water reflectance model, simulations of the spectral reflectance just above the water surface were made for:

- 400 combinations of the water constituents,
- 41 wavelengths: from 400 nm to 800 nm, by the interval of 10nm. (400, 410 …790, 800)
- Adding 3 different noise levels (2%, 5%, and 10%).

The generated 400 simulations based on water reflectance model without noise are shown below (Figure 2-6):

![Figure 2-6 400 spectral reflectance simulations](image)

In order to build a robust ANN, we have to import noise to this model (Zhang, Fell et al. 2003). Because the real measurements contain errors, we have to simulate also the errors. Based on the field measurements, a typical error level of 2% was detected. With the purpose of testing how much noise at the most can be accepted by the ANN, we tested data with 3 different noise level: 2%, 5%, and 10%. A single simulation with 2% noise level is shown below as comparison (Figure 2-7):

![Figure 2-7 2% noise imported to this model](image)

The red curve is the reflectance simulation with 2% measurement error, comparing with the blue curve, in which no noise is added.
Finally, the training dataset mixed 50% of ideal data generated from our Water reflectance simulations, and 50% of “noised” data. By this way, we expect the ANN we trained can provide accurate predictions for both cases: when the input data contain certain level of error or nearly no errors.

- **Validation Data**

In order to prevent the ANN from over fitting -- A case in which the error on the training set is driven to a very small value, but when new data is presented to the network, the error is large (Demuth, Beale et al. 2005) (Figure 2-8), we also generated validation data for 200 combinations of the water constituents. The same as training data, the range of the concentrations of the constituents is derived from Table 4. And the same noise level is also applied to half of the data randomly, as what we’ve done to the training dataset. If the ANN start to fit the noises in training data, the error on the validation set is become larger, then the training process stopped automatically. In this way, we can ensure the ANN provides generalized results.

![Figure 2-8 Over-fitting: the ANN started to fit the noises in training data (blue line), while the error on the validation data become larger (red line)](image)

- **In-situ Measurement Data Sets**

The in-situ measurements data used in this study come from the field sampling in Poyang Lake at October 16, 2005. Spectral remote sensing reflectance just above the water surface (ratio of upwelling radiance to downwelling radiance) was collected by a radiometer carefully. When we begin a measurement, the downwelling radiance on that particular time is also recorded by a white plate, so that the ratio can be normalized based on it.

It was a sunny day that day, and the wind speed is less than 2 meter per second. The sampling time during that day was from 11:30 AM to 12:25 PM while the solar zenith is ranged between 5 to 15 degrees (Observatory 2005). The water samples were taken into a dark container in order to prevent the Chlorophyll a (CHL) from acidification (Stich and Brinker 2005). Then the samples were transferred to the lab of school of resource and environmental science (SRES), Wuhan University the
next day. Then the samples were analyzed for deriving concentrations of SPM, CHL, and CDOM. Those measurements data are used for testing the accuracy of retrieval of the three constituents. Table 5 provides information on the in-situ measurements dataset chosen for this study.

Table 5 Measurements dataset collected in-situ

<table>
<thead>
<tr>
<th>Position</th>
<th>Spectral reflectance</th>
<th>Concentration of SPM</th>
<th>CHL</th>
<th>CDOM</th>
<th>Secchi-depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAR Light intensity of: +0 (above the surface), -0 (just beneath the surface), -10, -20… -90 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As we have mentioned before, we had measured the spectral reflectance 5 times in each plot independently. We use the average of them as a validation dataset to test the Water reflectance simulations in these positions. And we estimated the error of a typical measurement. A feeling of the error of spectral measurements can be obtained from the figure below (Figure 2-9):

![Figure 2-9 Errors of 5 times spectral reflectance sampling](image)

We found the error of a typical measurement can hardly larger than 2% from wavelengths of 400 nm to 800 nm in the field.

### 2.6.3. Determining ANN Architecture and Noise Adding for Optimal Water Constituent Retrieval

In order to find the best ANN that suit for the concentrations retrieval, the ANN predictions were compared to the validation data and the in-situ measurements by two error measures. First, the Pearson’s correlation coefficient: \( r^2 \). Second, root mean square error (RMSE) defined by:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [(SPM_{i}^{\text{prediction}}) - (SPM_{i}^{\text{in-situ}})]^2}
\]

Where SPM represents the SPM concentration, and the superscripts prediction and in-situ stand for derived and measured values.

The inputs of the ANN are the spectral reflectance at 41 wavelengths between 400 and 800 nm. So it’s fixed that the input layer contains 41 neurons. Until now, a systematic way to decide the “best” ANN
architecture and noise level fit for this particular problem hasn’t been developed. So we have to try the different combinations of the ANN architecture with different noise levels. Then the best performing ANN can be selected to retrieve the concentrations of water constituents.

2.6.4. Underwater Light Intensity Model

By now, the concentrations of water constituents (especially SPM) can be retrieved from spectral measurements above water surface. We then need to relate the underwater light intensity to the concentrations of constituents of the lake water.

In case II water, suspended particulate matter dominates the water constituents. In my study area, Poyang Lake, the concentration of SPM can even reach 700mg/L. On the other hand, CHL and CDOM have only concentrations of less than 0.02mg/L at most. Therefore, in the process to estimate the underwater light climate, we can ignore the variance of CHL and CDOM because very little contributions are caused by them.

We put some light intensity probes under the water at the same depth (10cm) but different locations with different SPM concentrations to collect the data of the light intensity. Here is the result (Figure 2-10):

![Figure 2-10 Relationship between SPM concentration & underwater light intensity](image)

Strong exponential relationship (R=0.89) is observed and the exponential regression formula is

\[ y = ae^{bx} \]

Formula 2-1

With coefficient data:

\[
\begin{align*}
a &= 863.12 \\
b &= -0.00276
\end{align*}
\]

Where \( x \) is the concentration of SPM in the unit of mg/L, and \( y \) is the underwater light intensity at depth of 10 cm, the unit is mol/m\(^2\)*s.

From this formula we could solve the light intensity in the depth of 10 cm by a given SPM concentration.
Then we examine the relationship between light intensity and water depth, in the same SPM concentration. Our data is recorded in the same location with the same SPM concentration (150mg/L) but different depth of water. We plotted the result below (Figure 2-11):

![Relation between water depth and light intensity](image)

Figure 2-11  Relation between water depth and light intensity

By an Exponential Fit, the \( R^2 \) can reach 0.99. The model is like:

\[
y = ae^{bx}
\]

Formula 2-2

Where \( x \) is the water depth in the unit of centimetre, and \( y \) is the light intensity in the unit of mol/m\(^2\)s.

And the coefficient Data:

- \( a = 1132.29 \)
- \( b = -0.0623 \)

Now we are able to calculate the underwater light intensity by any given SPM concentration and water depth. Combine Formula 2-1 and Formula 2-2 we can get:

\[
Z = \frac{ae^{by} ce^{dx}}{570.5231}
\]

Formula 2-3

Where \( x \) is the SPM concentration, \( y \) is the water depth; \( z \) is the light intensity underwater. And the coefficient Data:

- \( a = 1132.29 \)
- \( b = 0.0632 \)
- \( c = 863.12 \)
- \( d = -0.00276 \)

By this integrated formula, we can calculate the underwater light intensity at any depth, with any SPM concentration.
3. RESULT

3.1. FORWARD PROCESS

3.1.1. Model the IOPS of Nature Water

In the forward process, the absorption coefficient of pure water is plotted below (Figure 3-1):

![Figure 3-1 Absorption coefficient of pure water](image1)

The absorption coefficients of Pigment concentration (CHL) is calculated and plotted below (Figure 3-2):

![Figure 3-2 Absorption coefficient of CHL](image2)

We noticed that the absorption peaks appear in wavelength of 440nm and 675nm, similar as other articles had mentioned (Sathyendranath, Lazzara et al. 1987).
The absorption coefficient for suspended particulate matter (SPM) through wavelengths is computed according to the formula $a_{spm}(\lambda) = a_{spm}(400)\exp[-0.011(\lambda - 400)]$ as the curve below (Figure 3-3):

![Absorption coefficient for SPM](image)

**Figure 3-3** Absorption coefficient of SPM

The absorption coefficient for yellow matter (CDOM) from wavelength of 400 to 800nm, is shown as the figure below (Figure 3-4):

![Absorption coefficient for CDOM](image)

**Figure 3-4** Absorption coefficient for CDOM

For the backscattering coefficients of pure water, SPM and CHL, we plotted together in a figure below (Figure 3-5): (CDOM is not considered scattering the radiation),
For us to see how each of the three constituents affect the spectral reflectance above the water surface, by using the model developed in this chapter, we simulated the reflectance by varying only one concentration of the three one time (Figure 3-6, Figure 3-7, Figure 3-8).
Figure 3-7 Simulation 2: Changing the concentration of CHL

Figure 3-8 Simulation 3: Changing the concentration of CDOM

3.1.2. Validation for Forward Process

Theoretically, we have finished the forward process. We could give a simulation on spectral reflectance(s) above the water surface, when the concentration of SPM, CHL and CDOM is given. How well can it work? Does it simulate the real situation? We’ll find out in this section through 3 tests.

We obtained some spectral reflectance curve in the field through a radiometer, and we analyzed the constituents of lake water corresponding to those reflectance curves. Three groups of concentrations of the constituents are picked to validate the water reflectance model. The first group has a high SPM concentrations, the second group has a low SPM. The third group has a median value of concentration of SPM. We feed these data to the model we’ve built, and compare the 3 reflectance simulations with the real measurements. Table 6 shows the validation result for our light reflectance model.
MONITORING UNDERWATER LIGHT CLIMATE FOR AQUATIC VEGETATION IN POYANG LAKE BY REMOTE SENSING

Figure 3-9  Simulation test 1

Figure 3-10  Simulation test 2

Figure 3-11  Simulation test 3
In test 1, when the SPM concentration is high, the water reflectance model gives satisfactory simulation. The Root mean square error (RMSE) is only 0.0021, and the Pearson’s $R^2 = 0.97$. (Figure 3-9)

In test 2, when the SPM concentration is relatively low, the Water reflectance model doesn’t give prediction as good as it did in test 1, but still acceptable. The Root mean square error is 0.0299, and the $R^2 = 0.952$. Maybe it’s partly because noise occurred to the in-situ measurement for some reason. We can see noise occurring to the in-situ measurement curve (dotted one) after 750 nm and leads to some negative reflectance value which is unphysical and impossible. On the other hand, in the simulation, we could see a small absorption peak around 600 nm, and an absorption valley around 715, while the in-situ data doesn’t have such a peak and a valley. It’s possible that the CHL concentration from lab-work had been underestimated, because of the acidification process of chlorophyll a—the process turns chlorophyll a into a kind of acid so the CHL concentration decreased. (Figure 3-10)

In test 3, when the SPM concentration is not too large and not too small, the model performs well. The Root mean square error is only 0.0042, and the $R^2 = 0.952$. (Figure 3-11)

3.1.3. Conclusion for the Forward Process

From the results above, we can get the impression that our Water reflectance model gives satisfactory simulation on most cases, but not so well for relatively low SPM concentration. Our explanation is, when the SPM is not dominating the water so much, the reflectance affected by CHL become more important. However, we could not measure the CHL concentration accurately because of the acidification process of chlorophyll a. In this case, when the underestimated CHL concentration inputs to the model as a key parameter, the output of the model decreased its accuracy.

3.2. THE INVERSE PROCESS

3.2.1. Determining ANN Architecture for Optimal Water Constituent Retrieval

The architecture and the appropriate noise level as well as the above two error measures are given in the Table 12 in the appendix.

From the result displayed in the table, following conclusions can be drawn:
Regarding the retrieval of SPM concentration, the best result (the highest $r^2$ and the lowest RMSE) is case number 10 in the table, although all of the retrieval algorithms give satisfied performance. In this case (case 10), under the noise level of 2%, the validation $r^2$ value is 0.99, and the RMSE value is 1.89. The architecture of this ANN is 41-5-6-3, and the training method is BFGS quasi-Newton back-propagation.

Regarding the retrieval of pigment (CHL) concentration, the best performance (the highest $r^2$ and the lowest RMSE) is in case 13, while case 10 also gives satisfied results. In Case 13, under the noise level of 2%, the validation $r^2$ value is 0.97, and the RMSE value is 0.74. The architecture of this ANN is 41-10-4-3, and the training method is BFGS quasi-Newton back-propagation.

Regarding the retrieval of Yellow Matter (CDOM) concentration, the best result (the highest $r^2$ and the lowest RMSE) is given in case 13, while case 10 also gave satisfied results. In Case 13, under the noise level of 2%, the validation $r^2$ value is 0.98, and the RMSE value is 0.18. The architecture of this ANN is 41-10-4-3, and the training method is BFGS quasi-Newton back-propagation.

It has shown that the SPM concentration retrieval has the highest accuracy. Therefore, it’s the easiest one to retrieve. Between the retrieval of CHL and CDOM, there is no obvious evidence on which one is easier to retrieve. It all depends on algorithms.

For each of the retrieval of the three constituents, the noise level added to training dataset has a significant impact on the accuracy of the prediction only if the noise level is greater than 5%. That means, in certain degree, the ANN built for this particular application can tolerate an error smaller than 5% in its input data, so it has the feasibility to deal with the in-situ measurements which have the error less than 2%.

On one hand, the ANN in case 13 (architecture: 41-10-4-3) gives the best performance on CHL and CDOM retrieval, as we pointed out in conclusion 2 and 3. However, it spent 112 seconds to train the ANN using BFGS quasi-Newton back-propagation method. On the other hand, the ANN in case 10 (architecture: 41-3-6-3) gives the best prediction on SPM concentration, and second to the best prediction on both CHL and CDOM concentrations. Still, it only took 40 seconds to train this ANN using the same method. Therefore, when we evaluate the performance of the ANN-based constituents’ retrieval algorithms using in-situ data in the next section, we’ll pick these two ANN as candidates.

### 3.2.2. Evaluation the Performance of ANN-Based Inversion Algorithm

The candidates of the selected optimal ANNs (ANN in case 10 and 13) for the three constituent’s retrieval are assessed in three steps by applying them to 1) the “training data” in order to test if the ANN “learnt the model”; 2) the “validation data” in order to prevent the ANN from “overfitting”; 3) the “test data” which was collected in the field and have not been used for the ANN development in order to evaluate the ability of the ANN to retrieve concentrations based on real measurements.

The step 1 and 2 had been done in the last section. Here in the following tables, we highlight the performance with respect to the test data (in-situ) in the third step.

**Candidate 1: ANN (41-5-6-3):**
### Table 7: Evaluation the performance of ANN candidate 1

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Noise Level (%)</th>
<th>Training data for (N=400)</th>
<th>Validation data (N=200)</th>
<th>Test data for (N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>$r^2$</td>
<td>RMSE</td>
<td>$r^2$</td>
</tr>
<tr>
<td>SPM</td>
<td>2</td>
<td>1.37</td>
<td>0.99</td>
<td>1.89</td>
</tr>
<tr>
<td>CHL</td>
<td>2</td>
<td>0.82</td>
<td>0.96</td>
<td>0.88</td>
</tr>
<tr>
<td>CDOM</td>
<td>2</td>
<td>0.30</td>
<td>0.96</td>
<td>0.30</td>
</tr>
<tr>
<td>SPM</td>
<td>5</td>
<td>2.77</td>
<td>0.99</td>
<td>4.71</td>
</tr>
<tr>
<td>CHL</td>
<td>5</td>
<td>1.19</td>
<td>0.93</td>
<td>1.47</td>
</tr>
<tr>
<td>CDOM</td>
<td>5</td>
<td>0.33</td>
<td>0.95</td>
<td>0.30</td>
</tr>
<tr>
<td>SPM</td>
<td>10</td>
<td>3.55</td>
<td>0.99</td>
<td>10.07</td>
</tr>
<tr>
<td>CHL</td>
<td>10</td>
<td>2.27</td>
<td>0.77</td>
<td>2.75</td>
</tr>
<tr>
<td>CDOM</td>
<td>10</td>
<td>0.62</td>
<td>0.85</td>
<td>0.62</td>
</tr>
</tbody>
</table>

### Candidate 2: ANN (41-10-4-3):

### Table 8: Evaluation the performance of ANN candidate 2

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Noise Level (%)</th>
<th>Training data for (N=400)</th>
<th>Validation data (N=200)</th>
<th>Test data for (N=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>$r^2$</td>
<td>RMSE</td>
<td>$r^2$</td>
</tr>
<tr>
<td>SPM</td>
<td>2</td>
<td>1.37</td>
<td>0.99</td>
<td>2.45</td>
</tr>
<tr>
<td>CHL</td>
<td>2</td>
<td>0.65</td>
<td>0.97</td>
<td>0.74</td>
</tr>
<tr>
<td>CDOM</td>
<td>2</td>
<td>0.21</td>
<td>0.98</td>
<td>0.18</td>
</tr>
<tr>
<td>SPM</td>
<td>5</td>
<td>1.53</td>
<td>0.99</td>
<td>6.63</td>
</tr>
<tr>
<td>CHL</td>
<td>5</td>
<td>1.23</td>
<td>0.93</td>
<td>1.60</td>
</tr>
<tr>
<td>CDOM</td>
<td>5</td>
<td>0.49</td>
<td>0.91</td>
<td>0.61</td>
</tr>
</tbody>
</table>
From comparing the above tables (Table 7 & Table 8) of errors and correlation coefficients we could see, the ANN which has the architecture of (41-5-6-3) gives better performance on SPM, CHL and CDOM retrieval than the other one in most noise levels. Satisfactory performance is observed even through the “test data”, which is totally independent from the radiative transferring simulation. This is partly because the number and arrangement of the neurons in this network are just fine for this application: The ANN is large enough to solve the problem in reasonable accuracy, and compact enough to avoid over-fitting.

We should notice that this ANN (41-5-6-3) gives satisfied correlation coefficient on SPM and CHL retrieval, and relatively low on CDOM retrieval. The reasons behind will be discussed in chapter 4.

According to what we’ve just discussed, ANN (41-5-6-3) is chosen to retrieve concentrations of SPM, CHL, and CDOM, based on the Water reflectance model we’ve developed also in this study. The result of this evaluation for ANN (41-5-6-3) is shown in figures below for SPM, CHL and CDOM, respectively. (Figure 3-12, Figure 3-13, Figure 3-14)
Figure 3-12 Scatter plot showing the performance of the ANN based SPM retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set).
Figure 3-13  Scatter plot showing the performance of the ANN based CHL retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set).
Figure 3-14 Scatter plot showing the performance of the ANN based CDOM retrieval algorithms for the training data set, validation data set, and test data set (in-situ data set).
3.2.3. Conclusions for the Inverse Process

In this study, a methodology for the retrieval of three optically significant constituents from spectral reflectance has been developed. The artificial neural network (ANN) technique is applied to a set of spectral reflectance above the water surface which was generated from water reflectance model.

The ANN-based algorithm obtained through trainings can retrieve the concentrations of SPM, CHL and CDOM at the same time. This network has 4 layers: one input and one output layer, with two hidden layers in between of them. The input layer consists of 41 neurons, by which reflectances in different wavelength are inputted to the network. The first hidden layer contains 5 neurons and the second hidden layer has 6. This arrangement was considered the best architecture for retrieving SPM and CHL in the last section. The output layer contains 3 neurons, which provide the predictions for the three constituents all at once.

Applying the trained ANN-based algorithm to the test data which was collected in the field, the result for SPM retrieval is satisfactory. It gives a correlation of 0.871 (R=0.871) between predictions and the real values.

The result for CHL retrieval is OK too, the correlation between predicted value and the real measurements is 0.861 (R=0.861), however, the RMSE is relatively high (10.92 while the average concentration is 8.247). This is because our Water reflectance model which is developed in the forward process can not give an accurate reflectance series while the CHL concentration is low. It was the inborn fault of our Water reflectance model. Therefore, on the process of reversing it, the ANN gives negative values at some points which are impossible and unphysical.

And the ANN gives poor prediction on CDOM as we have expected (R=0.624). We have pointed out in the beginning of chapter 2, that when the suspended particulate matter dominates the water body, the spectral reflectance is also dominated by the characteristic spectrum of SPM, and it may overshadow the contribution of other constituents to the measured signal. In this study, CDOM concentration is relatively low (lower than 0.02mg/L compared with 400mg/L of SPM). As a result, the contribution of CDOM to the reflectance is overshadowed by SPM and its strong backscattering.

Another thing worth to mention is: an ANN with 41 neurons in its input layer is a little bit too much. We have performed a principal component analysis (PCA) to the inputs data first and selected only 7 most important components from within. After the PCA, the speed of training procedure is 4 times faster than before. But in order to keep the simplicity, in this thesis, we still use the ANN with 41 inputs.

3.3. UNDERWATER LIGHT INTENSITY MODEL

We developed the light intensity model in chapter 2 by which we are able to calculate the underwater light intensity by any given SPM concentration and water depth. The model is in the form of:
Where \( x \) is the SPM concentration, \( y \) is the water depth; \( z \) is the light intensity underwater. And the coefficient Data:

\[
a = -3.61e+009 \\
b = -3175064.3 \\
c = -0.063
\]

By this integrated formula, we can calculate the underwater light intensity at any depth, with any SPM concentration, as the Figure 3-15 below:

![Figure 3-15 Underwater light climate model](image)

In the figure above, X axis is the SPM concentration, with the unit of mg/L; Y axis it the water depth, with the unit of cm, Z axis is the light intensity, and the unit for light is mol/m²*s

We can see from this graph, the underwater light intensity is high, when the SPM concentration and the water depth are low, which means in the shallow and clear water, the vegetation in the bottom of lake can receive more sunlight.

To test the light intensity model, we compared the independent field data of light intensity and the prediction from the model, as figure shown below. The \( R^2=0.9604 \) and the root mean square error is 27.5827. The result shows the prediction is valid and with a high accuracy. (Figure 3-16)
Figure 3-16  Accuracy of predictions and measurements
4. APPLICATION TO THE TM IMAGERY (AN ATTEMPT)

4.1. INTRODUCTION

Based on the algorithms developed in this study, we could easily monitor the light climate underwater using a hyperspectral image. In this sense, the objective of this study had been achieved. However, in order to see how far this method can go, we attempted to apply it to TM satellite imageries, and hope we could generate a light climate map underwater with sufficient accuracy.

4.2. IMAGERY DATA

The TM imagery of Poyang lake area, China was taken at 2005-10-16 10:30 AM, the same day as the spectral reflectances above water surface were collected.

4.3. ATMOSPHERIC CORRECTION

Since we already had the hyper-spectral reflectance which was collected just above the water surface by a radiometer, the atmospheric correction is based on the ground measurements.

First, we transform the DN value in TM image to reflectance value using:

\[ L = DN \times \text{gain} + \text{bias} \]

Where \( L \) is the radiance of the object, DN is digital number in TM imagery. Gain and bias can be obtained in the head file of the image. (Chander, Held et al. 2004)

Then, we compute reflectance of objects from radiance:

\[ \rho = \frac{\pi d^2}{E_0 \cos \theta} \]

Where \( \rho \) is the reflectance of object, \( d \) is the astronomical unit, \( E_0 \) is the sun radiation at the top of atmosphere, \( \theta \) is the solar zenith angle which can be read in the head file of the imagery. Constant \( L \) for each band is listed in the Table 9 below:

<table>
<thead>
<tr>
<th>BAND</th>
<th>Band 1</th>
<th>Band 2</th>
<th>Band 3</th>
<th>Band 4</th>
<th>Band 5</th>
<th>Band 6</th>
<th>Band 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>L value</td>
<td>1969</td>
<td>1840</td>
<td>1551</td>
<td>1044</td>
<td>225.7</td>
<td>NULL</td>
<td>82.07</td>
</tr>
</tbody>
</table>

4.4. METHOD

Forward process

After reflectance correction, we use the hyperspectral reflectance data recorded by radiometer to simulate band1, band2, and band3 of TM data. Reflectance in certain wavelength were selected and averaged to simulate the corresponding bands. The Table 10 shows the wavelengths have been selected according to (TTE 2005):
Table 10  Wavelengths have been selected to simulate TM image

<table>
<thead>
<tr>
<th>Wavelength selected</th>
<th>Simulate for</th>
</tr>
</thead>
<tbody>
<tr>
<td>450-520nm</td>
<td>Band 1</td>
</tr>
<tr>
<td>520-600nm</td>
<td>Band 2</td>
</tr>
<tr>
<td>630-690nm</td>
<td>Band 3</td>
</tr>
</tbody>
</table>

Reverse process

In the reverse process, we also used the methods developed in the thesis; however, slight changes have been made to the ANN. The inputs of the ANN are adjusted to 3, instead of 41. And outputs are adjusted to 1. Because we only have 3 TM band simulations (red, green, blue), and only need one output (SPM).

For three reasons we limited the retrieval constituent of ANN to suspended particulate matter (SPM) only.

1. As we discussed before, when the SPM concentration is relatively high (in case II water), it can dominate the light intensity underwater.
2. When SPM dominated the water body, the remote sensing signal caused by backscattering of SPM overshadows the contribution of other constituents. For only 3 bands in TM image, it’s impractical to derive the signal contributed also by CHL and CDOM.
3. ANN couldn’t give outputs in principle equal or more than inputs with reasonable accuracy.

The ANN for the retrieval of SPM concentration in the study area is the previous part in this chapter. The architecture of this ANN is 3-5-6-1, and the training method is BFGS quasi-Newton back-propagation.

After the SPM concentration had been derived, with our light intensity model and water depth computed from DEM with water level, we could finally get the light climate map of the study area.

4.5. UNDERWATER LIGHT CLIMATE

Firstly, water depth map has been generated by knowing the water level of the sampling day (water level of Oct.16 is 12.1m above the sea level). (Figure 4-1)
We could see from the map that the bottom of the study lake is flat, vast area have the same water depth of 177-200cm. In the center of the lake, the elevation of the bottom even rises a little. Although the lake is shallow, when in the field, the bottom of the lake can’t be seen, because of the high concentration of the suspended matter.

By the retrieval methods we mentioned in Chapter 2, we’ve retrieved the SPM concentration map in the study lake (Figure 4-2). From this prediction map we could see the distribution of SPMs – the high concentration of SPM is distributed in the southeast part of the lake, and all the way to the southeast bank of the lake, while the relatively low concentration is spread in the west and north part of the lake. It’s easy to explain and it accords with the real situation well: the day before the sampling day (2005-10-15) was windy. The northwest wind which is prevailing in that season was blowing at the speed of 5m/s, from 19:30 to the morning of the next day (2005-10-16, 7:00). Because of the lake bank in the north and west side, the wind is not strong enough to disturb the water in the west and north part of the lake. On the contrary, the water in southeast part of the lake is disturbed strongly.
Although in large scale, the SPM retrieval fit the situation quite well, the retrieval cannot reach the expected accuracy. When we use the field data of SPM concentration to test the relationship between the real data and the prediction, it gives unacceptable result: the $R^2$ is only 0.31. The potential reason for that will be discussed in the next chapter.

Anyway, by now, we already have the water depth map and SPM concentration map (although with low accuracy). By applying the light climate model we developed in corresponding section, we are now able to map the light intensity of the study lake. (Figure 4-3)
In the light intensity map, we could see, the darker green represents that less sunlight can penetrate to the bottom; on the contrary, brighter green means more light can be received by the aquatic vegetation growing at the bottom. White color only exit in the fringe of the lake, because the water there is very shallow (<5cm) so the light can easily reaches to the bottom.

The light intensity at the bottom, as we mentioned before, is either affected by the water depth or by the water turbidity. In the middle of the lake, as we can tell from the map, the light intensity is a little stronger than the area around it. It’s because the elevation of the middle part of the lake is a little bit higher, so more sunlight can reaches to the bottom. And in the west side of the lake, the underwater light is also higher than the east side, that’s because of the relatively shallow and clear water.

For validation, in the field we attempt to get the light intensity at the bottom of the lake. However, the pity of this attempt is, the device for measuring light intensity can not give the value underwater when the light intensity is less than 1 mol/m²*m*s, while a large part of the lake bottom only have the light climate less than this level. Therefore we don’t have valid data to validate the result for the light intensity map retrieved by TM image. But anyway, the method we developed is validated by the hyper-spectral reflectance in Chapter 3, and gives satisfactory results.
5. DISCUSSION

5.1. GENERALLY DISCUSSION

In this study, we separated the whole modelling work into four steps:

1. Building of the water reflectance simulation model which can model the effect brings by the natural water and its constituents to the spectral remote sensing reflectance just above the water surface.
2. Training the selected ANN to learn the water reflectance simulation model and retrieve the concentrations of water constituents.
3. Build the empirical underwater light intensity model which can calculate the light intensity at the bottom of the lake by given SPM concentration and water depth.
4. Use the model developed in step 3, and give the inputs of SPM concentration retrieved from step2, plus the water depth derived from DEM, getting the underwater light climate for aquatic vegetation.

The discussion begins with the first step:

To build the water reflectance simulation model is in essence to find the relationship between the inherent optical properties and apparent optical properties. In case II water, the apparent optical properties are mainly caused by SPM which is the dominator of the water body. The leaving water radiance is largely influenced by the backscattering of SPM. However, in some wavelength, when the concentration of SPM is not very high, other optically significant water constituents also show their contributions (like in 690nm, the CHL shows its absorption peak).

By the forward process, we’ve found the different water constituents affect the IOP of water in different ways. The variation on concentration of SPM causes absorption coefficient of water most significantly changes in the wavelength of 670 to 690. Therefore, if we use only one band of image to retrieve SPM concentration by empirical method, the band selection should in the range of 690 to 720.

With the concentration of SPM become higher, the sensitivity of the absorption coefficient of SPM becomes lower. When the concentration is higher than 700mg/L, the IOPs doesn’t change much. That means if the concentration of SPM is extremely high, it becomes very hard for us to retrieve it. So, 700mg/L should be the upper limit for retrieving SPM concentration.

And we can tell from the figure which shows us the absorption coefficient. One interesting thing is that with the concentration of CHL increase, the absorption peak in 670nm caused by CHL moves to larger wavelength. That means we couldn’t use a single band for CHL retrieval by empirical method.
In the inverse process, it has been shown that the SPM concentration is the easiest constituents to be retrieved among the others. It can be retrieved by the ANN with only one hidden layers with 10 neurons on it with high accuracy.

When we validate the performance of the ANN-based algorism, we found the correlation coefficients between the ANN prediction and the test data is notably lower than the correlation coefficients between the ANN prediction and the valid data. For the ANN we finally used (41-5-6-3), with the noise level of 2%, the two numbers are 0.75 and 0.99 for SPM concentration. This does not mean that the ANN doesn’t fit the model well. Oppositely, the ANN fit the water reflectance model quite well ($r^2 = 0.99$). It means that the water reflectance model we built in step one doesn’t fit the in-situ measurements precisely, because of the complexity of the nature. Based on this factor, we can predict the ANN method has the potential to perform better, only if we improve the algorithm of the water reflectance model.

From the retrieval process, we can also tell the SPM and CDOM are retrieved with lower accuracy than that of SPM. This is because the concentrations of the CHL and CDOM are so much lower than that of SPM. For:

$$A = a_w + C_s a_s + C_c a_c + C_y a_y$$

$$B = b_w + C_s b_s + C_c b_c$$

If the concentration of one constituents, say $x$, are low, the term $C_s a_s$ or $C_s b_s$ is small. Which means the absorption or backscattering caused by $x$ is just a very small part of the total contributions. Therefore, the contribution by $x$, (just like CHL and CDOM) can be easily overshadowed by other constituents (like SPM). If we want to have all of the three constituents retrieved with the same accuracy, we must keep, or rather say, find a water area in which: $C_s a_s \approx C_c a_c \approx C_y a_y$.

In step 3, we have used an empirical regression for the prediction of underwater light intensity. The prediction shows: when the water depth is greater than 70 cm, even without SPM, less than 10 mol/m²m*s light photons can reach to the bottom. We know this can’t happen in a clear lake or in the ocean where the light can penetrate deeply. That’s because when we building the model, we assumed that only SPM changes in concentration has an impact on water transparency, while other constituents remain unchanged in their concentrations. So this model is suitable for applying it to Case II waters, and it can perform well while the SPM concentration is relatively high.

One way to solve this problem is to solve the radiative transfer equation and build a semi-physical or pure physical model based on it (Hoge, Lyon et al. 2003; Nalli and Smith 2003). By this method, the light underwater can be calculated accurately. Unfortunately, the radiative transfer equation is so complicated that only approximate solution can be derived by some methods (Fell and Fischer 2001; Sathyendranath, Cota et al. 2001; Alton, North et al. 2005). But if we need the universality of the light intensity model, the using of physical or semi-physical model is a trend. If we work on the turbid water or the SPM concentration doesn’t varies much, the empirical model is enough to predict the light climate at the bottom of lakes.

The last thing is, in the attempt of computing light climate underwater in Dahuchi, a part of the Poyang Lake, the accuracy for sediment retrieval is poor. Why our model which gives excellent
accuracy for hyper-spectral application does not function well when works with TM image? The possible reasons for that are discussed below:

The main reason for that is: reflectances calculated using TM image doesn’t fit well with the reflectance measured by radiometer in any wavelength or average of wavelengths. After the TM image has been atmospheric corrected, the correlation between the radiometer measurements and TM image in Red band, Green band, and Blue band is only 0.736, 0.521 and 0.704, respectively. In other words, the large errors are due to the imprecise atmospheric correction.

Another reason for that is the glitter, which means the downwelling sunlight reflected by level water surface as a mirror, and changes the direction to the reflected ray path, where the remote sensor is coincidentally in the path. The reflected ray brings no information about water body, and they are noises strong enough to overshadow the water leaving radiance which contains information on the water constituents.

Even the atmospheric correction have been performed carefully, and no glitter in the TM image, we still can not extract all three water constituents from only 3 useful band (RGB, from 400nm-800nm) by our model. Because the bands in TM are too broad, much useful information are lost. For example, the absorption peak around 680nm can’t be discerned in band 3 (630nm – 690nm). If the DN value in band 3 increases, we can’t tell it’s because of the CHL concentration decreasing or the SPM concentration increasing. Put it in another way, the TM image does not contain enough information for us to retrieve 3 constituents at once.

5.2. ASSUMPTIONS AND LIMITATIONS IN THE MODEL

5.2.1. Water reflectance model

The assumptions in the water reflectance model:

1. The lake is deep enough or turbid enough so the bottom reflectance doesn’t dominate the water leaving reflectance.
2. The absorption coefficient in 400nm \( a_{SPM}(400) \) is around 0.1, as the typical absorption coefficient of SPM in lake water.
3. The species of phytoplankton in the study lake and its absorption coefficient is similar to the one studied in the French project “EUMELI”.

The limitation of the water reflectance model:

1. The input SPM concentration is in the range of (50-1000mg/L)
2. The water surface is considered level. So it can only be applied to lake water with the wind speed less than 3m/s.

5.2.2. ANN retrieval model

The assumptions in the ANN retrieval model:

1. The retrieved concentrations of constituents should be in the range of (Table 11):
Table 11  Range of retrieved concentrations

<table>
<thead>
<tr>
<th>Components</th>
<th>Unit of Concentration</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPM</td>
<td>Mg/m$^3$</td>
<td>1</td>
<td>1000</td>
</tr>
<tr>
<td>CHL</td>
<td>g/m$^3$</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>CDOM</td>
<td>Mg/m$^3$</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

2. The noise level of spectral reflectance above the water surface is less than 5%.

The limitation of the ANN retrieval model:

1. The ANN retrieval model gives relatively low precision when the concentrations are low, especially for CDOM and CHL.
2. The model can not retrieval the constituents whose backscattering are optically overshadowed by other constituents’.
3. The architecture of the model maybe need to be changed when applying the model to other water areas.

5.2.3. Light intensity model

The assumptions in the light intensity model:

1. The light intensity is greatly depended by the concentration of SPM in case II water, which is by definition dominated by suspended particulate matter. In this case, we can ignore the variance of CHL and CDOM because very little contribution is caused by them.
2. The intensity of the upwelling light radiance above the water surface is around 2000 mol/m$^2$m*s. That is, the weather of the area is sunny and the sun angle is around 90 degree.

The limitation of the light intensity model:

1. This is a light intensity model just for the case II water. Because it is regressed from the data of case II water, so the attenuation function by average concentration of CHL and CDOM is already took into account by the model.
2. The light intensity model gives better prediction when the SPM concentration is not too low (lower than 10mg/L).
6. CONCLUSION

6.1. CONCLUSION

In this thesis, a methodology for monitoring the underwater light intensity in case II water by remote sensing approaches has been derived. The work is in three steps: first, a semi-physical model (water reflectance model) has been built to generate spectral reflectance simulations by given concentrations of water constituents as inputs. Second, a retrieval method is derived by applying ANN technique to a large set of spectra simulations which are generated by the first step. Third, a light intensity model (which developed also in this thesis) applied to the retrieved concentrations, giving the result of underwater light intensity. The results show that this method is a promising technique for the retrieval of water constituents, and also give satisfactory performance on monitoring the light climate underwater by inputs of hyper-spectral remote sensing reflectance. The biggest advantage of this method is that we don’t need a large sample set to train the ANN for retrieving constituents. All of the water samples we took and analyzed are just used for validation and test rather than build a regression model. In this way, we can save a lot of money and time. The conclusions may be summarized as follows:

1. The method is successfully used to monitoring the underwater light climate in lake water (case II water).
2. A semi-physical model of the IOP in case II water is proposed in order to simulate enough training data for the development of the ANN based retrieval method.
3. An ANN based method is applied successfully to derive the concentrations of suspended particulate matter, pigment and yellow matter from spectral remote sensing reflectance just above the surface of lake water.
4. An underwater light intensity model was developed to calculate the light intensity at any depth of water, with any SPM concentrations (in the range of 0-700mg/L).
5. The method is applied to monitor the light climate in study area as an attempt. Reasonable results are obtained. However, more works are required to modify and validate the particular application.

6.2. FUTURE WORK

(1) By knowing more about the inherent optical property and radiative transfer theory, the Water reflectance of Case II water should be updated, and then the model should provide more accurate result and fit the real measurements better.

(2) With more information about the species of phytoplankton in the study area, the prediction of CHL concentration could be more reliable.

(3) In this study, one pity is we limited by satellite data source. Applying the ANN based retrieval model to hyperspectral imagery can greatly increase the accuracy for concentration retrieval and light climate prediction.
(4) The light intensity model we developed in the end of chapter 2 will be widely validated in different areas by in-situ measurements. But it fits the measurements in the study area really good.

(5) We could also take the effects contributed by pigment and yellow matter into account to the light intensity model, when the SPM does not dominate the water body. More accurate result may be achieved.
REFERENCES


APPENDIX

1. Code

Part one: Forward process
spectral remote sensing reflectance above the water surface generator

clear all;
load realmeasure;
%====================water-air transmition rate====================
waterair_transmit=0.96;
SAMPLNO=400;
Noiselevel=0.02;
% Concentration of 3 matter...Initial value
%==============Randomly select initial value for concentrations=============
concen_SPM=rand(SAMPLNO,1)*700;
concen_YM=rand(SAMPLNO,1)*10;
concen_CHL=rand(SAMPLNO,1)*30;
%====================......Now the constant value....====================
abs.WATER= [0.0171 0.0162 0.0153 0.0144 0.0145 ...
0.0145 0.0156 0.0176 0.0196 ...
0.0257 0.0357 0.0477 0.0507 0.0558...]
abs.YM=
[0.52520175 0.456588467 0.396938944 0.34508214 0.3 ...
0.26087471 0.226735122 0.197114046 0.171362719 0.14897591 ...
0.129513157 0.11259333 0.097883938 0.085096208 0.073979089 ...
0.06431433 0.055912193 0.048607725 0.04257526 0.03673628 ...
0.031937551 0.027765173 0.024137882 0.020984467 0.018243019 ...
0.015859719 0.013787777 0.011986517 0.010420578 0.009059215 ...
0.007875703 0.006846807 0.005952328 0.005174076 0.004496873 ...
0.003910958 0.003400024 0.002981973 0.002671353 0.002393089 ...
0.002143811 0.001920499 0.001720449 0.001541237 0.001380692 ...
0.001236873 0.001108032 0.000992613 0.000889216 0.00079659 ...
0.000713613 0.000639279 0.000572688 0.000513033 0.000459593 ...
0.000411719 ...]
abs.CHL=
[0.0252 0.0293 0.0332 0.0368 0.0402...
0.039 0.0383 0.0388 0.038 0.0359...]
abs.SPM=
[0.01 0.008958341 0.008025188 0.007189237 0.006440364...]
0.005769498 0.005168513 0.004630131 0.00417829 0.00371567...
0.003328771 0.002981973 0.002671353 0.002393089 0.002143811...
0.001920499 0.001720449 0.001541237 0.001380692 0.001236873...
0.001108032 0.000992613 0.000889216 0.00079659 0.000713613...
0.000639279 0.000572688 0.000513033 0.000459593 0.000411719...
0.00038832 0.000304012 0.000295994 0.000265162 0.000237541...
0.000212797 0.000190631 0.000170774 0.000152985 0.000137049 ...
0.000122773];
% The absorption coefficient of Yellow Matter from wavelength 400-800, interval of 10.(400,410,420...800)

% The absorption coefficient of Suspended Particular Matter
% from wavelength 400-800, interval of 10.(400,410,420...800)

% The absorption coefficient of Pigment caused by
%CHL from wavelength 400-800, interval of 10.(400,410,420...800)
backsc.WATER={0.015878535 0.014271951 0.012860939 0.011617853 0.010519474...
0.009546222 0.00868152 0.007911282 0.007223496 0.006607886...
0.006055628 0.005559125 0.005111815 0.004708015 0.004342787...
0.004011832 0.003711396 0.003438193 0.003189339 0.002962299...
0.002754839 0.002564984 0.002390989 0.002231302 0.002084549...
0.001949503 0.001825071 0.001710277 0.001604246 0.001506196...
0.001415422 0.001331293 0.001253237 0.001180742 0.001113342...
0.001050618 0.00099219 0.000937712 0.000886872 0.000839384...
0.000794989};
% The backscattering coefficient* for pure water from wavelength 400-800, interval of 10.(400,410,420...800) * volume sca-coefficient * backscattering ratio

backsca.SPM={0.011283645 0.011172745 0.011065568 0. 010961905 0.010861564...
0.010764385 0.010670144 0.010578748 0.010490035 0.010403872...
0.010320136 0.010235403 0.010159494 0.01008238 0.010007277...
0.009934096 0.009862754 0.009793174 0.009725282 0.009659019...
0.009594292 0.009531066 0.009469275 0.009408864 0.009349781...
0.009291976 0.009238712 0.009180017 0.009125777 0.009072642...
0.009020575 0.008969538 0.008919498 0.008870422 0.008822278...
0.008775036 0.008728668 0.008683146 0.008638445 0.008594539...
0.008551404};
% The backscattering coefficient for SPM from wavelength 400-800, interval of 10.(400,410,420...800)

backsca.CHL={0.005547591 0.005452525 0.005361322 0. 005273737 0.005189548...
0.00510855 0.005030555 0.004952917 0.004882897 0.004812926...
0.004745341 0.004680015 0.00461832 0.004555681 0.004496946...
0.004439075 0.004383437 0.004329463 0.004277074 0.004226199...
0.00417677 0.004128721 0.004081993 0.004036528 0.003992725...
0.003941918 0.00389702 0.003856287 0.003826398 0.003787495...
0.003749538 0.003712492 0.003676323 0.003640998 0.003606486...
0.003572757 0.003539795 0.003507542 0.003476003 0.003445144...
0.003414943};
% The backscattering coefficient for CHL from wavelength 400-800, interval of 10.(400,410,420...800)

Part two: Inverse process
Training the ANN to retrieve the water constituents

```matlab
% t=[concen_SPM concen_CHL concen_YM];
% netold=newf([minmax(reflect_cuv),[12,3],{'tansig','purelin','trainlm'});
% netold=init(netold);

net=newf([minmax(training.P),[5,6,3],{'tansig','purelin','purelin'},'trainbfg']);
net.trainParam.show = 20;
net.trainParam.epochs = 2000;
net.trainParam.goal=0.05;

%====== smooth ========
net.performFcn = 'msereg';
net.performParam.ratio = 0.5;

%=================================
net = init(net);
[net,tr1]=train(net,training.P,training.T,[],[],[],test);
%[net,tr1]=train(net,reflect_cuv,t);

%========Principle component analysis. Preprocessing=============
[TrainingPnorm,meanp,stdp] = prestd(training.P);
[ptrans,transMat] = prepca(TrainingPnorm,0.001);

% 0.001; This means that prepca eliminates those principal components
% that contribute less than 0.1% to the total variation in the data set.

%--------when inputting------
pnewn = trasd(training.P,meanp,stdp);
pnewtrans = trapca(pnewn,transMat);
a = sim(net,pnewtrans);

%========Mean and Stand. Dev. Preprocessing=======
[pn,meanp,stdp] = prestd(reflect_cuv);
[ptrans,transMat] = prepca(pn,0.0001);
[pn,meanp,stdp,tn,meant,stdt] = prestd(training.P,training.T);

% see the test result
predict=sim(net,test.P);
predspm=predict(1,:);
plot(test.T(1,:),predspm,+)

% post trained
predict=sim(net,test.P);
predspm=predict(1,:);
RMSE=mean(std([predspm;test.T(1,:)]));
[m,b,r] = postreg(predspm,test.T(1,:));
correlation=r;

Part three: Validation
net.trainParam.show = 20;
net.trainParam.epochs = 1200;
net.trainParam.goal=0.1;
net.performFcn = 'msereg';
net.performParamRatio = 0.5;

net = init(net);
[net,tr1]=train(net,training.P,training.T,[],[],[],test);

predict=sim(net,test.P);
predchl=predict(2,:);
predcdom=predict(3,:);
vRMSEspm=mean(std([predspm;test.T(1,:)]));
vRMSEchl=mean(std([predchl;test.T(2,:)]));
vRMSEcdom=mean(std([predcdom;test.T(3,:)]));

%=========Graph==========
subplot(2,2,1);
[mspm,bspm,rspm] = postreg(predspm,test.T(1,:));
subplot(2,2,2);
[mchl,bchl,rchl] = postreg(predchl,test.T(2,:));
subplot(2,2,3);
[mcdom,bcdom,rcdom] = postreg(predcdom,test.T(3,:));
correlation_spm=rspm;
```
correlation_chl=rchl;
correlation_cdom=rcdom;
vRMSEspm;
vRMSEchl;
vRMSEcdom;
R2spm=correlation_spm*correlation_spm;
R2chl=correlation_chl*correlation_chl;
R2cdom=correlation_cdom*correlation_cdom;
%-------------------------- training data-------------
predict=sim(net,training.P);
predspm=predict(1,:);
predchl=predict(2,:);
predcdom=predict(3,:);
RMSEspm=mean(std([predspm;training.T(1,:)]));
RMSEchl=mean(std([predchl;training.T(2,:)]));
RMSEcdom=mean(std([predcdom;training.T(3,:)]));

subplot(2,2,1);
[mspm,bspm,rspm] = postreg(predspm,training.T(1,:));
subplot(2,2,2);
[mchl,bchl,rchl] = postreg(predchl,training.T(2,:));
subplot(2,2,3);
[mcdom,bcdom,rcdom] = postreg(predcdom,training.T(3,:));
correlation_spm=rspm;
correlation_chl=rchl;
correlation_cdom=rcdom;

training_RMSE_SPM=RMSEspm;
training_RMSE_CHL=RMSEchl;
training_RMSE_CDOM=RMSEcdom;
training_R2spm=correlation_spm*correlation_spm;
training_R2chl=correlation_chl*correlation_chl;
training_R2cdom=correlation_cdom*correlation_cdom;
result=[training_RMSE_SPM training_R2spm vRMSEspm R2spm training_RMSE_CHL training_R2chl ...
vRMSEchl R2chl training_RMSE_CDOM training_R2cdom vRMSEcdom R2cdom]

%-------------Real data-------------
predict=sim(net,real.P);
predspm=predict(1,:);
predchl=predict(2,:);
predcdom=predict(3,:);
RMSEspm=mean(std([predspm;real.T(1,:)]));
RMSEchl=mean(std([predchl;real.T(2,:)]));
RMSEcdom=mean(std([predcdom;real.T(3,:)]));
subplot(2,2,1);
[mspm,bspm,rspm] = postreg(predspm,real.T(1,:));
subplot(2,2,2);
[mchl,bchl,rchl] = postreg(predchl,real.T(2,:));
subplot(2,2,3);
[mcdom,bcdom,rcdom] = postreg(predcdom,real.T(3,:));
correlation_spm=rspm;
correlation_chl=rchl;
correlation_cdom=rcdom;

real_RMSE_SPM=RMSEspm
real_RMSE_CHL=RMSEchl
real_RMSE_CDOM=RMSEcdom
real_R2spm=correlation_spm*correlation_spm
real_R2chl=correlation_chl*correlation_chl
real_R2cdom=correlation_cdom*correlation_cdom

Part four: Light intensity model

xspm=1000;
ydepth=100;
x=[0:10:xspm];y=[0:2:ydepth];
xbig=ones(ydepth/2+1,1)*x;
ybig=(ones(xspm/10+1,1)*y)';
mesh(xbig,ybig, (1132.29*exp(-0.0632*ybig)*863.12*exp(-0.00276*xbig))/(863.12*exp(-0.00276*150)))
## 2. Tables

Table 12 Determining ANN Architecture for Optimal Water Constituent Retrieval

<table>
<thead>
<tr>
<th>Case No</th>
<th>Architecture (input-hidden-output) (%)</th>
<th>Noise Level (%)</th>
<th>Training data for SPM (N=400)</th>
<th>Validation data for SPM (N=200)</th>
<th>Training data for CHL (N=400)</th>
<th>Validation data for CHL (N=200)</th>
<th>Training data for CDOM (N=400)</th>
<th>Validation data for CDOM (N=200)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td>$r^2$</td>
<td>RMSE</td>
<td>$r^2$</td>
<td>RMSE</td>
<td>$r^2$</td>
</tr>
<tr>
<td>1</td>
<td>41-10-3</td>
<td>2</td>
<td>2.58</td>
<td>0.99</td>
<td>2.88</td>
<td>0.99</td>
<td>1.86</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>41-10-3</td>
<td>5</td>
<td>3.15</td>
<td>0.99</td>
<td>4.99</td>
<td>0.99</td>
<td>2.12</td>
<td>0.80</td>
</tr>
<tr>
<td>3</td>
<td>41-10-3</td>
<td>10</td>
<td>4.47</td>
<td>0.99</td>
<td>8.67</td>
<td>0.98</td>
<td>2.97</td>
<td>0.63</td>
</tr>
<tr>
<td>4</td>
<td>41-15-3</td>
<td>2</td>
<td>3.41</td>
<td>0.99</td>
<td>8.08</td>
<td>0.99</td>
<td>2.55</td>
<td>0.71</td>
</tr>
<tr>
<td>5</td>
<td>41-15-3</td>
<td>5</td>
<td>2.89</td>
<td>0.99</td>
<td>5.96</td>
<td>0.99</td>
<td>1.80</td>
<td>0.85</td>
</tr>
<tr>
<td>6</td>
<td>41-15-3</td>
<td>10</td>
<td>3.08</td>
<td>0.99</td>
<td>6.20</td>
<td>0.99</td>
<td>1.80</td>
<td>0.85</td>
</tr>
<tr>
<td>7</td>
<td>41-3-6-3</td>
<td>2</td>
<td>1.82</td>
<td>0.99</td>
<td>2.27</td>
<td>0.99</td>
<td>1.29</td>
<td>0.92</td>
</tr>
<tr>
<td>8</td>
<td>41-3-6-3</td>
<td>5</td>
<td>3.05</td>
<td>0.99</td>
<td>4.63</td>
<td>0.99</td>
<td>1.89</td>
<td>0.84</td>
</tr>
<tr>
<td>9</td>
<td>41-3-6-3</td>
<td>10</td>
<td>5.42</td>
<td>0.99</td>
<td>9.41</td>
<td>0.98</td>
<td>4.80</td>
<td>0.10</td>
</tr>
<tr>
<td>10</td>
<td>41-5-6-3</td>
<td>2</td>
<td>1.37</td>
<td>0.99</td>
<td>1.89</td>
<td>0.99</td>
<td>0.82</td>
<td>0.96</td>
</tr>
<tr>
<td>11</td>
<td>41-5-6-3</td>
<td>5</td>
<td>2.77</td>
<td>0.99</td>
<td>4.71</td>
<td>0.99</td>
<td>1.19</td>
<td>0.93</td>
</tr>
<tr>
<td>12</td>
<td>41-5-6-3</td>
<td>10</td>
<td>3.55</td>
<td>0.99</td>
<td>10.07</td>
<td>0.98</td>
<td>2.27</td>
<td>0.77</td>
</tr>
<tr>
<td>13</td>
<td>41-10-4-3</td>
<td>2</td>
<td>1.37</td>
<td>0.99</td>
<td>2.45</td>
<td>0.99</td>
<td>0.65</td>
<td>0.97</td>
</tr>
<tr>
<td>14</td>
<td>41-10-4-3</td>
<td>5</td>
<td>1.53</td>
<td>0.99</td>
<td>6.63</td>
<td>0.99</td>
<td>1.23</td>
<td>0.93</td>
</tr>
<tr>
<td>15</td>
<td>41-10-4-3</td>
<td>10</td>
<td>1.08</td>
<td>0.99</td>
<td>9.71</td>
<td>0.98</td>
<td>2.30</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- RMSE: Root Mean Square Error
- $r^2$: Coefficient of Determination

**Note:** The table above presents the results of different cases for varying ANN architectures and noise levels, showing the performance metrics for training and validation datasets for SPM, CHL, and CDOM.