DETERMINATION OF CHLOROPHYLL-A DISTRIBUTION IN LAKE EYMIR USING REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS WITH HYBRID INPUTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ONUR YÜZÜGÜLLÜ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ENVIRONEMENTAL ENGINEERING

JANUARY 2011

Approval of the thesis:

DETERMINATION OF CHLOROPHYLL-A DISTRIBUTION IN LAKE EYMIR USING REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS WITH HYBRID INPUTS

submitted by **ONUR YÜZÜGÜLLÜ** in the partial fulfillment of the requirements for the degree of **Master of Science in Environmental Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Göksel N. DEMİRER _____
Head of Department, **Environmental Engineering**

Assoc. Prof. Dr. Ayşegül AKSOY _____
Supervisor, **Environmental Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Celal. F. Gökçay _____
Environmental Engineering, METU

Assoc. Prof. Dr. Ayşegül Aksoy _____
Environmental Engineering, METU

Prof. Dr. Can Ayday _____
Satellite and Space Sciences Res. Inst., Anadolu University

Dr. Emre Alp _____
Environmental Engineering, METU

Dr. Barış Kaymak _____
Environmental Engineering, METU

Date : 6 January 2011

I hereby declare that all the information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name, Last Name: Onur YÜZÜGÜLLÜ

Signature:

ABSTRACT

DETERMINATION OF CHLOROPHYLL-A DISTRIBUTION IN LAKE EYMIR USING REGRESSION AND ARTIFICIAL NEURAL NETWORK MODELS WITH HYBRID INPUTS

YÜZÜGÜLLÜ, Onur
M.Sc., Department of Environmental Engineering
Supervisor: Assoc. Prof. Dr. Ayşegül AKSOY

January 2011, 152 pages

Chlorophyll-a is a parameter which can be used to understand the trophic state of water bodies. Therefore, monitoring of this parameter is required. Yet, distribution of chlorophyll-a in water bodies is not homogeneous and exhibits both spatial and temporal variations. Therefore, frequent sampling and high sample sizes are needed for the determination of chlorophyll-a quantities. This would in return increase the sampling costs and labor requirement, especially if the topography makes the location hard to reach. Remote sensing is a technology that can aid in handling of these difficulties and obtain a continuous distribution of chlorophyll-a concentrations in a water body. In this method, reflectance from water bodies in different wavelengths is used to quantify the chlorophyll-a concentrations. In previous studies in literature, empirical regression models that use the reflectance values in different bands in different combinations have been derived. Yet, prediction performances of these models decline especially in shallow lakes. In this

study, the spatial distribution of chlorophyll-a in shallow Lake Eymir is determined using both regression models and artificial neural network models that use hybrid inputs. Unlike the models generated before, field measured parameters which can influence the reflectance values in remotely sensed images have been used in addition to the reflectance values. The parameters that are considered other than reflectance values are photosynthetically active radiation (PAR), secchi depth (SD), water column depth, turbidity, dissolved oxygen concentration (DO), pH, total suspended solids (TSS), total dissolved organic matter (TDOM), water and air temperatures, wind data and humidity. Reflectance values are obtained from QuickBird and World View 2 satellite images. Effect of using hybrid input in mapping the reflectance values to chlorophyll-a concentrations are studied. In the context of this study, three different high-resolution satellite images are analyzed for the spatial distribution of chlorophyll-a concentration in Lake Eymir. Field and laboratory studies are conducted for the measurement of parameters other than the reflectance values. Principle component analysis is applied on the collected data to decrease the number of model input parameters. Then, linear and non-linear regression and artificial neural network (ANN) models are derived to model the chlorophyll-a concentrations in Lake Eymir. Results indicate that ANN model shows better predictability compared to regression models. The predictability of ANN model increases with increasing variation in the dataset. Finally, it is seen that in determination of chlorophyll-a concentrations using remotely sensed data, models with hybrid inputs are superior compared to ones that use only remotely sensed reflectance values.

Keywords: Lake Eymir, regression models, artificial neural network, remote sensing, chlorophyll-a

ÖZ

HİBRİD GİRDİLİ REGRASYON VE YAPAY SİNİR AĞLARI MODELLERİ İLE EYMİR GÖLÜNDE KLOROFİL-A DAĞILIMININ BELİRLENMESİ

YÜZÜGÜLLÜ, Onur
Y. Lisans, Çevre Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Ayşegül AKSOY

Ocak 2011, 152 sayfa

Klorofil-a, trofik derecenin belirlenmesinde kullanılabilen bir parametredir. Bu nedenle, klorofil-a konsantrasyonlarının izlenmesi gerekmektedir. Ayrıca, klorofil-a 'nın su kütlelerindeki heterojen dağılımı mekansal ve zamansal değişimler göstermektedir. Bu sebeple klorofil-a konsantrasyonlarındaki değişimlerin detaylı ve sıklıkla izlenmesi gerekmektedir. Sıklıkla tekrarlanan ölçümler yüksek maliyet ve iş gücü gerektirmektedir. Özellikle topoğrafya'nın ulaşımı zorlaştırdığı alanlarda örnekleme daha da zorlaşmaktadır. Uzaktan algılama ise bunun gibi çalışmalarda klorofil-a konsantrasyonunun su kütlelerinde sürekli dağılımının elde edilmesine olanak sağlar. Bu yöntem dahilinde farklı dalga boylarındaki yansıma değerlerinin kullanılarak klorofil-a konsantrasyonuna ulaşılabilir. Daha önceki çalışmalar göze alındığında deneysel modellerin kullanılmış olduğu görülmektedir. Deneysel modellerde farklı bantlardaki yansıma değerlerinin farklı kombinasyonları ile çalışılmıştır. Ancak, önceki çalışmalarda model performansının sığ göllerde düşmüş olduğu gözlenmiştir. Bu çalışmada sığ bir göl olan Eymir Gölü'nde klorofil-a

konsantrasyonun mekansal dağılımı incelenmiştir. Çalışmada gerçekleştirilen modellemede doğrusal ve doğrusal olmayan modellere ek olarak yapay sinir ağları modeli de kullanılmıştır. Kullanılan modellerde önceki çalışmalardan farklı olarak hybrid girdi kullanılmıştır. Bu çalışmada klorofil-a konsantrasyonunun belirlenmesi için uygulanmış olan modellerde yansıma değerlerine ek olarak, yansıma değerlerini değiştirebilecek saha parametreleri de eklenmiştir. Bu parametreler fotosentez için aktif ışık (FAI), secchi derinliği (SD), su derinliği, bulanıklık, çözünmüş oksijen (ÇÖ) konsantrasyonu, pH, toplam askıda katı madde (AKM), toplam çözünmüş organic madde (TÇOM), hava ve su sıcaklıkları, rüzgar hızı ve nemdir. Yansıma değerleri Quickbird ve Worldview 2 uydularının görüntülerinden elde edilmiştir. Çalışmada ayrıca hybrid girdiler yansıma değerleri ile birlikte kullanılarak klorofil-a dağılımının haritalandırılması çalışılmıştır. Çalışma kapsamında Eymir Gölü'ndeki klorofil-a dağılımının belirlenmesi amacıyla üç adet yüksek çözünürlüklü uydu görüntüsü kullanılmıştır. Yansıma değerlerine ek olarak diğer parametrelerin ölçülmesi amacıyla saha ve laboratuvar çalışmaları gerçekleştirilmiştir. Ayrıca parameter sayısını azaltabilmek için veri üzerinde temel bileşen analizi uygulanmıştır. Bir sonraki basamakta ise Eymir Gölü'ndeki klorofil-a konsantrasyonu regrasyon ve yapay sinir ağları modelleri ile modellenmiştir. Çalışmanın sonuçları göstermiştir ki yapay sinir ağları modeli, regrasyon modellerine oranla daha iyi bir tahmin kapasitesine sahiptir. Ayrıca söz konusu tahmin kapasitesi veri kümesindeki sapma arttıkça yükselmektedir. Son olarak, hibrid modellerin sadece tayfsal yansımadan oluşan modellerden üstün olduğu görülmüştür.

Anahtar Kelimeler: Eymir Gölü, regresyon modelleri, yapay sinir ağları, uzaktan algılama, klorofil-a

To my parents

ACKNOWLEDGEMENTS

The author would like to express his sincere gratitude to his thesis supervisor Assoc. Prof. Dr. Ayşegül Aksoy for her unlimited support, valuable criticism, and endless and forever patience throughout this study.

The author would also thank to committee members Prof. Dr. Celal F. Gökçay, Prof. Dr. Can Ayday, Dr. Emre Alp and Dr. Barış Kaymak for their precious suggestions and contributions to this study.

The author gratefully acknowledges the financial support from TUBITAK ÇAYDAG (Project Code: 109Y201) and METU Scientific Research Projects Fund (Project Code: 07-02-2010-00-01)

The author would like to thank to Mr. Tolga Pilevneli for his direct support in all steps of this research.

The author would like to thank to his colleagues Miss Firdes Yenilmez, Mr. Murat Varol, Miss Seçil Ömeroğlu, Miss Elif Küçük and Miss Gizem Naz Çalışkan for their support while carrying out site measurements and laboratory analysis.

The author would like to thank to Mr. Hakan Moral for his technical assistance throughout the study.

The author would like to thank to Miss Yücel Erbay, Mr. Tuluhan Sipka, Mr. Cengiz Doğangönül, Mr. Tolga Özbilge and Mr. Tolga Alkevli for their support in analyzing satellite imageries.

The author would like extend his special thanks to his friends Miss Gönül Hızalan, Miss Zuhal Çam, Miss Gülce Akbaş, Miss Sevinç Tunçağıl, Miss Hande Bozkurt and Mr. Sinan Aşçı and Mr. Kerem Talu for their understanding and continuous morale support.

The author feels himself responsible to mention his deep gratitude to his parents, his brother, his sister, his aunt, his uncle and his grandparents.

TABLE OF CONTENTS

ABSTRACT..... iv
ÖZ..... vi
ACKNOWLEDGEMENTS..... ix
TABLE OF CONTENTS..... x
LIST OF TABLES..... xii
LIST OF FIGURES..... xv

CHAPTERS

1. INTRODUCTION **1**
2. LITERATURE REVIEW & THEORETICAL BACKGROUND **4**
2.1. Study Area 4
2.2. Eutrophication 8
2.3. Historical Progress of Water Quality in Lake Eymir 10
2.4. Remote Sensing of Chlorophyll-a 11
2.5. Artificial Neural Networks (ANNs) 15
2.6. Previous Studies on Remote Sensing of Chlorophyll-a 16
3. METHODS and MATERIALS **31**
3.1. Information about Satellite Images 32
3.2. Field Study 35
3.2.1. Locations of Sampling Points 37
3.2.2. In-Situ Measurements 39
3.2.3. Water Sampling 43
3.3. Laboratory Analyses 43
3.3.1. Chlorophyll - a 43
3.3.2. Total Suspended Solids 44
3.3.3. Color 45
3.4. Image Analysis 46
3.5. Data and Principle Component Analysis 47

3.5.1. Modeling of chlorophyll-a using hybrid inputs 48
3.5.1.1. Linear and Non-Linear Regression Models 48
3.5.1.2. MATLAB Neural Network Models.....49
4. RESULTS and DISCUSSION.....53
4.1. Statistical Analysis of Field Data.....53
4.2. Analysis of the Relationships Between Chlorophyll-a and Band Reflectance Values.....58
4.3. PCA Results 63
4.4. Modeling of Chlorophyll-a 67
4.4.1. Modeling of Chlorophyll-a for Field 1 68
4.4.2. Modeling of Chlorophyll-a for Field 2 77
4.4.3. Modeling of Chlorophyll-a for Field 3 86
4.4.4. Modeling of Chlorophyll-a for Combined Data 97
4.4.5. Overall Analysis 113
5. CONCLUSIONS and RECOMMENDATIONS FOR FUTURE STUDIES 118
REFERENCES 121
APPENDICES
A. SPATIAL CHLOROPHYLL-A DISTRIBUTIONS 130
B. IMAGES of LAKE EYMIR 141
C. CORRELATION MATRICES and EIGEN VECTORS.....147

LIST OF TABLES

TABLES

Table 2-1: Summary table for the previous studies 30
Table 3-1: Satellite Imageries and Field Work Dates 32
Table 3-2: Meteorological Data between September 9th and 18th, 2009. 33
Table 3-3: Meteorological Data between April 26th and May 2nd, 2010. 33
Table 3-4: Meteorological Data between July 28th and August 2nd, 2010.....33
Table 3-5: The Spectral Bandwidths of the Quickbird Satellite 34
Table 3-6: Sampling Points Coordinates for Field Study 1 and 2 38
Table 3-7: Sampling Points Coordinates for Field Study 3 39
Table 4-1: Parameter vs. Field Study.....54
Table 4-2: Statistical Values for Field Study 1.....55
Table 4-3: Statistical Values for Field Study # 2.....56
Table 4-4: Statistical Values for Field Study # 3.....57
Table 4-5: Statistical Values for All Data.....58
Table 4-6: Band Correlations (R values) with Chlorophyll-a Concentration 59
Table 4-7: Band Ratio Correlations with Chlorophyll-a Concentration.....59
Table 4-8: Results of Literature Equations 62
Table 4-9: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 1) 64
Table 4-10: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 2) 64
Table 4-11: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 3) 65
Table 4-12: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (All Fields).....66
Table 4-13: Parameter Comparison Table 66

Table 4-14: Field 1: Comparison of the Linear Regression Models with Different Inputs.....69
Table 4-15: Applicability of Linear Regression Models of Field 1 to Field 2 and 3 ...71
Table 4-16: Field 1: Comparison of the Non-Linear Regression Models with Different Inputs.....72
Table 4-17: Applicability of the Non-Linear Regression Models of Field 1 to Field 2 and Field 3.....73
Table 4-18: Field 1: Statistical Information for ANN Model with Hybrid Inputs 74
Table 4-19: First Field Study Result Summary Table for Hybrid Models 77
Table 4-20: Second Field Study Linear Regression Results Comparison Table.....78
Table 4-21: Applicability of Linear Regression Model of Field 2 to Field 1 and 3.....80
Table 4-22: Field 2: Comparison of the Non-Linear Regression Models with Different Inputs.....81
Table 4-23: Applicability of the Non-Linear Regression Models of Field 2 to Field 1 and Field 3.....82
Table 4-24: Second Field Study ANN Results Comparison Table.....83
Table 4-25: Second Field Study Result Summary Table for Hybrid Models 86
Table 4-26: Field 3: Comparison of the Linear Regression Models with Different Inputs.....87
Table 4-27: Applicability of Linear Regression Model from Field 3 to Field 1 and 2.90
Table 4-28: Field 3: Comparison of the Non-Linear Regression Models with Different Inputs.....91
Table 4-29: Applicability of Non-Linear Regression Model of Field 3 to Field 1 and 2 93
Table 4-30: Field 3: Statistical Information for ANN Model with Hybrid Inputs 94
Table 4-31: Third Field Study Result Summary Table for Hybrid Models 97
Table 4-32: Combined Dataset: Comparison of the Linear Regression Models with Different Inputs 98
Table 4-33: Applicability of Linear Regression Model obtained using the combined dataset to Fields 1, 2 and 3.....100
Table 4-34: Non-Linear Regression Results Comparison Table for All Data 101
Table 4-35: Applicability of Linear Regression Model for the Combined Data to Fields 1, 2 and 3 103
Table 4-36: ANN Results Comparison Table for All Dataset 104

Table 4-37: Application of the ANN model with Hybrid Inputs of the Combined Dataset to Field 1.....107
Table 4-38: Results Comparison Table for Field 2 (combined model).....109
Table 4-39: ANN Results Comparison Table for All Dataset, Third Field Data 111
Table 4-40: Summary Table for the Models with Hybrid Inputs 113
Table 4-41: Hybrid Model Results Comparison Table 114
Table A-1: Correlation Matrix of Parameters Measured in First Field Study 147
Table C-2: Eigen Values of Parameters Measured in First Field Study 148
Table C-3: Eigen Vectors of Parameters Measured in First Field Study.....148
Table C-4: Correlation Matrix of Parameters Measured in Second Field Study.....149
Table C-5: Eigen Values of Parameters Measured in Second Field Study.....150
Table C-6: Eigen Vectors of Parameters Measured in Second Field Study.....150
Table C-7: Correlation Matrix of Parameters Measured in Third Field Study.....151
Table C-8: Eigen Values of Parameters Measured in Third Field Study.....152
Table C-9: Eigen Vectors of Parameters Measured in Third Field Study.....152

LIST OF FIGURES

FIGURES

Figure 2-1: Location of Lake Eymir 5
Figure 2-2: Surface water inputs and outputs to Lake Eymir 6
Figure 2-3: Spectral curve for chlorophyll-a.....14
Figure 2-4: Structure of Artificial Neural Network 15
Figure 3-1: The Spectral Bands for Quickbird 2 and World View 2 Satellites 34
Figure 3-2: Sampling Points for Field Study 1 and 2 37
Figure 3-3: Sampling Points for Field Study 3 38
Figure 3-4: YSI 6600 EDS Probe 40
Figure 3-5: Kestrel 4500 Anemometer 41
Figure 3-6: PAR Sensor.....42
Figure 3-7: Secchi Disc.....42
Figure 3-8: Van Dorn Water Sampler.....43
Figure 4-1: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 1).....69
Figure 4-2: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Field 1) 70
Figure 4-3: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 1).....72
Figure 4-4: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Field 1) 73
Figure 4-5: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 1).....74
Figure 4-6: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 1) 75
Figure 4-7: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 1) 75
Figure 4-8: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 1) 76

Figure 4-9: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 2) 79
Figure 4-10: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Field 2) 79
Figure 4-11: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 2) 81
Figure 4-12: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Field 2) 82
Figure 4-13: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 2) 83
Figure 4-14: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 2) 84
Figure 4-15: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 2) 84
Figure 4-16: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 2) 85
Figure 4-17: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 3) 88
Figure 4-18: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 4 – Band Reflectance Inputs (Field 3) 88
Figure 4-19: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 8 – Band Reflectance Inputs (Field 3) 89
Figure 4-20: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 3) 92
Figure 4-21: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with 4 – Band Only Reflectance Inputs (Field 3) 92
Figure 4-22: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 8 – Band Only Reflectance Inputs (Field 3) 93
Figure 4-23: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 3) 94
Figure 4-24: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 3) 95
Figure 4-25: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 3) 95

Figure 4-26: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 3) 96
Figure 4-27: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Combined Dataset) 98
Figure 4-28: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Combined Data) 99
Figure 4-29: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Combined Data) 101
Figure 4-30: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Combined Data) 102
Figure 4-31: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Combined Data).....104
Figure 4-32: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set – Combined Data) 105
Figure 4-33: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set – Combined Data) 105
Figure 4-34: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set – Combined Data) 106
Figure 4-35: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 1) 107
Figure 4-36: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 1) 107
Figure 4-37: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 1) 108
Figure 4-38: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 1) 108
Figure 4-39: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 2) 109
Figure 4-40: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 2) 109
Figure 4-41: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 2) 110
Figure 4-42: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 2) 110

Figure 4-43: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 3) 111
Figure 4-44: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 3) 111
Figure 4-45: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 3) 112
Figure 4-46: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 3) 112
Figure 4-47: Stereo Image of Lake Eymir 116
Figure 4-48: Satellite Pass Path and Wave Directions in Lake Eymir 117
Figure A-1: First Field Study Measured Chl-a Distribution 130
Figure A-2: First Field Study Linear Regression Predicted Chl-a Distribution 131
Figure A-3: First Field Study % Deviation of Linear Regression Predicted from Measured Chl-a Distribution 131
Figure A-4: First Field Study Non-Linear Regression Predicted Chl-a Distribution 132
Figure A-5: First Field Study % Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution 132
Figure A-6: First Field Study ANN Predicted Chl-a Distribution 133
Figure A-7: First Field Study % Deviation of ANN Predicted from Measured Chl-a Distribution 133
Figure A-8: Second Field Study Measured Chl-a Distribution 134
Figure A-9: Second Field Study Linear Regression Predicted Chl-a Distribution 134
Figure A-10: Second Field Study Deviation of Linear Regression Predicted from Measured Chl-a Distribution 135
Figure A-11: Second Field Study Non-Linear Regression Predicted Chl-a Distribution 135
Figure A-12: Second Field Study Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution 136
Figure A-13: Second Field Study ANN Predicted Chl-a Distribution 136
Figure A-14: Second Field Study Deviation of ANN Predicted from Measured Chl-a Distribution 137
Figure A-15: Third Field Study Measured Chl-a Distribution 137
Figure A-16: Third Field Study Linear Regression Predicted Chl-a Distribution 138
Figure A-17: Third Field Study Deviation of Linear Regression Predicted from Measured Chl-a Distribution 138

Figure A-18: Third Field Study Non-Linear Regression Predicted Chl-a Distribution 139
Figure A-19: Third Field Study Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution 139
Figure A-20: Third Field Study ANN Predicted Chl-a Distribution.....140
Figure A-21: Third Field Study Deviation of ANN Predicted from Measured Chl-a Distribution.....140
Figure B-1: Lake Eymir First Study Satellite Imagery 141
Figure B-2: Lake Eymir Second Study Satellite Imagery.....142
Figure B-3: Lake Eymir Third Study Satellite Imagery.....143
Figure B-4: Stereo Image of Lake Eymir 144
Figure B-5: Stereo Image of Lake Mogan and Lake Eymir.....145
Figure B-6: Digital Elevation Model of Lake Eymir 146

CHAPTER 1

INTRODUCTION

In today's world, there are several water bodies that suffer from eutrophication problem. The eutrophication problem is defined as the excess loading of water bodies with nutrients originating primarily from anthropogenic sources. Eutrophication can cause major changes in environment and water ecosystem. Main sources of nutrients are the increase in human population, agricultural production and industrial growth. As a result of eutrophication, major changes are observed in aquatic plants and species due to decrease in DO concentrations in water. The most important indicator of eutrophication in water bodies is the excess growth of algae and larger aquatic plants. This may make a fresh-water source non-potable and increase treatment cost due to filtration problems. (Yang et al., 2008)

In order to assess the water quality in a lake, monitoring and management systems are needed. After gathering datum, models can be employed to make predictions about the status of water quality for various load reduction or input scenarios. However, monitoring by means of field work and laboratory analysis can be costly, especially when a high number of samples are required for determination of spatial distribution of water quality parameters of concern. This problem can be valid for chlorophyll-a as well, which is an important parameter that indicates the abundance and concentration level of algae. Algae concentrations, in return, give an idea about the eutrophication state, water quality and pollution level.

Chlorophyll-a determination may necessitate extensive sampling especially in eutrophic lakes. This is due to the fact that spatial distribution of chlorophyll-a in eutrophic lakes is not homogeneous. Chlorophyll-a is contained in different amounts in different algal species. Moreover, algal community can be concentrated in specific locations in water body. On top the of those, depending on several water

quality parameters, the concentrations may change temporarily as well as spatially. In extreme cases, variations can be observed during a day too. Therefore, rigorous sampling is required in order to determine chlorophyll-a concentrations. Further inconveniences can be experienced if sampling has to be applied at a location hard to reach. These in return may increase sampling costs and labor requirements. However, remote sensing technology can provide means to overcome these inconveniences.

In remote sensing, satellite images can be used to determine spatially distributed parameter values by using the spectral reflectance information at image pixels. An advantage of remote sensing is its ability to generate a continuous parameter distribution in the area of concern. By using a small number of samples, predictions can be made for a greater area. As a result, sampling costs can be minimized. As sampling costs are reduced, a higher monitoring frequency can be feasible.

As remote sensing of chlorophyll-a will be discussed in the next chapter, there are several studies in literature that use remotely sensed data for the prediction of chlorophyll-a concentrations (Zhang et al., 2002; Cauwer et al., 2004; Cannizzaro et al., 2006; Ekercin, 2007; Wang et al., 2008; Cui et al., 2008; Lin et al., 2008). However, prediction performance decreases especially for shallow lakes due to factors that may impact the reflectance values other than chlorophyll-a such as bottom reflectance and resuspension of bottom sediment (Panda et al., 2004). As a result, new approaches are required to increase the chlorophyll-a prediction performance using remotely sensed data.

In this study, algal (chlorophyll-a) concentrations in Lake Eymir, an eutrophic lake located at 20 km south of Ankara, are determined using models that use both remotely sensed data and field parameters that may impact the reflectance values in remotely sensed images as hybrid inputs. The field parameters considered in model development are temperature, specific conductivity, total dissolved solids (TDS), salinity, dissolved oxygen concentration (DO), pH, total suspended solids (TSS), color, photosynthetically active radiation (PAR), secchi depth (SD), total water column depth, wind speed, wind direction, humidity and air temperature. For modeling of chlorophyll-a, linear and non-linear regression and artificial neural network (ANN) models are used. As a result, differently from the typical models in literature that use only remotely sensed data for chlorophyll-a determination, hybrid

input is used in prediction of chlorophyll-a concentrations. This study concentrates mainly on the performances of regression and ANN models with hybrid input data in modeling of chlorophyll-a concentrations.

Parameter values that are obtained from field, laboratory and image analysis are analyzed using PCA (Principle Component Analysis). By applying PCA, the number of parameters that will be used in models with hybrid inputs are determined. Then, chlorophyll-a concentrations are modeled by regression and ANN models that use hybrid dataset. Chlorophyll-a modeling is performed by only remotely sensed data as well for comparison. Results are compared between hybrid input models and band only models based on chlorophyll-a prediction performances. Finally, spatial distribution of chlorophyll-a is generated using the models derived.

The organization of this thesis is as follows. In the second chapter, literature review and background of this study is given. Within the chapter, information about study area, eutrophication, water quality in Lake Eymir, remote sensing of chlorophyll-a and relevant previous studies and artificial neural networks are provided. Later, in the third chapter, methods and materials that are used throughout the study are given. This chapter includes information about the satellite images used, details of field and laboratory studies and corresponding image and data analyses. In the fourth chapter, results for regression and neural network models are given and discussed. Finally, in the fifth chapter, conclusions and recommendations for further studies are given.

CHAPTER 2

LITERATURE REVIEW & THEORETICAL BACKGROUND

2.1. Study Area

Lake Eymir is located at 20 km south of Ankara. It is located in 39.28 N and 32.30 E, at an altitude of 969 m (Tan, 2002). Its location is depicted in Figure 2-1. Also stereo (3-D) images of Lake Eymir is also given in Figure B-4 and Figure B-5 with digital elevation model of the area in Figure B-6. Lake Eymir is an important recreational area for Ankara and the Middle East Technical University (METU). It is situated in METU Campus area since 1958 and has been used as the primary drinking water source for METU until 1990 (Demirci et al., 2005).

Formation of the lake is predicted to be as a result of tectonic activities that took place in 4th geological age (Diker, 1992) and alluvium build up (Camur et al., 1997). The lake area is approximately 1.25 km^2 with approximate dimensions of 3000 m x 500 m x 5.5 m (Camur et al., 1997). It has 13 km of shoreline (Tan et al., 2005). The average volume of the water in the lake is $3.88x10^6$ m^3 (Karakoc et al., 2003). Changing volume provides a retention time of 1.8 to 23 years (Beklioglu et al., 2003). This value is reported to change between 3.4 to 35.4 years by Tan (2002). Tan (2002) also reported that the water depth in Lake Eymir in 2002 was in the range of 4.3 to 6.0 m. In addition to this, the lake has an average water volume of $3.88x10^6$ m^3.

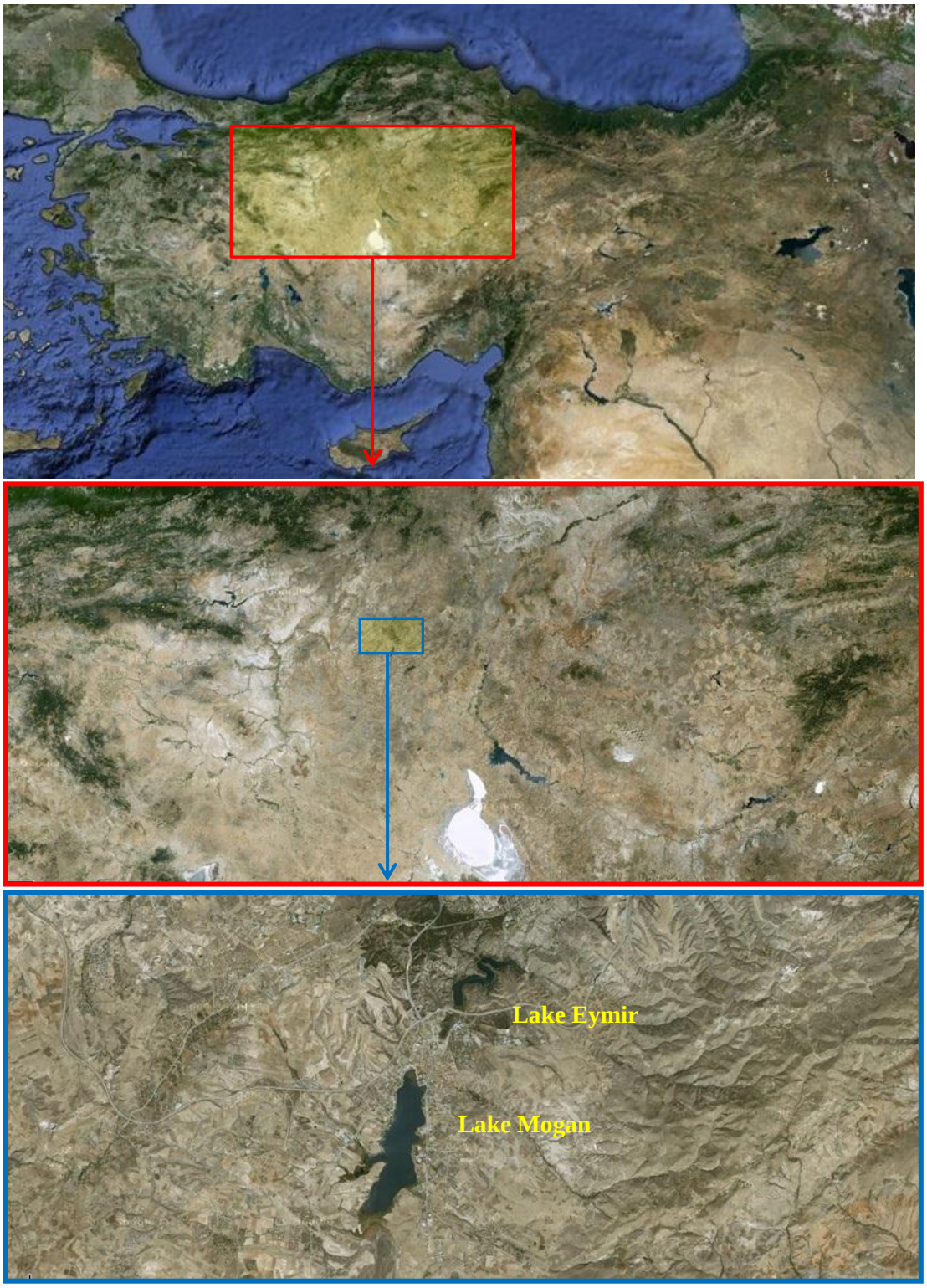

Figure 2-1: Location of Lake Eymir

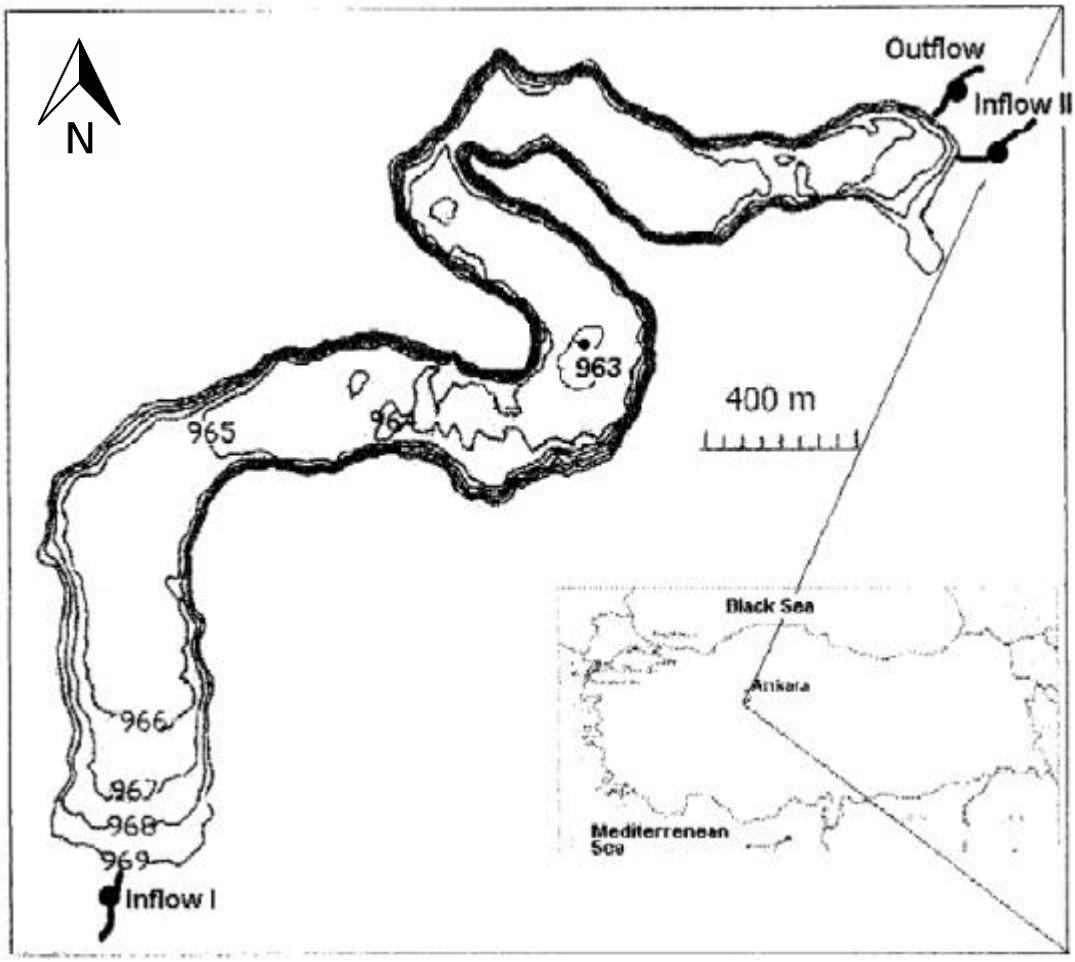

Figure 2-2: Surface water inputs and outputs to Lake Eymir
(Taken from Altınbilek et al., 1995)

Lake Mogan generates the major input to Lake Eymir. The natural channel connecting these two lakes is modified by building a gated concrete channel to control the inflow from Lake Mogan to Lake Eymir for the purpose of sustaining the water level in Lake Mogan. A wetland is formed at the inlet that receives input from Lake Mogan. Therefore, the wetland, which is rich in nutrients, is fed by Lake Mogan. Kışlakçı Creek provides water inflow to Lake Eymir in wet seasons (Yagbasan et al., 2009). In addition, an underground water inflow at northern part feeds the lake at an approximate flowrate of 17 L/s (Altınbilek et al., 1995). Besides these inflows, Diker (1992) reports that the drainage area of Lake Eymir contains some of the regions in the Elma Dag and this provides additional water inflow originating from the mountain to the lake. Also, precipitation can be included in the water budget of the lake. The average precipitation in the area is approximately 390 ± 76 mm/year. The basin can be defined as semi-arid, based on precipitation and

vegetative properties (Ozaydın et al., 2001). As depicted in Figure 2-2, the lake drains into Imrahor Creek at south of the lake. Evaporation and groundwater discharge can be counted as the other outflows. In a yearly basis, the average evaporation amount is 1092.2 mm. According to Altınbilek et al (1995), there is a discharge to underground layers in an average amount of 2 L/s.

Naturally, the water level in Lake Mogan is 3 m higher than for Lake Eymir. Yet, there are fluctuations in water levels of both lakes that vary seasonally. In Lake Eymir, the water level varies by 0.5 – 1.0 m (Yagbasan et al., 2009). 98% of the water inflow to Lake Eymir is supplied by Lake Mogan (EPASA, 2006). But, in recent years due to the low water level in Lake Mogan, the gate on the channel that connects two lakes has been kept shut and there has been no water passage from Lake Mogan to Lake Eymir. However, on April 2010, the water depth in the lake increased suddenly due to the incoming water from Lake Mogan. This was as a result of the broken gate that controls the inflow from Lake Mogan. The problem with the gate arose due to the pressure resulting from large amount of water in Lake Mogan. As a result of this event, the depth of water in Lake Eymir increased by approximately 1 m within few days. Further with heavy precipitation, the increase in the water depth of Lake Eymir reached to 1.5 – 1.7 m. In April 2010, the channel was rehabilitated and the gate was reinstalled by the Gölbaşı Municipality (An Concrete Canal Between Mogan and Eymir are Refined, 2010).

The largest residential area that is located in the vicinity of the lake is the Gölbaşı District with an approximate 2009 population of 86,749 people (TUIK, 2010). Moreover, there are two settlements close by. These are TEAS (Turkish Electricity Transmission Corporation) settlement with an approximate population of 5000 capita and the Police Academy. In addition to these settlements, there are also some small scaled industries around the lake and a municipal landfill site is present in the basin, which is currently not used. The landfill is located nearby the channel that connects Lake Mogan and Lake Eymir.

Starting from 1970s, Lake Eymir became a discharge area for the untreated wastewaters of the Gölbaşı Municipality. Especially, southern part of the lake was impacted. The discharges have caused eutrophication problem in the lake. Due to the importance of the lake for its diversity and being one of the few natural recreational areas around Ankara, the area surrounding the lake and 245 km^2 of its catchment basin is declared as “Special Environmental Protection Area” by the Decree of Cabinet of Ministers with the declaration number of 90/1117 in

22.10.1990. In year 1995, a by-pass line was constructed to divert the input from the Gölbaşı Municipality sewage system to the outflow of Lake Eymir for the purpose of reducing the pollution load to the lake (Altınbilek et al., 1995; Beklioglu et al., 2003). However, since the by-pass line had not been operated continuously, the lake continued to be a receiving body. Besides this input, the untreated wastewater of TEAS settlement reached to the lake as well. TEAS was connected to main wastewater collector in 2001 (Tan, 2002).

Considering the nutrient rich wastewater discharges that find way to the lake, there is a high potential for algal growth. In addition, the shallow depth of water in the lake enhances the conditions for eutrophication. The lake has been declared as highly eutrophic (Elahdab, 2006; Beklioglu et al., 2003; Beklioglu et al., 2008; Karul et al., 2000; Karakoç et al., 2003; Yenilmez et al., 2010; Beklioglu et al., 2010). There is abundance of submerged and surface plants (Beklioglu et al., 2003; Muluk et al., 2005; Tan et al., 2005).

2.2. Eutrophication

The word eutrophic originates from Greek language. “Eu” means well and “trophe” means nourishment / nutrient (Yang et al., 2008). The eutrophication problem was first realized in 1800s. Due to waste deposition in streams, hypoxia (oxygen depletion) and anoxia (lack of oxygen) were observed. Eutrophication leads to excessive algal growth and it is one of the important problems in lake ecosystems that are located especially in developing areas, close to residential areas, fertilized and irrigated lands.

According to the investigations conducted by UNEP (United Nations Environmental Protection), 30 – 40% of the lakes in the world are subject to problems related to eutrophication. This situation can cause decrease in sunlight penetration to deeper parts of water column. As a result, photosynthesis might be retarded for the aquatic plants in lake bottom. Eutrophication also effects dissolved oxygen (DO) concentrations. Very low levels of DO as a result of eutrophication may lead to massive deaths of aquatic species. In the abundance of nutrients, mass volume of algae forms a thick green layer in the lake surface, called as the “green scum”. In some cases this algal population might release toxins and decompose organic matters into harmful gases that harm fish and other aquatic organisms

(Yang et al., 2008). Along with decomposition of organic matters, hypoxia and anoxia problems can be seen depending on the turbulence of water system (De Jonge et al., 2002)

Larger water bodies have larger assimilative capacities for nutrient loadings from anthropogenic sources. Smaller water bodies are more susceptible to the eutrophication problem. (De Jonge et al., 2002). The eutrophication is inevitable in the lifetime of lakes. Geldiay (1949) states that the life of lakes can be defined as, oligotrophic state – eutrophic state – marshy place – marshy ground – meadow marshy ground – soil. Anthropogenic sources (industrial wastewater, municipal wastewater) fasten the change from one stage to another. (Larsson et al., 1985).

Eutrophication mainly originates from excess nitrogen (N) and phosphorus (P) species loadings into the water body. However, availability of light, temperature of water, micro nutrients and turbulence are among other factors that impact algal growth and eutrophication process. Based on the conducted studies in literature, the most favorable situation for eutrophic conditions can be defined as: high total nitrogen (TN) and total phosphorus (TP) concentrations with slow currents in the water body and suitable environmental factors such as increased temperature that boost microbial activity (Yang et al., 2008). As a result of eutrophication, fertility of freshwaters increases. Fertilized waters require extensive filtering due to their offensive aesthetic properties like transparency, smell and taste, before their usage for economic purposes. Besides, recreational areas with eutrophic freshwaters are possible health-hazard sites (Fish, 1969)

Eutrophication related studies started to gain importance after 1960s. According to Smith et al (2006), eutrophication concept changed in recent years. Smith et al. (2009) states that eutrophication is a natural phenomenon and there is no need to follow its stages unless there is an artificial effect which would increase the rate of algal growth and water quality deterioration. The artificial effects are due to macro nutrient loadings from point and non-point sources, such as carbon, nitrogen and phosphorus.

2.3. Historical Progress of Water Quality in Lake Eymir

Although the focus of this study is not the current status of water quality in Lake Eymir, a basic review of its progress is provided below since chlorophyll-a concentrations have been impacted by the deterioration of water quality in the lake.

The first detailed studies in Lake Eymir were conducted by Chaput in 1931 Woltereck and Mann in 1934, and Calvi and Woltereck in 1936 (as cited in Geldiay, 1949). Later on, Geldiay studied Lake Eymir between 1946 and 1947 (Geldiay, 1949). These studies provided sufficient water quality information in order to evaluate its trophic state. Water quality studies conducted in Lake Eymir mainly concentrated on concentrations of chlorophyll-a, phosphorus, nitrogen, DO and total suspended solids.

Lake Eymir was classified as eutrophic in previous studies (Diker, 1992; Altınbilek et al., 1995). Based on the study that was conducted by Geldiay (1949), DO concentrations were measured as low as 0 mg/L. The anoxic condition was also supported by Diker (1992). In the study conducted by Diker (1992), low concentrations of DO were observed in bottom layers during winter. Moreover, the productivity of algal community went down with decreasing temperature which was reflected in chlorophyll-a concentrations. In spring, as water temperature increased, chlorophyll-a concentrations increased. As a result of photosynthesis, DO concentrations were higher (Geldiay, 1949).

In eutrophication, macro nutrients (especially phosphorus and nitrogen) are the major contributors to algal productivity. Changes in the amounts of macro nutrients effect the lake water quality directly. Phosphorus and nitrogen concentrations in Lake Eymir were measured in several studies (Diker, 1992; Altınbilek et al., 1995; Tan, 2002; Beklioğlu, 2002). Results of these studies revealed that similar variations were observed in phosphorus and nitrogen concentrations. From 1992 to 1995, phosphorus and nitrogen concentrations decreased on a yearly average basis. This decrease in nutrient concentrations was related to the diversion of incoming wastewater to the lake (Tan, 2002). In return, chlorophyll-a concentrations declined expectedly as well (Diker, 1992; Tan, 2002)

In 1998 and 2000, biomanupulation was applied in Lake Eymir to improve its water quality. In biomanupulation, fish types that are present in the lake were modified (Tan, 2002). In later years, water quality was monitored by Ozen (2006).

Ozen (2006) claimed that both of the biomanupulation applications resulted in a decline in nutrient concentrations. Phosphorus and nitrogen concentrations exhibited an increasing trend from 1997 to 2003. However, after 2003, there was a large increase in phosphorus concentrations (Muluk, 2005), while nitrogen concentrations were stable. In 2004, the yearly average chlorophyll-a concentration reached its maximum value compared to the ones observed in the preceding 10 years. However, high concentrations lasted for only one year. In 2005, nutrient concentrations were lower in the lake (Ozen, 2006). A positive correlation was observed between nutrient and chlorophyll-a concentrations which are in line with the results of Altınbilek et al. (1995).

Changing water quality in Lake Eymir is evaluated based on total suspended solid concentrations and secchi disc depth as well (Altınbilek et al., 1995; Beklioglu, 2002; Muluk, 2005; Ozen, 2006). Studies revealed that the total suspended solid concentration is closely related to both chlorophyll-a concentration and secchi disc depth. Starting from 1995, the yearly average secchi disc depth values have been changing drastically. Until the first biomanupulation application in 1998, the secchi disc values were reaching up to a maximum of 4 m. Following the first biomanupulation, the average secchi disk depths decreased until 2004. Same as other mentioned parameters, 2004 was the year with significant changes. In 2004, secchi disc depth values increased for one year and then decreased again in 2005 (Ozen, 2006). These changes in the secchi disc depths can be explained by variations in total suspended solid concentrations. Total suspended solid concentrations showed a decreasing trend similar secchi disc depth until the first biomanupulation. Following that, total suspended solid concentrations were stable until 2005. Unlike the secchi disc depth, total suspended solid concentrations increased in year 2005. When these parameters are compared with respect to yearly averaged chlorophyll-a concentrations, the general trends between parameters were as expected. Increase in chlorophyll-a concentrations resulted in decreasing secchi disc depths with increasing total suspended solid concentrations.

2.4. Remote Sensing of Chlorophyll-a

Remote sensing can be defined basically as: gathering information from an object or area from a distance, without being in a contact with the corresponding

object or area (Barrett et al., 1992). By the use of remote sensing technology, models can be generated and used repeatedly for obtaining specific information from corresponding areas. Moreover, distributions can be obtained using lower frequencies of field sampling with the help of generated models. This situation provides an advantage for monitoring of large areas. Besides, remote sensing can ease problems concerning sampling of sites with challenging conditions such as harsh topography. Finally, since sampling requirements may be reduced, sampling costs can be decreased.

The concept behind remote sensing technology is to use the absorbance and reflectance of electromagnetic waves by a given material or an object. Every object has some potential for absorbance in specific wavelengths, which also means each object absorbs different amounts of energy in different wavelengths. This is referred to the spectral signature.

In the remote sensing, satellite image data that are retrieved by means of optical and thermal sensors placed in different platforms are used. Retrieved electromagnetic wave intensity is converted to numbers that vary based on the radiometric resolution of the sensor; 0 corresponds to black and 2^n corresponds to white where n defines the bit value of the image. The values in the range of 0 to 2^n corresponds to color tones between black and white. In the image, there will only be one value per pixel (the smallest unit of the image) per band and this is called as the digital number (DN) of the corresponding pixel for the corresponding band. In remote sensing studies, the main aim is to obtain a relationship between DNs and measured field data in exactly the same coordinates (geometric correction is required for accurate coordinate values) that define both the pixel location and the point location on earth. There are several studies present in literature that report empirical relationships between DNs and the field data. Yet, most of these relationships are unique to the cases considered. Remote sensing have been applied for environmental monitoring and management such as; trophic state detection for surface waters, determination of chemical composition of water and soil surfaces, climate analysis, erosion management, land use determination, etc. (Elahdab, 2006)

As a photon penetrates to water body, mainly three mechanisms are observed; namely, absorption, scattering and diffraction. The occurrence of these mechanisms mainly depends on the state of water. Water bodies are classified into classes; Case 1 and Case 2. The Case 1 is the open ocean in which the major

optically active constituent is the chlorophyll, and water lacks suspended particles. On the other hand Case 2 waters are coastal, estuary or inland waters including lakes. In Case 2, there is a complex relationship between the reflectance and water quality parameters. This complexity is mainly due to the presence of chlorophyll-a, suspended particles, colored dissolved organic matter (CDOM), terrigenous particles, anthropogenic particles and mainly due to their high concentrations. Those structures are able to cause high amounts of scattering. (Sudheer et al., 2006 and Kishino et al., 2005). In water bodies, the blue region of light penetrates into deeper distances due to its lower wavelength. This region is mainly used in order to observe the deeper parts of water columns. On the other hand, in order to study the depths closer to surface, mainly higher wavelengths (the red and infrared region of light) are used. In addition to this, near-infrared region also provides land-water distinction (Blanco et al., 2007). According to Blanco et al (2007), in Case 2 waters, there is high amount of radiance measured in both the visible and infrared portions of light due to presence of dissolved and suspended particles.

Another main idea behind remote sensing technology is to determine the degree of light that is reflected. For water bodies, as depth increases, light density decreases due to absorbance of light. The light density can be determined by measuring the photosynthetically active radiation (PAR). The degree in which the light intensity decreases is called as the attenuation. The attenuation is a function of absorption and scattering.

In remote sensing of waters, passive remote sensors which detect emitted or reflected electromagnetic energy from a natural source (i.e. sun), can be used. Based on the reflectance properties of materials, only a portion of the incoming light is reflected. In water bodies, this reflectance is mainly due to the reflectance from water and scattering from atmospheric particles. This reflected light is mainly affected by the intensity of solar light, angle of reflectance and the optical properties of material. The optical properties are known to be independent of the intensity or angle of reflectance and depend on the presence of suspended and dissolved particles in water.

In remote sensing studies on water bodies, mainly the visible region is used which has a wavelength ranging from 390 to 750 nm. Besides the visible region, near-infrared region is also used in some studies which scans a light window around 800 nm. In the visible region, the red portion of light attenuates more quickly than

the blue light portion due to its higher wavelength. After scattering and absorption, reflected light portion is recorded by sensors on a platform (i.e. a satellite).

Main light absorbers in water bodies can be listed as algal species, inorganic and organic particles and water. Algal communities including phytoplankton are capable of photosynthesis. So, as they receive sun light, they can use it for photosynthesis. Because of their structure, they appear as green as they reflect the green portion of light. However, they absorb the red and blue portions of light in high amounts (Figure 2-3). Algae mainly absorb the red portion of light. On the other hand, dissolved organic particles mainly absorb in the blue range and highly reflect the red portion of light. In addition to algae and dissolved organic particles, inorganic particles mainly act as a reflector for the incoming light. This reflectance can also be defined as turbidity. Blanco et al (2007) supports that water bodies have dark colors in the visible region. In the presence of suspended and dissolved particles in Case 2 waters, brightness increases with the increase in turbidity. In contrast, for Case 1 waters, darker color means deeper water column while lighter color means shallower water column.

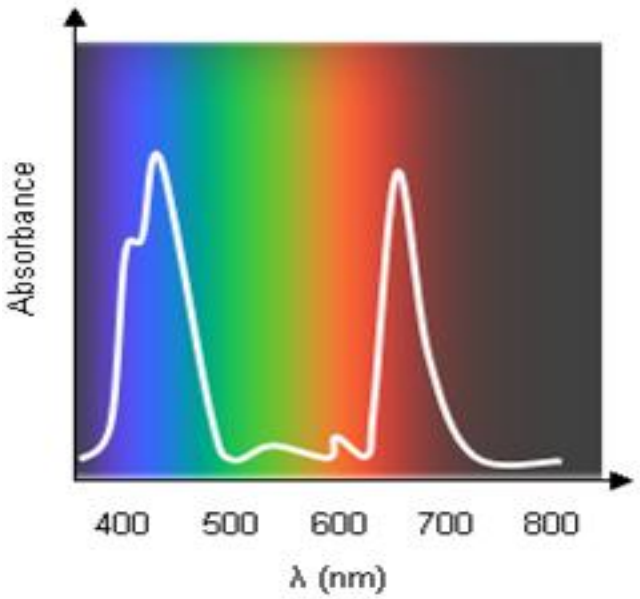

Figure 2-3: Spectral curve for chlorophyll-a

2.5. Artificial Neural Networks (ANNs)

The main idea behind ANNs is to mimic the functioning of human brain in modeling a phenomenon. In human brain, neurons collects signals from other neurons through structures called dendrites. Then, a neuron sends a responsive electrical signal through axons that splits to several branches. Each branch ends with a structure called a synapse. The synapse converts the incoming electrical signal from an axon into electrical effects that generates the activity in connected neurons. As a neuron receives electrical activity, it sends another signal to the following axon. Here, learning process occurs by variations in the effectiveness of electrical activities (Stergiou et al., 2010). An ANN is a mathematical representation of this natural processing in the neural network of a brain to solve a problem.

ANN model provides a non-linear adaptive processing of data in a mathematical structure. This structure is called as the network architecture or mesostructure. In this structure there are two building units which are processing units (neurons or nodes) and weighted connections (links). An ANN structure representation is given in Figure 2-4. This structure contains processing units and layers that are connected to each other to form a network. Data enters to network from the input layer, then passes through hidden layers and finally reaches to output unit(s) (Kneale et al., 2004). Each layer is composed of a number of neurons which are connected to the neurons in the proceeding layer by weighted links. The ANN is an alternative technique for analysis and modeling. This technique can be employed for data classification, pattern determination and signal processing (Keiner, 1998).

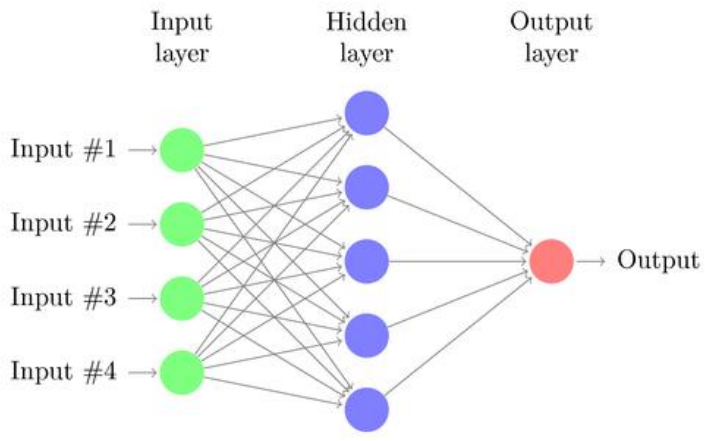

Figure 2-4: Structure of Artificial Neural Network

There are three main steps in ANN modeling. The first step is the training step in which a part of the data set is used to derive (or learn) the possible patterns between input and output data. Pattern search is provided by comparing input (predictor) and output (predictand) data. Training stage of ANN model is cautious and focused for changes in the knowledge structure of a system. In the training stage, the connections between neurons are adjusted for a better fit. There are two common ways of learning in an ANN model. These are the supervised and unsupervised learning. The supervised learning requires an association of each input data with an output data. On the other hand, in unsupervised training, the output is a set of clusters. The aim of the training stage is to minimize output error by use of different algorithms like “search the error surface” and “descend the gradient”. The validation step searches whether overtraining (or memorization) has occurred or not. Finally, the validation step is followed by the testing stage. The testing stage corresponds to verification of the model. Basically in this stage, the prediction capability of the ANN model is tested against the observation data. This capability is measured by Pearson’s correlation coefficient (R^2) and root mean square error (RMSE) (Kneale et al., 2004). Based on the studies by Musavi et al (2002), Zhang et al (2002), Cui et al (2008), ANNs are highly applicable for Case 2 water bodies when a relationship is sought between certain water quality parameters and remotely sensed data.

2.6. Previous Studies on Remote Sensing of Chlorophyll-a

Chlorophyll-a concentrations are modeled using remotely sensed data in several studies. In these studies several sensors have been used. One of the most common satellite used to retrieve data is Landsat satellites (Wang et al., 2008; Keinel, 1998; Zhang, 2002; Bilge et al., 2003; Liu et al., 2003; Sudheer et al., 2006; Yang and Du, 2008). These satellites have a spatial resolution of 60 m (Keinel et al., 1998). Besides Landsat, MODIS is another satellite that has been frequently used (Wu, 1998; Moisan et al., 2005; Lin et al., 2003; Cauwer et al., 2004; Dall’Olmo et al., 2005; Pozdnyakov et al., 2005; Wu et al., 2009; Wang et al., 2010). It has a spatial resolution of 1000 m (Wu et al., 1998). In most of the open water studies, SeaWIFS satellite has been employed (Moisan et al., 2005; Gross et al., 1999; Dien et al., 2002; Gross et al., 2000; Musavi et al., 2002; Marullo et al., 2002; Cauwer et al., 2004; Dall’Olmo et al., 2005; Pozdnyakov et al., 2005). SeaWIFS satellite has a

spatial resolution of 1.1 km (Gross et al., 1999). Besides these satellites MERIS (Cauwer et al., 2004) with 300m spatial resolution, ASTER (Kishino et al., 2005) with 15-90 m spatial resolution, SPOT (Liu et al., 2003) with 20 m spatial resolution and IKONOS (Ekercin, 2007; Omerci et al., 2009) with 4 m spectral resolution have also been employed.

The subject, water quality monitoring, has been discussed in detail in the book “Remote Sensing and Digital Image Processing” which was published in 2005 (Moisan et al., 2005). Moisan et al. (2005) studied mainly on Case 1 waters using remote sensing. In the space there are 2 main satellites that were send in order to monitor water quality, SeaWIFS and MODIS. Most of the previous studies were conducted by these satellites. However in these studies beside remotely sensed imagery, numerical models were also required. First remote sensing studies for water quality monitoring stated in 1970’s with Bernstein et al (1977). The parameters, which are applicable for MODIS satellite, are Chlorophyll-a, TSS, CDOM that can be measured by empirical and semi-analytical algorithms. As well as empirical and semi-analytical model, there are deterministic models which are based on differential equations and numerical integration techniques.

Since Landsat 7 satellite was launched and became operational, remote sensing has been applied in water quality management studies actively (Wang et al., 2008). Remote sensing provided spatial and temporal distributions of environmental parameters in a more rapid and simultaneous manner. In water quality monitoring using remote sensing, mainly three parameters are optically active; chlorophyll-a, suspended solids and CDOM (Panda et al., 2004). Chlorophyll-a is being used as a primary pigment-index for phytoplankton concentrations and productivity. Suspended sediment is also another common type of pollutant in waters in terms of weight and volume (Baruah et al., 2001).

For modeling of chlorophyll-a concentration in both Case 1 and Case 2 water bodies, linear and non-linear empirical, semi-analytical and analytical models have been introduced. Empirical models are based on statistical relationship between parameters. Parameters that are used in empirical models are radiance leaving water, reflectance ratios of spectral bands and field measured chlorophyll-a concentrations. On the other hand, semi-analytical models uses reflectance values to relate essential optical properties of water bodies such as absorption and backscattering coefficients. Those coefficients are originating from presence of phytoplankton, colored dissolved organic matter, detrital material and suspended

particles (Garcia et al., 2006). In previous studies, most of the empirical models are based on in band ratios in several forms. Analysis of the previous studies indicate that empirical algorithms work better in Case 1 waters rather than Case 2 waters. Second types of models are semi-analytical models and they give better results in Case 2 waters compared to empirical models (Moisan et al., 2005). Use of ANNs in mapping of remotely sensed data into water quality parameters is initially practiced by Gross et al (2000) and Zhang et al (2003).

Liu et al (2003) mentions that chlorophyll-a mainly absorbs at 450-670 nm. High chlorophyll-a concentrations are results in higher absorbance in blue band with lower absorbance in green wavelength.

There are lots of studies that are considering correlation of chlorophyll-a with spectral reflectance obtained from satellite images in the literature. Some of the studies are specific for a location that is called local studies and some can be generalized to a variety of lakes called transportable studies. The transportable studies are not sensitive as local studies and they are spatially and temporarily independent. Transportable studies make routine studies applicable and can give idea about the changes within a time range.

In the study that was conducted by Blanco et al (2007), the reflectance of chlorophyll-a for different wavelengths were studied and it was found out that: chlorophyll-a absorbs 98% of incoming light in the wavelength range of 400 nm and 500 nm. Also, again in visible and near-infrared range this value goes up to 99%. However, 550 nm results in highest reflection due to lower absorption for green light.

Baruah et al (2001) and Sudheer et al (2006) reported that, in the previous studies chlorophyll-a was studied with regression, ratio models and primary component analysis (PCA) with different bands. Due to these linearity problems non-linear relations were searched with artificial neural network models (ANN). Also ANN is such a tool that provides relationship between set of variables as inputs and outputs. According to Slade et al (2002), algorithms retrieved from satellite could provide continuous mapping between satellites originated observation vector and geophysical vector of interest. As ANNs are capable of generating an interaction between two vectors without any initial knowledge, they can produce a continuous mapping between desired two data sets. Moreover based on the study by Panda et al (2004), R^2 values up to 0.94 was found for chlorophyll-a parameter in the remote sensing studies with ANN models while with linear regression only gives 0.31.

First studies regarding phytoplankton growth and their monitoring from space were studied by Ryther et al in 1957. Within the study, authors tried to relate the rate and amount of photosynthesis and phytoplankton concentrations in water bodies. They tried to estimate amount of phytoplankton by relating its concentration to rate of photosynthesis and light. For this study, they mainly dealt with temperature differences throughout the lake surface. In order to solve this problem, they took samples from different depth and surfaces where temperature was distributed homogeneously. They tried to improve the relationships, between chlorophyll-a and photosynthesis rate that were derived in 1940's and 50's. For this study, several samples were taken from coastal parts of Woods Hole located in the Massachusetts. For the chlorophyll-a extraction acetone method was used and its concentration was determined spectrophotometrically. In the final stage of the study, they attempted to find the phytoplankton growth by using radiation, transparency and chlorophyll-a concentration. They found out that only radiation values are needed besides SDD in order to measure growth rate of phytoplankton. This study was one of the keystones of water quality monitoring studies that relying on radiation or reflectance values, which leads us to remote sensing studies.

In 1998, Keiner modeled chlorophyll-a concentrations in the Delaware Bay by using artificial neural networks. For this purpose, Thematic Mapper sensor provided by Landsat Thematic Mapper was used. As regression analysis has problems in modeling non-linear situations, neural network models were tried to be used within this study. For this study multilayer, feed-forward, and error back propagation model was used. In every station chlorophyll-a was measured. Chlorophyll-a concentrations were in the range of 2.7 µg/L and 56.3 µg/L. Acetone extraction method was applied for the chlorophyll-a and concentration was measured flourometrically. In the analysis dark pixel subtraction method was used for atmospheric correction. This study contains both regression analysis and neural network modeling. For the regression analysis, pixel values were taken to be independent and chlorophyll-a concentration was taken as dependent variable. Results for both regression and neural network analysis for this study are given in Table 2-1.

Again in 1998 Wu studied Texas Gulf Coast for changes in the water quality parameters by using satellite imagery gathered from MODIS satellite. In the study PCA (Principal Component Analysis) method was applied in order to generate uncorrelated output bands from different images. The generated PC bands were linear correlations of original spectral bands and were uncorrelated. From the PCA

results it was found out that first four PC were defining 97.3% of the total dataset. Beside this first PC defines 87.5% of the dataset itself. As a result of the study it was found out that use of PCA eases the studies when highly correlated separate datasets are present.

As it was mentioned before ANN's can easily be applied to both case 1 and case 2 waters. In a study that was conducted by Gross et al in 1999, ANN model was used in order to determine the relationship between chlorophyll-a and spectral reflectance values taken from SeaWIFS sensor that is located in SeaStar satellite. SeaWIFS sensor is capable of measuring reflectance values in the 5 different wavelengths (412, 443, 490, 510 and 555 nm). In this study, only case 1 waters were used due to fact that they covers approximately 90% of all waters and chlorophyll-a is the only optical parameter that affects the reflectance values.

In the study, both input and output values were normalized in order to prevent numerical saturation in the both data sets (69 data pairs). Five different spectral bands provided 5 different input arrays and chlorophyll-a concentration provided output layer. Number of hidden layers and present hidden neurons were optimized by trial and error process and found as two hidden layers that contains six and four sigmoid (hidden) neurons respectively. Weights for the transfer functions were randomly chosen between -1 and 1. In addition to this learning, rate was set to be 0.01 throughout the training process. The results showed that the applicability of neural network models to case 1 waters is possible and easily applicable, the results of this study is given in the Table 2-1.

Vietnam Waters located in China were studied in 2002 by Dien et al. in order to determine the distribution of chlorophyll-a by using satellite imagery obtained from SeaWIFS satellite. Field works within the study continued for 2 years and samples were taken as they could represent the corresponding season. As a result of regression analysis that was applied for in-situ and SeaWIFS estimated data showed significant correlation with R^2 value of 0.719.

In the next year, 2000, Gross et al studied the transfer functions between chlorophyll-a and reflectance values in the ANN models. In order to obtain reflectance values, SeaWIFS sensor imagery with 5 bands were obtained. As a first stage of the study, the optical model "Morel" was used. Through the study variability in the phytoplankton type was neglected. So, reflectance values were assumed to be only affected by chlorophyll-a concentration. While using "Morel" model, there were sets of data were generated; one with no noise (Type 1) and one with noise

added by Gaussian expression (Type 2). In the second stage multi-layered perceptrons also called ANN models were tested for the optimal transfer function. As it was mentioned before transfer function between the layers of the network can be linear and non-linear (sigmoid). In this study, the neural network that was used by Gross et al in 1999.

In the order to train the model, ANN model was set to minimize the mean quadratic error by using least-square method. Next, models NN-1 with Type 1 data and NN-2 with Type 2 data were observed. It was found out that both sets of data can fit to polynomial pattern, but it was also mentioned that increasing in the chlorophyll-a concentration in NN-1 model results in worse fit in polynomial structure. Finally, in-situ data were taken into consideration. In this case, NN-2, which was trained with noisy data, gave better fit that NN-1 model in in-situ measurements.

Water quality parameters were studied in 2000, by using remote sensing technologies (SeaWIFS sensor) for coastal zone (case 1) management purposes. The shallow-water reflectance model (SWRM) used in this study. Durand et al observed feasibility and performance of linear and non-linear (ANN, simulated annealing and Chebyshev expansion) models in remote observation of chlorophyll-a. After the modeling stage, used algorithms were tested in the sensitivity analysis. So, it was found out that non-linear models were highly applicable in water quality purposes in remote sensing studies.

Baruah et al (2001) studied lake Kasumigara in Japan that is a eutrophic lake (Trophic State Index = 64, 90% of the lake is eutrophic). In the study water samples are taken from 29 different points. Atmospheric correction is applied by using dark-pixel subtraction method due to clear weather. Later steps land mask is applied by using NDWI. In the lake lowest chlorophyll-a concentrations are seen in January and February.

In the analysis Back-Propagation Neural Network is used for chlorophyll-a in order to increase the performance of the model. In the model only one hidden layer is used because higher number of hidden layers may lose the capacity to generalize by over fitting. The inputs that are reflectance values undergo two different functions before output values. First function provides summation that sums up the products of inputs and corresponding weights of the links. Later this sum is added to a bias and passes to a second function called squashing function.

As a result of the study, R^2 and MSE values of linear and logarithmic models are also tried and compared with results of the neural network model. Best regression models are given below. Results are given in Table 2-1.

$$\ln\big(Chlorophyll-a\ (\mu g/L)\big) = 48.31 - 1.52\,TM2 - 5.579[TM3/TM1]^2 \qquad \text{(Eq-1)}$$

Slade et al (2002) studied the interaction of remote sensing with ANN models in order to determine chlorophyll-a concentration. MLP (multi-layered perceptions) were trained with Levenberg-Marquardt algorithm. In this study a network with two hidden layers was chosen and searched for optimum number of hidden neurons. As a result of this study empirical (ensemble) (OC-4) models were compared with respect to ANN models. Results of the study are given in the Table 2-1.

Musavi et al (2002) tried to determine a relationship between observed chlorophyll-a concentration and pixel reflectance of SeaWIFS imagery. In this process both case 1 and case 2 waters were observed. This study also provides us to understand the difference between case 1 and case 2 waters in modeling point of view. In the context of this study, linear, empirical and non-linear models (2 semi-analytical, 15 empirical, artificial neural network) were tested. Used empirical models can be listed as: Morel 1, CalCOFI, CalCOFI cubic and OCTS-C.

Artificial neural network that was used in the study was trained by 90% of the data (919 pairs of reflectance). There was one hidden layer that contains 10 hidden neurons in the ANN model. This feed-forward model used hyperbolic tangent sigmoid functions for activation between layers. Results of the study for the empirical models and ANN are given in Table 2-1. (Musavi et al.,2002)

Mediterranean Sea was observed in 2002 by Marullo et al for chlorophyll-a retrieval. In this study SeaWIFS sensor was used and empirical algorithms were tested for oligotrophic and ultra-oligotrophic seas. In the studies, stations in the coastal regions were eliminated due to their possible effects as case 2 water bodies. After sampling from 45 station, extraction of chlorophyll-a was performed by using acetone method. In addition to this, 5 different bands (412, 443, 490, 510 and 555 nm) from SeaWIFS sensor were used. Finally Marullo et al found out that empirical models, which were tested, provided R^2 values higher than 0.93 with STD value of maximum 49 %.

Again in 2002, Zhang studied Gulf of Finland (case 1 water body) by using both satellite remote sensing and radar remote sensing. For this study Landsat TM (optical), ERS-2 SAR (radar) imagery, which were taken in 1997, were used in

separate studies and only empirical models were tested for their application to water quality management for coastal of areas of Finland. Both for satellite and radar imagery, irregularities in the water surface due to impurities and physical effects, optical properties are affected.

For the study in the Gulf of Finland, Zhang et al (2002) studied chlorophyll-a, by using empirical neural network model. In the field studies water temperature and salinity were also measured for 53 different sampling points. In the model there were 8 input arrays, 7 arrays for 7 different bands of Landsat TM and 1 band for SAR imagery. For one hidden layer, there were 5 hidden nodes (This value for hidden neurons was found with trial and error procedure from 2-10 hidden neurons). As activation function tanh and tansig were used. For the training of the model, values of 27 sampling point were used with Matlab Neural Network Toolbox and training process tried for 10 times with random weights. Levenberg-Marquardt method was used for backpropogation algorithm. After training, best weights with obtained smallest error was used for validation stage with remaining 26 data.

As a result of the study it was found out that chlorophyll-a had a low correlation with all bands of Landsat TM imagery. Empirical neural network model showed better results than regression models and the results are given in Table 2-1 (Zhang et al., 2002).

In 2003, Liu et al tried to quantify water quality parameters by remote sensing means with both empirical and semi-empirical approaches. It was mentioned that in shallower waters (like case 2 water bodies), it is hard to model water quality parameters with linear regression models. So, in most of the cases non-linear models (e.g. ANN) were used because empirical and semi-empirical models are generally site specific.

In this study, due to their higher spatial resolution, mainly imagery of Landsat and SPOT satellites were used in monitoring chlorophyll-a. As a result of the study it was found out that use of remote sensing imagery is an easy way to model water quality parameters by both linear and non-linear means.

Gulf of Bohai Sea in China was studied by Lin et al in 2003 for detection of algal blooms also called “red tide”. For this study both visible and near infrared regions of MODIS were used. After testing several models, algal blooms were detected empirically by use of “Normalized Difference Vegetation Index (NDVI)” in

which only Band 1 and Band 2 were used. As a result of the study, Lin et al (2003) proved that presence of algal bloom could be explained for positive NDVI values,.

Water quality parameters in the Beaver Reservoir (Northwest Arkansas) was modeled by Panda et al (2004) with ANN model and by use of remotely sensed images taken from Landsat TM in 2001 and 2002. Samples in the study were taken from surface, 1 m and 2 m below surface for 64 sampling points. Chlorophyll-a analysis was conducted in the lab for the water samples. In the chlorophyll-a analysis 90% acetone method was used. In the study both statistical and neural network models were tested that found out that there was no correlation between single DN and chlorophyll-a concentration. Other results are given in the Table 2-1.

Heege et al (2004) studied Lake Costance, which is located in the boarder of Switzerland, Germany and Austria, (Case 2 water) for chlorophyll-a. In this study modular inversion program (MIP) were used to observe its performance over satellite imagery. MIP is a specialized tool for modeling of hydrobiological parameters obtained from multi-spectral and hyper-spectral images. Program architecture includes algorithms that relate water quality parameters with their reflectance values. As the model has modular capabilities, different functions are easily used within the model.

For the application of the study, that was conducted by Heege et al (2004), airborne scanner, Daedalus AADS1268 (11 channels), was used. The data was obtained in a period of 6 years (1996 – 2002). In the modeling stage only 4 channels were used (2 – 5), that are located in the following wavelengths: 485, 560, 615, 660 nm. As a result of the study, applicability of the MIP was shown in water quality parameters. But, its applicability was limited.

In year 2004, Cauwer et al studied one of the main problems that were faced in the Southern North Sea (case 2), anthropogenic eutrophication due to excess loading of nitrogen and phosphate since 1950's. For this problem, they tried to find a way to monitor and model eutrophication situation by using remote sensing technology. For the chlorophyll-a concentrations SeaWIFS and MODIS imagery were used by obtaining blue/green band ratio. This ratio showed that in Case 1 waters, chlorophyll-a concentration gave its peak value at wavelength of 440 - 443 nm. For the field studies, 107 water samples were taken from 2 meter depth and analyzed for chlorophyll-a after filtering through glass-fiber filter by using high performance liquid chromatography (HPLC). Resulted chlorophyll-a concentrations were modeled by 4 different algorithms: SeaWIFS OC4v4 (Empirical chlorophyll-a

algorithm), MODIS Default (Empirical chlorophyll-a algorithm), MERIS (ANN) and GR03 (turbid water algorithm). Results for the corresponding algorithms are given in the Table 2-1.

As a result of the study, the difficulty of chlorophyll-a modeling in case 2 waters proved again after testing 3 different non-linear models.

In a study conducted in the Tokyo Bay by Kishino et al (2005), back propagation ANN was used in order to model the chlorophyll-a distribution in the water surface with respect to ASTER Imagery. Input parameters are taken to be radiances from different bands of ASTER/VNIR and zenith angle of the sun with output parameter, chlorophyll-a. Two hidden layers were chosen, one with 80 and the other with 45 neurons. It should be mentioned that ASTER satellite has no blue band that is an important concern for chlorophyll-a determination. Another point is that bandwidths of ASTER are large (80 to 200 nm) with respect to other ocean sensors (10 to 20 nm).

Again in 2005, Dall'Olmo et al studied applicability of red and near-infrared bands that are located in the SeaWIFS and MODIS sensors for the determination of chlorophyll-a concentration mainly in Case 2 waters. For the study, two different type of water bodies (sand pit and reservoir), which are located in the Eastern Nebraska from 251 sampling stations, were chosen. In the first phase of the study, reflectance values were retrieved by using a hand spectrometer in the wavelength range of 400 – 900 nm with a spectral resolution of 1.5 nm. In every sampling station water samples were taken. In the laboratory analysis, chlorophyll-a was determined fluorometrically after ethanol extraction. In the second phase of the study, data set were divided into two for training (136 data) and validation (115 data). By using training data set, non-linear least-squares best fit was applied between chlorophyll-a concentration and reflectance values of different bands. Then, this data set was trained for three different non-linear models. In the final stage trained models were tested for validation data sets and best results observed in this study are given in Table 2-1. As a result of the study it was found out that proposed models were applicable to determination of chlorophyll-a in both Case 1 and case 2 water bodies. It was reported that use of red and near-infrared bands in the chlorophyll-a determination provided valuable information based on characteristics of phytoplankton.

Another study in this topic was conducted by Sudheer et al (2006). Beaver Reservoir is Arkansas, USA (Case 2 water body) was studied by remote sensing for

chlorophyll-a concentration distribution. They used multilayer perception (MLP) with back-propagation algorithm. MLP technique involves finding optimal weight factor for the network as the other techniques mentioned before. Training network tries to provide a global solution to the weight matrix located the hidden layers. In this study there was only one hidden layer containing 10 hidden neurons.

Landsat TM bands were linearly correlated with chlorophyll-a concentration and best results were found to be 0.14 for TM1 and 0.31 for TM3. On the other hand for the ANN study, best fitting was observed by the combination of TM1 and TM2 bands. Addition of TM3 adds a negative potential. It was also observed that TM4, TM5 and TM7 bands did not contain much information chlorophyll-a modeling. Results for the study are given in the Table 2-1.

Cannizzaro et al studied remote sensing reflectance in shallow waters in 2006 by remote sensing in order to estimate chlorophyll-a concentrations. In this study, shipboard datum were obtained in 27 research trips to Tongue of Ocean, West Florida Shelf, Providence Channel, Great Bahamas Bank and shores of Lee Stocking Island between years 1998 – 2001. Samples, which were collected, analyzed fluorometrically for chlorophyll-a concentration after hot methanol extraction. Resulted chlorophyll-a concentrations were tested by several empirical and semi-empirical algorithms. Semi empirical model that was used relied on spectral reflectance in 550 nm and ratio of different spectral bands. Used model was mainly applicable to case 1 waters, but in this study it was modified for case 2 water by including parameters like water depth. As a result of the model run that contains 450 data for chlorophyll-a concentrations, ratios of the bands 443 and 490 nm gave better results than any other ratios tested for shallow Case 2 waters.

In 2007, Blanco et al studied Chesapeake Bay by using remote sensing techniques for chlorophyll-a concentration. For the remote sensing purposes both satellite and aerial photos were gathered in summer months for the specified sites. Their aim was to define a relationship between remotely sensed data and chlorophyll-a concentration. The results of this study were not shown in the conference though.

Pozdnyakov et al studied Great Lakes in 2005 for chlorophyll-a concentration by using satellite imagery obtained from SeaWIFS and MODIS. SeaWIFS Images were taken in July and September 2003. The images were analyzed with “MUMM-MSL12” software for atmospheric correction. In this study, both Levenberg-Marquardt and artificial neural network models were tested for applicability in

monitoring purposes. In this study, multi-layered perceptron neural network called “HDF2CPA” was used for correlation between in-situ and pixel values. For the transfer function between the layers, non-linear sigmoid function was used. In the neural network structure that was used there were three layers. There was equal number of neurons to number of spectral bands in the satellite. In the hidden layer, 30 hidden neurons were used and that was decided as a result of trials. Finally, in the output layer there was 3 neurons corresponds to chlorophyll-a, total suspended solids and dissolved organic carbon. Training was continued until rms error reached to 0.00049 as a result of 10,000 trials.

Golden Horn in Istanbul / Turkey was studied in 2007 by Ekercin in order to obtain a relationship between remotely sensed data retrieved from IKONOS-2 satellite and water quality parameters. For this purpose they have used multiple regression algorithms in order to explore the possible relationships between IKONOS data and water quality parameters. In the present data level of significance was found to be, $p = 0.05$. Water samples that were used in the study were taken from 9 permanent sampling stations. Analysis was considered changes in the chlorophyll-a concentration. For this analysis all of 4 bands were used and R^2 value was found to be 0.9924. Final analysis considered TSS concentrations and R^2 value between in-situ and model results was found to be 0.9724 (Table 2-1).

In China, optimization method for highly turbid water based on relations between the visible bands and in-situ data was applied. Models were generated for chlorophyll-a concentration. For the parameters ANN and semi-analytical models was also tried. (Le, 2009)

Again in China, Cui et al (2008) studied Bohai Sea after Lin et al (2003), for estimation of chlorophyll-a concentration by modeling it with artificial neural networks with respect to satellite originated reflectance data. In the corresponding study 50 different measurement stations were obtained and only 42 of them were used in this study due to excess amount of turbidity in other 8 stations. Gathered data set were divided into two part in which 32 of the stations were used in neural network training and testing and remaining 10 were used in NN model precision validation. For the modeling phase 3 layered backpropogation neural network were trained by inputting reflectance values and targeting chlorophyll-a concentrations. For the transfer function between the hidden layers hyperbolic tangent sigmoid and logarithmic sigmoid function, and for the output layer linear transfer function was used. In order to observe the performance of the model, three different parameter

used that are: root mean square Error (RMS), mean relative error (MRE) and Pearson correlation coefficient (R^2). As a result of the study, the neural network including four different bands (443 nm, 490 nm, 555 nm and 670 nm) as input, 3 layered including 7 hidden neurons was found to be the best model representing the situation with results given in Table 2-1. As a final stage for the study, noise analysis is also conducted in order to gather sensitivity of the model and the model were found to be stable for the noise.

Wang et al (2008) used artificial neural networks for the determination of chlorophyll-a concentration in Feitsui reservoir, located in China, by using Landsat ETM+ data. For this study 24 different water samples were taken from 0 and 0.8 m depth. Chlorophyll-a concentration in the samples were analyzed flourometrically after acetone extraction. In the next stage statistical analysis were applied to the dataset. As a result of the first linear regression analysis, highest correlation was found between band 7 and chlorophyll-a concentration that is 0.74. For the neural network training dataset 16 randomly selected data was used, while remaining 8 was used for testing purposes. There were 6 neurons in the input layer with 4 neurons in the hidden layer. Results for the studies are summarized in the Table 2-1.

In 2009, Yang and Du tried to model water quality parameters by remote sensing (Landsat 7 imagery) using artificial neural network approach in Mobile Bay. For this purpose they have conducted a regression analysis between image and chlorophyll-a concentrations. Results of this study are given in the following Table 2-1.

Golden Horn located in Istanbul / Turkey was studied again but this time by Ormeci et al. in 2009 for the determination of chlorophyll-a amount by using IKONOS satellite imagery. For this purpose multiple linear regression analysis was applied. As a result of the first study that was conducted with 9 sampling stations R^2 value between in-situ measurements and model estimated chlorophyll-a concentrations were found to be 0.88. In next 3 experiments, that was conducted with 8 different sampling points, maximum R^2 value was found to be 0.91. Final 3 experiments were conducted by using 7 sampling points and maximum R^2 value was calculated to be 0.97. Ormeci et al. (2009) point out the fact that both distribution and number of in-situ data effects regression analysis and the accuracy of the map.

Another study related to monitoring of water quality parameters by using remote sensing techniques were conducted in 2009 by Wu et al. In the corresponding study that was carried out in the Chaohu Lake in China data retrieved from MODIS satellite was used. On the other hand chlorophyll-a concentrations were measured in the field studies. The results of the field studies were analyzed with different bands of MODIS satellite by using back-propagation artificial neural network and genetic algorithm neural network while also tested with linear regression models. For the field study twelve different sampling sites were chosen and 48 different sets of MODIS data were used in model where 40 of them used for calibration purposes and 8 of them used for verification. In linear regression analysis Pearson correlation method was applied and significant relation ($P<0.01$) between reflectance values and water quality parameters were found. In neural network studies, 2 different methods were tested. As a result of the study they found out that methods can be ranked as following based on their precision in defining the dataset: Genetic Algorithm > Back-Propagation Neural Network > Multivariate Regression > Linear Regression.

One of the latest studies in this subject was conducted by Wang et al (2010). In this study Apalachicola Bay, located in Florida, was studied for temporal and spatial changes in the chlorophyll-a and TSS concentration by using MODIS imagery. For this study, 19 different sampling points were chosen for October 2002 and 11 points for April 2006 and water samples were taken from 0.5 depth of water column from every station and analyzed for chlorophyll-a concentration. In the next step, which is regression analysis, chlorophyll-a concentrations were chosen to be dependent variable and independent variables were reflectance of single band (or ratio), log transformed band ratio and logarithmically transformed band ratio. As a result of the study it was found out that chlorophyll-a concentrations were affected by physical factors like: Wind, tidal forcing, water flows, upland land use and changes in the land cover.

Table 2-1: Summary table for the previous studies

	Study	Chlorophyll-a (µg/L)		
		R^2	RMSE	MSE (%)
NN	Keiner, 1998	0.94	3.18	-
Regression		0.51	9.22	-
ANN 1	Gross et al., 1999	-	2.11	18.59
ANN 2		-	1.69	23.29
NN	Baruah et al., 2001	0.93	1.53	8.10
Regression		0.31	4.39	53.29
Empirical	Slade et al., 2002	0.35	1.27	-
ANN 1		0.50	0.87	-
ANN 2		0.63	0.79	-
Linear	Musavi et al., 2002	0.61	0.23	-
CalCOFI		0.92	0.19	-
CalCOFI Cubic		0.92	0.19	-
OCTS-C		0.93	0.19	-
Morel 1		0.92	0.18	-
ANN		0.93	0.17	-
NN	Zhang et al., 2002	0.92	0.47	-
Regression		0.67	0.96	-
Regression	Dien et al., 2002	0.72	-	-
NN	Panda et al., 2004	0.55	0.08	-
Regression		0.37	-	-
GR03	Cauwer et al., 2004	0.90	-	-
MODIS		0.52	-	-
OC4v4		0.60	-	-
MERIS			-	-
SeaWIFS (670, 765)	Dall'olmo et al., 2005	0.86	12.90	-
MODIS (667, 748)		0.90	11.00	-
MODIS (678, 748)		0.85	13.50	-
NN	Sudheer et al., 2006	0.53	12.34	-
Regression		0.09	-	-
Multiple Regression	Ekercin, 2007	0.99	-	-
NN	Cui et al., 2008	0.88	0.54	0.22
Statistical		0.74	0.99	0.28
Regression Radiance	Yang and Du, 2009	0.94	3.60	-
Regression Reflectance		0.93	0.74	-
ANN Training	Wang et al., 2008	0.72	-	-
ANN Testing		0.82	-	-

CHAPTER 3

METHODS and MATERIALS

In this study, the distribution of chlorophyll-a in Lake Eymir was modeled by regression and ANN models that use remotely sensed data and measured field parameters together as hybrid inputs. For this purpose, three different high-resolution satellite images that were taken at different seasons (September 2009, April 2010 and July 2010) were utilized. As a result, performances of the models were tested for different conditions. Since weather was mostly cloudy and there was no chance of obtaining a clear image, no winter time image was used. As will be discussed in detail in sub-section 3-2, field studies were conducted after the images were captured. In field, in-situ measurements were realized and water samples were taken at predefined sampling points for laboratory analysis. These points are located using a Global Positioning System (GPS) receiver. The satellite images were orthorectified by the private company beforehand and were analyzed for the reflectance values at corresponding sampling points. In order to calculate the reflectance, atmospheric correction was applied on the images. Later, all obtained data were compiled in an Excel file. XLStat software was used in order to apply Principle Component Analysis (PCA) on the dataset to determine which parameters would be used in modeling of chlorophyll-a concentrations in Lake Eymir. XLStat also provided the basic statistical analysis for the dataset. Then, chlorophyll-a concentrations were modeled with linear and non-linear regression models, as well as ANN models. Finally results were compared and maps were produced using a GIS software, ArcMap 10. Details of these steps are given in below sub-sections.

3.1. Information about Satellite Images

The first stage of the study was the acquisition of satellite imageries. In this study, 3 different high-resolution satellite images were used. The reason behind seeking high-resolution images was the size of the study area. As Lake Eymir has a surface of 1.25 km^2, lower resolution images might result in lower sensitivity in analysis. The first two satellite images were obtained from Quickbird 2 and the third one was obtained from World View 2 satellite. In Table 3-1 below, the imagery dates and their corresponding field work dates can be seen.

Table 3-1: Satellite Imageries and Field Work Dates

#	Satellite Name	Image Acquisition Date	Field Work Date
1	Quickbird 2	September 9th, 2009	September 18th, 2009
2	Quickbird 2	April 26th, 2010	May 2nd, 2010
3	World View 2	July 28th, 2010	August 2nd, 2010

In all field works there was a lag time between the image and field work dates. The delays in field work were due to the facts that the image acquisition dates were not announced beforehand, and when the images were taken, the information was not conveyed immediately. However, although the ideal case is to perform the field work on the date of image acquisition, in literature the time gap between imagery date and field studies for biological events such as algal blooms have been 7-10 days (Bricaud et al., 2002). Moreover, no significant weather variations were observed during the time periods between the image and field work dates. Therefore, it was deemed that during these periods, chlorophyll-a concentrations were not affected significantly. This assumption was validated by checking the meteorological conditions in Ankara between the image and field work dates via weather reports. These reports were obtained from WunderGround archive (Ankara, Turkey Forecast: Weather Underground, 2010). In Table 3-2, Table 3-3 and Table 3-4, weather reports are presented. In addition, it was assumed that during the time gap, there was no significant change in nutrient availability that would impact the growth of algae.

Table 3-2: Meteorological Data between September 9th and 18th, 2009.

	9.09	10.09	11.09	12.09	13.09	14.09	15.09	16.09	17.09	18.09
Mean Temperature (C)	18	16	18	20	18	16	16	16	15	16
Max Temperature (C)	24	25	26	23	22	23	23	23	23	24
Min Temperature (C)	11	7	11	16	15	10	8	10	7	7
Precipitation	1	0	0	0.1	0.3	0	0	0	0	0
Dew Point (C)	10	9	10	11	13	9	9	8	6	4
Average Humidity	62	60	58	61	76	64	67	59	56	48
Maximum Humidity	82	88	88	82	94	94	94	88	93	87
Minimum Humidity	36	22	24	32	43	29	26	19	20	14
Sea Level Pressure (hPa)	1015.08	1016.33	1012.51	1009.98	1011.79	1019.27	1019.96	1017.65	1016.81	1014.08
Wind Speed (km/h)	9	7	9	14	10	7	5	9	6	5
Max Wind Speed (km/h)	30	24	17	28	28	18	15	26	18	18
Visibility (km)	10.6	10.5	10.6	10.8	10.3	10.6	11	10.9	11.7	11.4

Table 3-3: Meteorological Data between April 26th and May 2nd, 2010.

	26.4	27.4	28.4	29.4	30.4	1.5	2.5
Mean Temperature (C)	13	12	10	10	10	10	12
Max Temperature (C)	19	17	17	17	18	19	21
Min Temperature (C)	7	6	3	3	3	1	2
Precipitation (cm)	0	0	0	0	0	0	0
Dew Point (C)	-1	1	2	1	-2	-6	1
Average Humidity	41	49	56	50	44	36	47
Maximum Humidity	76	71	81	81	87	70	81
Minimum Humidity	10	20	32	18	10	4	15
Sea Level Pressure (hPa)	1018.86	1018	1016.31	1015.85	1014.67	1014.23	1014.73
Wind Speed (km/h)	17	18	14	11	15	9	7
Max Wind Speed (km/h)	28	33	24	20	24	18	15
Visibility (km)	11	10.7	10.9	10.9	11.6	11.3	11.4

Table 3-4: Meteorological Data between July 28th and August 2nd, 2010.

	28.07	29.07	30.07	31.07	1.08
Mean Temperature (C)	26	26	26	28	28
Max Temperature (C)	34	30	34	37	38
Min Temperature (C)	17	21	17	18	19
Precipitation (cm)	0	0	0	0	0
Dew Point (C)	14	15	14	12	8
Average Humidity	49	56	50	42	29
Maximum Humidity	82	73	77	78	56
Minimum Humidity	12	32	20	9	6
Sea Level Pressure (hPa)	1011.5	1012.94	1014.58	1015.04	1013.56
Wind Speed (km/h)	6	16	16	9	6
Max Wind Speed (km/h)	20	26	24	18	18
Visibility (km)	12	11	12.7	12.1	12.4

The Quickbird Satellite is developed to be an agile, stable platform with high accuracy for remote sensing of the Earth. It is an implementation of BGIS 2000, which is a Ball Aerospace Global Imaging System. This system allows gathering panchromatic images with 0.61 m pixel sizes and multispectral images with 2.5 m pixel sizes. Quickbird has a swath width of 16.5 km, large field of view (FOV), high contrast (MTF), high signal to noise ratio and 11 bit dynamic range. The Quickbird satellite was launched in 18 October 2001 and today it is operated by the Digital Globe Company. It has an orbital altitude of 450 km and it is mainly used for

mapping, military surveillance, weather research and environmental monitoring purposes. Its revisiting period was given as 2-3 days depending on the latitude with an orbital period of 93.4 minutes (Quickbird, 2008). Information about the spectral bandwidths of the Quickbird satellite is given in Table 3-5:

Table 3-5: The Spectral Bandwidths of the Quickbird Satellite

Bandwidth	Upper Wavelength (nm)	Lower Wavelength (nm)
Blue	450	520
Green	520	600
Red	630	690
Near – Infrared	760	900

World View 2 is the other satellite whose image was also used in this study. This satellite was launched in October 2009 and became online in January 2010. The World View 2 has high resolution in both multispectral and panchromatic bands. Resolutions are 1.84 m for multispectral and 0.46 m for panchromatic images. Beside the standard 4 spectral bands (Blue, Green, Red, Near Infrared), it has 4 additional ones (Coastal Blue, Yellow, Red Edge, and Near Infrared 2). The band windows are shown in Figure 3-1 in comparison to the band windows of QuickBird Satellite. The World View 2 has a swath width of 16.4 km with 11-bit of dynamic range. Its orbital altitude is set to be 770 km with the revisiting period of 3.7 days. (World View 2, 2009).

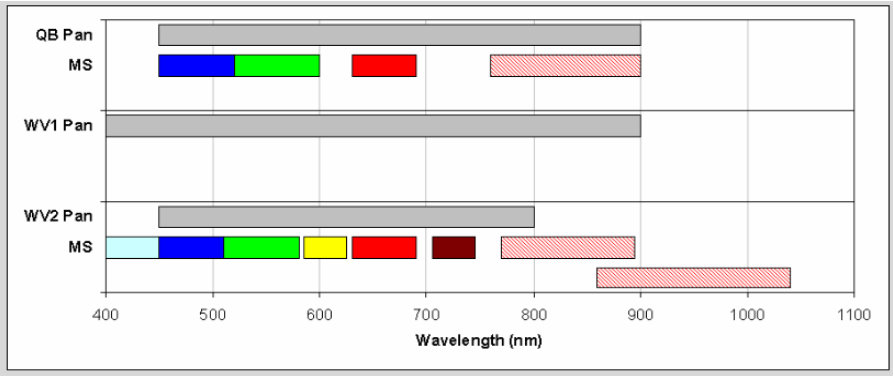

Figure 3-1: The Spectral Bands for Quickbird 2 and World View 2 Satellites

The first satellite imagery of Lake Eymir was taken by Quickbird 2 on September 9th, 2009 from an altitude of 450 km. The image was taken at 8:45:47 AM (GTM+2). At this time, the cloud cover over the whole image area was given as 1.3%, while there was no cloud present over the lake. The second image was taken by Quickbird 2 as well on April 26th, 2010 from an altitude of 450 km. This image was taken at 8:34:31 AM (GTM+2). The cloud cover was reported to be 0 %. The third and the final image was taken by World View 2 on July 28th, 2010 form an altitude of 770 km. The image was taken at 9:10:04 AM (GTM+2). The cloud cover was 0%. The satellite images are depicted in Figure B-1, Figure B-2 and Figure B-3. From herein, the images taken on September 9th 2009, April 26th 2010, and July 28th 2010 will be referred to as the first, second, and third images, respectively. Same chronology will be used for the corresponding field works (Table 3-1) as well.

All images were obtained as multispectral bundles and as pan sharpened images. Multispectral images captures reflected data at specific wavelengths in electromagnetic spectrum. Sensors are sensitive to different wavelengths in visible range (450 nm – 690 nm) and near infrared range (750 nm – 900 nm). By using multispectral images, details that cannot be identified by human eye can be retrieved. Panchromatic images are sensitive to all wavelengths of the visible portion of electromagnetic spectrum. Also in satellite imageries, panchromatic images have higher spatial resolution than multispectral images (Hough, 1991). Retrieved images had a UTM projection zone 36 with WGS 84 datum, while purchased images had Geographic Projection and European Datum 1950. For re-sampling in the imagery, nearest neighbor method was used in order to avoid any change in the reflectance values.

3.2. Field Study

Following the acquirement of the images, field works were conducted on the dates presented in Table 3-1. Sampling points in the field were determined beforehand and located in the field using a Magellan Sportrak GPS Receiver (±3 m accuracy) (GPS Review – Magellan SporTrak, 2004). The accuracy of the GPS receiver depends on several factors including satellite position, presence of noise in radio signal, interference from something that has the same operation frequency, atmospheric conditions such as charged particles and water vapor, surrounding mountains and buildings, and atomic clocks of satellites. A GPS system requires 4

different satellites in order to work properly. In a GPS system, each satellite transmits a signal with specific intervals that identifies its location. Each signal that is received by a GPS receiver arrives with a small time difference. The distance between the receiver and the corresponding satellite is determined by calculating the time lag between each signal. As the receiver calculates the distances with respect to 4 different satellites, it can calculate its position in a 3-D space by intersecting the imaginary spheres that are centered by each satellite (How Does GPS Work?, 1998).

In field studies, the following activities were conducted.

- Conducting in-situ measurements at sampling points for determination of:
  - Temperature
  - pH
  - Conductivity
  - Dissolved Oxygen
  - Turbidity
  - Wind speed
  - Relative Humidity
  - Wind direction
  - Air temperature
  - Dew point temperature
  - Photosynthetically active radiation (PAR)
  - Secchi disc depth
  - Water column depth

  Water sampling at 2 depths (at lake surface and at 0.50 m depth) at sampling points using 500 mL sampling bottles for determination of:
  - Chlorophyll-a
  - Total suspended solids
  - Color

Parameters that are measured throughout this study were chosen based on their potential effects on reflectance values and chlorophyll-a concentrations. For example, chlorophyll-a concentration is known to change based on varying humidity in the surrounding atmosphere (Chappelle et al., 1992). Additional to water quality parameters, atmospheric parameters were also included in this study due to their

impacts on chlorophyll-a concentrations and reflectance values. For instance, wind direction and wind speed are effective over reflectance values.

3.2.1. Locations of Sampling Points

The number of sampling points was chosen so as to cover the whole area of the lake with sufficient precision. In previous studies on remote sensing of water quality parameters, the density of sampling (the ratio of sampling points per square kilometer) ranged between 0.002 samples/km^2 (Zhang et al., 2002) to 3.2 samples/km^2 (Elahdab, 2006). In this study, 30 sampling locations were employed over the lake area of about 1.25 km^2, which corresponds to a sampling density of 24 samples/km^2. The locations of these sampling points are chosen arbitrarily to evenly cover the study area. Similar sampling locations were utilized for the first and second field studies as depicted in Figure 3-2. For the third field study, the sampling points presented in Figure 3-3 were employed. The adjustment was made to have a more homogeneous distribution of sampling points. Since a boat was used as the sampling vehicle, the sampling points were chosen to be away from the shore, to avoid the interference of rooted aquatic plants and shallow depths with sampling activities. The locations and coordinates of the sampling points depicted in Figures 3-2 and 3-3 are also given in Tables 3-6 and 3-7, respectively.

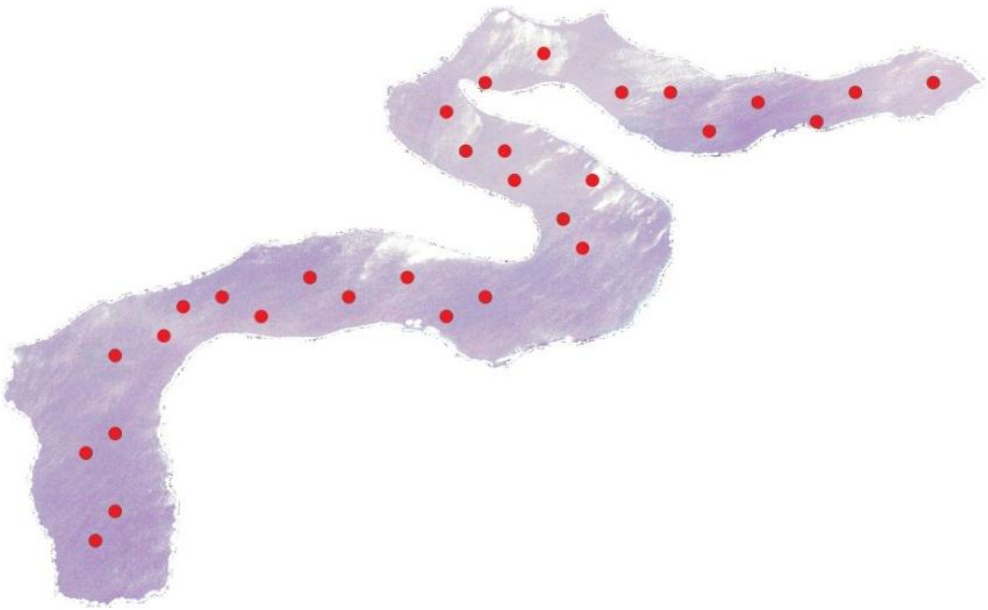

Figure 3-2: Sampling Points for Field Study 1 and 2

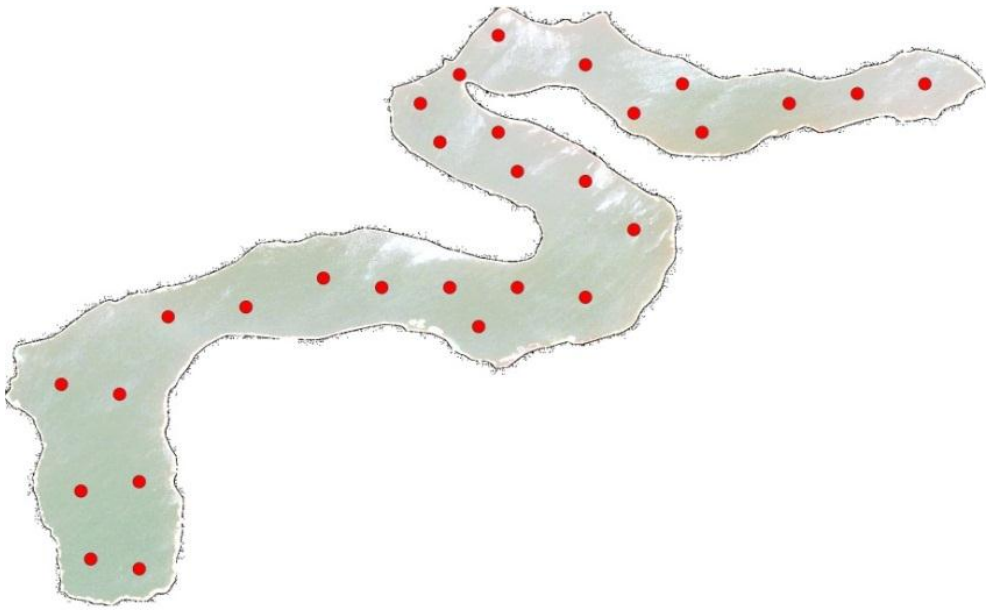

Figure 3-3: Sampling Points for Field Study 3

Table 3-6: Sampling Points Coordinates for Field Study 1 and 2

	North			East				North			East		
	Deg	Min	Sec	Deg	Min	Sec		Deg	Min	Sec	Deg	Min	Sec
Point 1	39	49	5	32	49	7	*Point 16*	39	49	38	32	49	55
Point 2	39	49	8	32	49	9	*Point 17*	39	49	42	32	49	58
Point 3	39	49	14	32	49	6	*Point 18*	39	49	42	32	49	50
Point 4	39	49	16	32	49	9	*Point 19*	39	49	45	32	49	49
Point 5	39	49	24	32	49	9	*Point 20*	39	49	45	32	49	45
Point 6	39	49	26	32	49	14	*Point 21*	39	49	49	32	49	43
Point 7	39	49	29	32	49	16	*Point 22*	39	49	52	32	49	47
Point 8	39	49	30	32	49	20	*Point 23*	39	40	55	32	49	53
Point 9	39	49	28	32	49	24	*Point 24*	39	49	51	32	50	1
Point 10	39	49	32	32	49	29	*Point 25*	39	49	51	32	50	6
Point 11	39	49	30	32	49	33	*Point 26*	39	49	47	32	50	10
Point 12	39	49	32	32	49	39	*Point 27*	39	49	50	32	50	15
Point 13	39	49	28	32	49	43	*Point 28*	39	49	48	32	50	21
Point 14	39	49	30	32	49	47	*Point 29*	39	49	51	32	50	25
Point 15	39	49	35	32	49	57	*Point 30*	39	49	52	32	50	33

Table 3-7: Sampling Points Coordinates for Field Study 3

	North			East				North			East		
	Deg	Min	Sec	Deg	Min	Sec		Deg	Min	Sec	Deg	Min	Sec
Point 1	39	49	3	32	49	6	*Point 16*	39	49	42	32	49	57
Point 2	39	49	2	32	49	11	*Point 17*	39	49	43	32	49	50
Point 3	39	49	10	32	49	5	*Point 18*	39	49	47	32	49	48
Point 4	39	49	11	32	49	11	*Point 19*	39	49	46	32	49	42
Point 5	39	49	21	32	49	3	*Point 20*	39	49	50	32	49	40
Point 6	39	49	20	32	49	9	*Point 21*	39	49	53	32	49	44
Point 7	39	49	28	32	49	14	*Point 22*	39	49	57	32	49	48
Point 8	39	49	29	32	49	22	*Point 23*	39	49	54	32	49	51
Point 9	39	49	32	32	49	30	*Point 24*	39	49	54	32	49	57
Point 10	39	49	31	32	49	36	*Point 25*	39	49	49	32	50	2
Point 11	39	49	31	32	49	43	*Point 26*	39	49	52	32	50	7
Point 12	39	49	27	32	49	46	*Point 27*	39	49	47	32	50	9
Point 13	39	49	31	32	49	50	*Point 28*	39	49	50	32	50	18
Point 14	39	49	30	32	49	57	*Point 29*	39	49	51	32	50	25
Point 15	39	49	37	32	50	2	*Point 30*	39	49	52	32	50	32

3.2.2. In-Situ Measurements

A YSI 6600 EDS multi-parameter sonde (Figure 3-4) was used in order to measure pH, temperature (^{o}C), conductivity (µs/cm), DO (mg/L), and turbidity (NTU) with respect to depth. The sonde was lowered in the water column slowly and readings were recorded at every 3 seconds. The water column depth is measured by the use of water pressure over the sonde. Use of the multi-parameter sonde provided a lot of advantages in field work including reduced time for in-situ measurements since a number of parameters were recorded at the same time and it acted as a data storage device.

Calibrations for each parameter were conducted before field work. Calibration standards were supplied from the Hach Lange Company. For pH, 3-point calibration (pH=4, 7 and 10) was applied. For conductivity, again 3-point calibration was employed using 2 standard solutions and ultrapure water (for zero value). For turbidity calibration, hydrazine sulfate and hexamethylenetetramine solution (with 4000 NTU) was prepared and diluted for 100 and 250 NTU values.

Ultrapure water was used for 0 NTU value. DO calibration was conducted at field. Saturated water vapor pressure was used for the automatic calibration of dissolved oxygen.

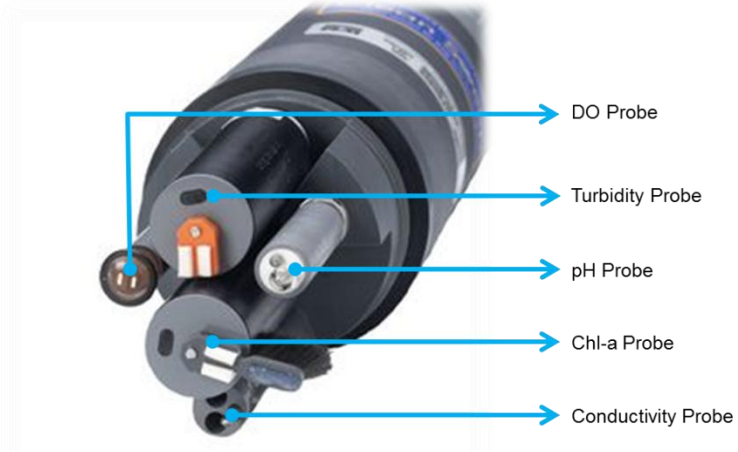

Figure 3-4: YSI 6600 EDS Probe

The operating principle of the pH probe on YSI sonde is similar to a battery. Voltage is generated between two points based on the ionic concentration in water. The inside of the electrode is filled with a neutral solution. The outside structure of the pH probe is a glass membrane that allows passage of H^+ ions. When the probe is submerged to an acidic environment, H^+ ions passes through inside and generates a positive charge around the electrode. This increase is measured and reported as the pH. In the case of a basic solution, H^+ ions present inside the probe in neutral environment passes through the glass membrane to outside and leaves a negative charge around the electrode. As a result, the pH value is calculated by the decrease in the total charge.

In aqueous environments electricity is conducted to some extent. Conductance means the ability of liquid to conduct electricity. The conductivity sensor on the YSI sonde is made up of two separate electrodes and a temperature sensor. When both of the electrodes are submerged to a solution, a low voltage is given to the liquid, and the magnitude of generated current is measured as the conductivity. Temperature is effects conductivity.

The DO probe is composed of a membrane, a platinum electrode and a silver electrode. The idea behind the operation is to measure the difference between oxidation reduction potentials. Potassium chloride solution is used in order to provide sufficient conductivity. Working principle of the DO probe is based on

applied voltage to the electrodes. A current is generated between two electrodes and oxygen that passes through the membrane is reduced in the potassium chloride solution. The generated potential due to reduced oxygen is measured and converted to the amount of DO.

Turbidity is measured by the scattering of light. There are two parts (light source, receiver) installed on the turbidity probe. The first part generates the light and the second part collects the light that is reflected from the suspended particles. The ratio between the generated and received lights is calculated as turbidity value.

Anemometer was the another equipment used for in-situ measurements. It was used in order to measure weather related parameters. In this study, a Kestrel 4500 multi-parameter anemometer (Figure 3-5) was used in order to measure the parameters related to weather conditions. This instrument can measure wind speed, air temperature, wind chill, relative humidity, heat stress index, dew-point temperature, wet bulb temperature, altitude, wind direction, head and tail wind. Kestrel 4500 has a capability of storing 1400 data in its memory which makes it useful equipment for field work.

Figure 3-5: Kestrel 4500 Anemometer

Photosynthetically active radiation (PAR) is an important parameter that affects algal growth (Spears et al., 2008). PAR can be defined as the light that is available for photosynthesis. This parameter is measured by a LICOR LI-193SA PAR sensor (Figure 3-6) with respect to the depth of water column.

Figure 3-6: PAR Sensor

Secchi disc measurements were made using a standard black and white disc of approximately 20 cm diameter. It was lowered to water and the depth at which black and white parts are no more distinguishable was recorded as the Secchi Disc depth in units of meters. Secchi depth indicated the depth light was attenuated. The Secchi depth value mainly affected from the presence of suspended particles in water, which may also impact the reflectance values in pixels of the satellite images. The instrument that is used for Secchi disc depth determination can be seen in Figure 3-7. This disc is also used in measurement of the depth of water at sampling points. The disc was lowered until it touched to the bottom and the depth was recorded.

Figure 3-7: Secchi Disc

3.2.3. Water Sampling

Water sampling in the field is conducted using Van-Dorn horizontal water column sampler as seen in Figure 3-8. This equipment allows obtaining water samples at given depths. Caps are present on both sides. The sampler is submerged to water to the desired depth. Then, a heavy subject on the rope, called as the messenger, is released which hits on the assembly that keeps the caps open and closes them. As a result, water that is present in the sampler is contained within. The water sample in the Van-Dorn is then taken to the boat and the sample is transferred to numbered plastic sample bottles. In this study, samples were taken at each sampling points at two depths; surface and 0.5 m below surface.

Figure 3-8: Van Dorn Water Sampler

3.3. Laboratory Analyses

The water samples taken by the Van Dorn sampler are analyzed for

- Chlorophyll-a
- Total suspended solids, and
- Color

at the laboratory using the standard methods provided below. In spectrophotometric analyses, the DR/2400 Hach Lange Spectrophotometer was used. Weighing of the samples is realized by Sartorius Basic (BA210S).

3.3.1. Chlorophyll - a

The chlorophyll-a concentrations were determined by ISO 10260, 1992 Standard Ethanol Extraction Method. The summary of the steps followed in this spectrophotometric method is given below.

- 250 mL water sample is taken from the well mixed sample,
- Water sample is filtered through a glass-fiber filter,
- The filter is inserted into the chlorophyll-a jar and 20 mL of pure ethanol is poured over it,
- The jar is heated to 76 ^{o}C in a hot water bath for 15 minutes,
- Then, it is cooled down to room temperature,
- The filter is removed from the jar and sample is taken into a 1 cm x 1 cm rectangular for determination of absorbance.
- Absorbance values of the sample is read and recorded at 665 and 750 nm against ethanol blank,
- Then, 0.1 ml, 0.1 N HCl acid is added to the sample and kept for 35 minutes for reaction,
- Absorbance values for the acidified sample is read and recorded at 665 and 750 nm values against ethanol blank,

For the calculation of chlorophyll-a concentration (in micrograms per liter), the following equation is used.

$$Chlorophyll - a\left({}^{\mu g}/_{L}\right) = {}^{(A - A_a) * 29.6 * V_e}/_{V_n * 1} \quad \text{(Eq- 2)}$$

where,

A = Absorption difference between 665 nm and 750 nm before acid addition (-)

A_a = Absorption difference between 665 nm and 750 nm after acid addition (-)

V_n = Sample volume (L)

V_e = Ethanol volume that is used per sample (mL)

3.3.2. Total Suspended Solids

The total suspended solids concentration was determined using the standard method 2540D (Greenberg et al., 1998). The steps of this gravimetric method are summarized below.

- 2.5 - 200 mL water sample is taken from the well mixed sample. The taken sample volume is recorded as “V” in mL,
- 0.45 µm Whatman filter is dried in the drying oven (103 – 105 ^{o}C) for 1 hour and then kept in desiccator until the analysis is performed,
- Before the analysis, dried filters are weighted in an analytical balance and the weights are recorded as “B” in milligrams,
- The water sample is filtered through the filter,
- Filter containing the particles on top is dried in the drying oven (103 – 105 ^{o}C) for 24 hours and then kept in the desiccator for 1 hour,
- The dried filter containing the particles is weighted in the analytical balance and the weight is recorded as “A” in milligrams.

For the calculation of suspended solids concentration in the water sample (in milligrams per liter), the following equation is used.

$$Total\ Suspended\ Solid\ Concentration\ \left(\frac{mg}{L}\right) = {(A-B)*1000}/{V} \qquad \text{(Eq-3)}$$

where,

A = Weight of dried filter containing filtered particles (mg)

B = Weight of dried filter with no particles (mg)

V = Sample volume (mL)

3.3.3. Color

Color is measured only for the first field study. Based on PCA results, this parameter was omitted from other field studies in order to decrease sampling requirements and analysis costs. Color can be defined as the apparent or true colors. In the apparent color, there are some dissolved particles and suspended materials. So, by filtering or centrifuging, the true color of the sample can be obtained. In this study, the true color determination was applied based on the methodology given for Platinum-Cobalt Standard Method in Standard Methods for the Examination of Water and Wastewater and NCASI, Technical Bulletin No. 253 (1971). For the analysis DR/2400 Hach Lange Spectrophotometer is used. Steps for the spectrophotometric analysis are given below.

- Color program 120 is chosen on the Hach Spectrophotometer for analysis at 455 nm,
- For blank, 50 mL of deionized water is filtered through 0.45 µm Whatman filter and a cuvette is filled with about 10 mL of it,
- 50 mL of sample is filtered through 0.45 µm Whatman filter and a cuvette is filled with about 10 mL of this filtered sample,
- The blank is placed in the cell holder and the spectrophotometer is set to zero by reporting 0 PtCo
- The cuvette containing the filtered sample is placed into the cell holder and color is reported as PtCo.

3.4. Image Analysis

Image analysis is conducted to obtain the reflectance values at the given pixels of the satellite images.

In order to analyze the images for digital numbers and present interferences, ENVI software was used. The unused area in the satellite image was removed and only the lake area was taken as an ROI (Region of Interest). ROI is a subset in whole dataset which is identified for a specific purpose. Therefore, the lake area was cropped and isolated. ENVI was used in order to obtain DNs (Digital Numbers) of different bands of the image

ENVI was used in order to gather pixel and radiance values from the geospatial image of Lake Eymir. ENVI was also used to represent DEM (Digital Elevation Model) and stereo (3-D) image. Corresponding stereo and DEM images are given in Figure B-4, Figure B-5 and Figure B-6.

Presence of aerosols, humidity and particles in the atmosphere lead to changes in spectral reflectance. Atmospheric correction is applied to obtain the ground leaving reflectance from sensor radiance. In order to eliminate the atmospheric effects, atmospheric correction is applied by using ENVI and Microsoft EXCEL. There are two main steps in the atmospheric correction process. In the first stage, properties of atmospheric constituents are taken into consideration. In the second and the final stage those properties are used in order to calculate the

surface reflectance by using transfer algorithms. As a result of atmospheric correction following problems can be overcome.

- Radiance difference originating from different sensors in different satellites.
- Change in the sensed reflectance due to atmospheric conditions like cloud cover and particles present in the atmosphere.

For atmospheric correction, dark object subtraction method (Eq-4) was used on all bands. In order to obtain reflectance values, first histograms of the image were generated for the corresponding bands. Later, zero values in the histogram were removed from the dataset. Zero values interfere with the real reflectance values since it represents perfect absorption or perfect reflection. Finally, the minimum and the maximum of histogram values were chosen to aid in conversion of DNs in pixels to reflectance values (Beisl et al., 2008). The following equation represents the conversion calculation.

$$Reflectance\ Value = \frac{Histogram\ Value - Minimum\ Histogram\ Value}{Maximum\ Histogram\ Value} \quad \text{(Eq-4)}$$

Histogram values were matched to DN values in the range of 0-255. So, by generating a band-specific linear equation, DN values were converted to reflectance values that are used in the modeling of chlorophyll-a.

3.5. Data and Principle Component Analysis

Field data and remotely sensed data are analyzed statistically to evaluate the relationships between parameters. Furthermore, the minimum, maximum, mean and standard deviations for the measured parameters were determined. For this purpose the data analysis tool of Microsoft EXCEL was used. Moreover, the spatial distribution of field parameter values was mapped using ArcMap 10.

Following the basic statistical analysis, PCA was conducted using XLStat which can be integrated to Microsoft EXCEL as an add-in. PCA helps to identify effective factor groups that have the most explanatory characteristics over the system. PCA can be defined as a method for reducing the dimensionality by applying covariance analysis between factors. In addition to this, a principle

component is a linear combination of weights that are optimized for observed parameters. Since the numbers of parameters that can potentially be used as the independent variables of the regression models are high, PCA is applied in order to identify the importance of parameters in defining the system. Moreover, PCA is applicable in cases where unknown trends present in a dataset needs to be uncovered.

3.5.1. Modeling of chlorophyll-a using hybrid inputs

3.5.1.1. Linear and Non-Linear Regression Models

Linear and non-linear regression models were applied in order to observe the performance of these approaches in modeling of the chlorophyll-a concentrations using hybrid inputs. These models are more traditional to ANNs. Therefore, results obtained from regression models also constitute a baseline for comparison for ANNs. Both linear and non-linear regression models were studied using XLStat. For both, two input data cases were considered. These were;

- PCA eliminated hybrid data (band reflectance and field parameter values)
- Band reflectance values only

The basic structures for the linear and non-linear regression models are provided in Equations 5 and 6, respectively.

$$Chlorophyll - a\ (\mu g/L) = C + \sum_{i=1}^{11}(P_i * X_i) \quad \text{(Eq-5)}$$

$$Chlorophyll - a\ (\mu g/L) = \sum_{i=1}^{11}\left(P_i * {X_i}^{K_i}\right) \quad \text{(Eq-6)}$$

where,

C = constant

P_i = regression coefficient for parameter i

X_i = value of parameter i

K_i = exponent for parameter X_i

In above equations, C , P_i and K_i values are set by XLStat software while minimizing RMSE and maximizing R^2 values between observed and predicted chlorophyll-a concentrations. X_i values represent the model parameters. Based on the limitations in the XLStat software, up to 11 different input parameters could be used in regression model development.

3.5.1.2. MATLAB Neural Network Models

ANN models were derived using "MATLAB Neural Network Toolbox". As mentioned before, ANN models are developed through generating a relationship between inputs and outputs by changing the weights that connect the neurons in different layers to each other through weighted links. Similar to biological neural network, artificial neural networks are designed to learn in order to find best solutions. Its behavior can be simplified as presence of independent elements, their connections and weights of the connections. In MATLAB, weights are automatically adjusted in training stage by using specific learning rule.

The neural network toolbox in MATLAB has its own in-line functions and graphical representations for created network. After network is created it can be trained and simulated with respect to given data. In ANN model development, all input and output datasets were normalized between [0,1] or [-1,1]. The data set was divided into three sets: training, test, and validation. Training data set is used to derive the pattern by use of training functions. Training functions that are present in the MATLAB library are mathematical expressions which allows user to adjust networks weights and bias values automatically and globally. In the structure, weights are used to improve the learning stage. Besides weights, bias values allow small shifts in learning function for a better fit. In neural networks, one of the ways to minimize errors is to update weights and biases based on output values. This update is provided by back-propagation algorithms. In back-propagation, there are two main steps. The first stage is the propagation phase. In propagation phase, MATLAB uses the training dataset as the input data and calculates the outputs. In the output calculation activation function controls the strength of output with providing non-linearity. Then, delta values (error) between input, hidden and output data are calculated. Delta calculation is provided by using Delta rule which is a gradient decent learning rule. By use of gradient decent rule, local minimum in a solution set is found by following the negative steps in gradient direction. In the

second stage, weights are updated based on the calculated delta values. To achieve this delta value and result of activation function of the input are multiplied for obtaining gradient of weight dataset. Finally weight value is updated in the opposite direction of the gradient. Because of this different training algorithms provide different learning rates.

Neural network toolbox has several built-in back-propagation training algorithms. Built-in algorithms are listed below (Demuth et al., 2009):

- Batch training with weight and bias learning rules (trainb)
- Powell-Beale conjugate gradient back-propagation (traincgb): This algorithm updates weights and bias values based on conjugate gradient back-propagation with Powell-Beale restart.
- Fletcher-Powell conjugate gradient back-propagation (traincgf)
- Polak-Ribiére conjugate gradient back-propagation (traincgp): This algorithm updates weights and bias values based on conjugate gradient back-propagation with Polak-Ribiére updates.
- Gradient descent back-propagation (traingd): In this algorithm weights and biases are updated based on the direction of negative gradient of the performance function. The performance of this algorithm depends on the number of trials, computation time, minimum gradient, maximum error and the learning rate.
- Gradient descent with adaptive learning rule back-propagation (traingda): This algorithm updates weights and bias values based on gradient decent with adaptive learning rate.
- Gradient descent with momentum back-propagation (traingdm): This algorithm updates weights and bias values based on gradient decent with momentum.
- Gradient descent with momentum and adaptive learning rule back-propagation (traingdx): This algorithm combines traingda and traingdm and updates weights and bias values based on gradient decent with momentum and adaptive learning rate.
- Levenberg-Marquardt back-propagation (trainlm): This algorithm updates weights and bias values based on Levenberg-Marquardt optimization.
- One Step Secant Algorithm (trainoss): This algorithm updates weights and bias values based on one-step secant method.

- Resilient back-propagation (trainrp): This algorithm updates weights and bias values based on resilient algorithm.
- Scaled conjugate gradient back-propagation (trainscg): This algorithm updates weights and bias values based on scaled conjugate gradient method.

The learning rate of ANNs is also affected by the use of transfer functions. Transfer function is mathematical representation of relationships between input and output data. There are three main transfer functions present in the toolbox: linear, threshold and sigmoid. The linear transfer function is used for conditions, where input values are proportional to outputs. The threshold transfer function is used where output can be changed based on previously defined threshold value. Sigmoid transfer function is used when output value changes continuously in non-linear manner. Also MATLAB provides sub-training sigmoid transfer functions to relate input and output data as tangent or logarithmic. Even if output and input values changes between minus infinity and plus infinity, the transfer functions allow the network to operate in a predefined range. The tangent sigmoid function generates relationship between input and output values scaled in the range of -1 to 1. On the other hand, logarithmic sigmoid transfer function provides a distribution between 0 and 1 (Demuth et al., 2009).

In this study, a set of back-propagation training algorithms were tested using an automated ANN architecture builder by Moral (2002). This script provides the best training algorithm, the optimum number of hidden layers and neurons in hidden layers among several alternatives. Besides the training algorithms, log-sigmoid transfer function was used. In order to improve the learning rate of ANN, normalization was applied for input and output layers. The script of Moral (2002) was modified to scan more alternatives at a single run of the script. Two different "for" loops are present in the modified script. The first loop changes the number of hidden neurons present in one layer, which starts with 1 and ends in 9. The second loop changes the number of trials (epochs) for a specified number of hidden layers and hidden neurons.

Datasets used in model runs are divided in to three subsets: training, testing and validation. First, the data set was ranked according to the sampling point numbers. Within that dataset, all of the odd-numbered data were set for training purposes. For the testing subset, data with the rank that corresponds to 2+4n are

chosen (n=0,1,2…). Finally, the validation set which is used for the validation of the ANN model is generated using the data with the rank 4+4n (n=0, 1, 2…). By this ranking, spatially homogeneous distribution was provided for all three datasets.

Performances of different ANN models are measured by RMSE and R^2 between the estimated and true values of chlorophyll-a. If R^2 value for all dataset (training, test, validation) is higher than 0.8, than that ANN model is selected for further consideration. In the next stage, selected ANNs are analyzed for the performances only based on the validation, and validation and test datasets. Pearson's coefficient (p) is also calculated for the given model results in this part. Smaller p means lower probability in obtaining a good model performance by chance. In the final stage, the ANN model with the best R^2 and lowest p value is used for simulation purposes.

CHAPTER 4

RESULTS and DISCUSSION

4.1. Statistical Analysis of Field Data

Table 4-1 summarizes the data obtained for different cases. In the table, measured or obtained parameters are presented by green shade while the ones that are not measured or obtained are in red. After completing field and laboratory studies along with image analysis, all data was collected in a file. Later, in order to understand basic relationships, XLStat software was used for statistical analysis. In statistical analysis, the minimum, maximum, mean, standard deviation and coefficient of variation (CV) values of the parameters were calculated. In addition, correlation coefficient (r) and coefficient of determination (R^2) between the parameters were examined.

In Table 4-2, the minimum, maximum, mean, standard deviation and CV values of the data set belonging to the first field study are given. In the table, the "(0)" and "(0.5)" indicators in the variable column (first column) are used to define surface and depth of 0.5 m, respectively. PAR 0/0.5 defines the ratio of the PAR value at surface to the PAR value at the depth of 0.5 m.

Statistical values given for the first field study in Table 4-2 showed that some of the parameters (surface chlorophyll-a, humidity and PAR ratios) were exhibiting relatively higher spatial variation over the sampling points. For these parameters, CV was greater than 0.4. However, for all parameters, CV values were less than 1.0, stating that the variations in parameter values were not significantly high. The correlation matrix for the first field study revealed that at 95% confidence level, there are high correlations (R^2>0.7) between all reflectance values of four bands, as expected. However, chlorophyll-a does not have significant correlations with any of the parameters. Chlorophyll-a exhibited the highest correlations with surface temperature, DO, pH, and depth (R^2=0.42, 0.41, 0.45 and 0.56, respectively).

Table 4-1: Parameter vs. Field Study

Parameter	Field 1 18/9/9	Field 2 2/5/10	Field 3 2/8/10	Parameter	Field 1 18/9/9	Field 2 2/5/10	Field 3 2/8/10
NIR – 2 Reflectance	S	S	S	Average Water Temperature	MS	MS	MS
NIR – 1 Reflectance	S	S	S	Surface Water Temperature	MS	MS	MS
NIR Reflectance	S	S	S	Average Conductivity	MS	MS	MS
Red Edge Reflectance	S	S	S	Surface Conductivity	MS	MS	MS
Red Reflectance	S	S	S	Average TDS	MS	MS	MS
Yellow Reflectance	S	S	S	Surface TDS	MS	MS	MS
Green Reflectance	S	S	S	Average DO	MS	MS	MS
Blue Reflectance	S	S	S	Surface DO	MS	MS	MS
Coastal Blue Reflectance	S	S	S	Average pH	MS	MS	MS
Chlorophyll-a Concentration	L	L	L	Surface pH	MS	MS	MS
TSS Concentration	L	L	L	Average Turbidity	MS	MS	MS
Color	L	L	L	Surface Turbidity	MS	MS	MS
Wind Speed	A	A	A	Secchi Disc Depth	F	F	F
Humidity	A	A	A	Water Column Depth	F	F	F
Air Temperature	A	A	A	PAR Values	F	F	F

S	*Parameters obtained from satellite imagery*	MS	*Parameters measured by multiparameter sonde*
L	*Parameters measured in laboratory*	F	*Parameters measured in field study*
A	*Parameters measured by anemometer*		

Table 4-2: Statistical Values for Field Study 1

Variable	Obs.	Minimum	Maximum	Mean	Std. Dev.	CV
NIR (-)	25	0.010	0.022	0.017	0.004	0.24
R (-)	25	0.030	0.052	0.042	0.006	0.14
G (-)	25	0.070	0.099	0.083	0.010	0.12
B (-)	25	0.025	0.046	0.037	0.006	0.16
Avg. Temp (^{o}C)	25	17.773	20.383	19.161	0.844	0.00
Sur Temp (^{o}C)	25	17.138	21.353	19.348	0.965	0.00
Sur Cond (ms/cm)	25	1.733	3.336	3.226	0.347	0.00
Avg. DO (mg/L)	25	5.439	9.638	7.220	1.106	0.15
Sur DO (mg/L)	25	2.868	12.073	8.362	2.347	0.28
Avg. pH (-)	25	8.928	9.039	8.983	0.031	0.00
Sur pH (-)	25	8.960	9.250	9.062	0.063	0.00
Avg. Turb. (NTU)	25	4.046	17.757	9.958	3.077	0.31
Sur Turb. (NTU)	25	3.775	11.000	7.337	1.910	0.26
Chl-a (0) (µg/L)	25	9.477	189.547	78.946	49.928	0.63
TSS (0) (mg/L)	25	7.333	22.667	13.560	3.752	0.28
TSS (0.5) (mg/L)	25	7.333	24.667	13.093	4.074	0.31
Secchi D. (m)	25	0.470	0.650	0.562	0.055	0.10
Depth (m)	25	1.780	4.150	3.360	0.570	0.00
Wind (m/s)	25	0.040	3.140	1.357	0.651	0.00
Humidity (%)	25	21.600	79.200	43.610	18.807	0.43
PAR 0/0.5 (-)	25	5.716	64.699	15.749	13.549	0.86

Table 4-3 summarizes the statistical analysis of the data for Field Study 2. In this field study, the parameters were impacted by the sudden rise in water depth in April. As can be seen while the average depth recorded in Field 1 was 3.4 m, this value jumped to 4.7 m. For this dataset, chlorophyll-a and turbidity parameters indicated relatively higher variations as well as the band reflectance values. Again, CV values were less than 1.0, stating that the variations in parameter values were not significantly high. At 95% confidence level, correlations among different bands were high and in positive direction. Chlorophyll-a did not have significant correlations with any of the parameters. The highest correlations were with average DO, TSS at 0.5 m, and humidity (R^2=0.25, 0.23, and 0.41, respectively).

In Table 4-4, results of the statistical analysis for Field 3 are provided. This dataset represents the summer conditions. Among the three field studies, Field 3 has the lowest average DO and surface chlorophyll-a concentrations. Moreover, water temperature and SD are the highest compared to other field studies. For Field 3, the average turbidity, surface chlorophyll-a, humidity and PAR ratios were the parameters exhibiting higher spatial variations in addition to the band reflectance values. Although all CV values were less than 1.0, chlorophyll-a concentrations, PAR ratios and band reflectance values had the highest variation in comparison to other field studies. At 95% confidence level, correlations between chlorophyll-a and

other parameters improved. Surface chlorophyll-a had the highest correlation with humidity, surface temperature, surface turbidity, SD, and PAR ratio (R^2=0.61, -0.69, 0.56, -0.59 and 0.51, respectively).

Table 4-3: Statistical Values for Field Study # 2

Variable	Obs.	Minimum	Maximum	Mean	Std. Dev.	CV
NIRr (-)	30	0.001	0.036	0.017	0.007	0.41
Rr (-)	30	0.007	0.033	0.018	0.006	0.33
Gr (-)	30	0.004	0.032	0.015	0.006	0.40
Br (-)	30	0.002	0.023	0.013	0.006	0.46
Avg Temp (^oC)	30	14.161	16.650	15.121	0.665	0.00
Sur Temp (^oC)	30	16.050	19.523	18.117	0.827	0.00
Sur Cond (ms/cm)	30	2.872	3.140	3.026	0.048	0.00
Avg DO (mg/L)	30	10.702	14.011	12.503	0.904	0.00
Sur DO (mg/L)	30	12.148	16.298	14.172	1.195	0.08
Avg pH (-)	30	8.225	8.611	8.532	0.074	0.00
Sur pH (-)	30	8.298	8.735	8.645	0.077	0.00
Avg Tur (NTU)	30	0.894	5.694	2.015	1.090	0.54
Sur Tur (NTU)	30	0.650	5.675	1.363	0.852	0.00
Chl-a (0) (µg/L)	30	16.585	92.404	43.438	16.998	0.39
TSS (0) (mg/L)	30	6.000	14.000	11.317	1.812	0.16
TSS (0.5) (mg/L)	30	5.000	15.000	11.467	2.474	0.22
Secchi D (m)	30	0.500	0.700	0.602	0.053	0.09
Depth (m)	30	2.750	5.700	4.717	0.689	0.00
Wind (m/s)	30	0.000	4.200	1.620	0.965	0.00
Humidity (%)	30	11.300	25.400	18.387	3.344	0.18
PAR 0/0.5 (-)	30	3.449	9.713	5.049	1.407	0.28

When the results of the statistical analyses are compared it is seen that different parameters were identified that have relatively higher correlations with surface chlorophyll-a concentrations. In fact, these correlation were not significantly high at 95% confidence level. However, it must be emphasized that the lake exhibited drastic changes during the course of this study. In Field 1, the impact of drought conditions was prevailing. Then, in Field 2, the impact of sudden increase in water depth by about 40% was in action. In Field 3, again the water input from Lake Mogan was stopped and summer conditions were prevailing. Nutrient availability for algal growth is very important in growth of algae. If nutrient concentrations were included in the statistical analysis, high correlations would probably have been obtained for chlorophyll-a. However, in this study, nutrients are not considered since their effect on absorbing or reflecting the light reaching to the water surface would not be significant. Rather, nutrients have indirect effect on remotely sensed data by impacting the amount of algae (or chlorophyll-a). Humidity exhibited higher spatial variation compared to other parameters. As field studies

start in the morning and takes about 5 hours to complete, humidity changes in air over the lake surface.

Table 4-4: Statistical Values for Field Study # 3

Variable	Obs.	Minimum	Maximum	Mean	Std. Dev.	CV
NIR-1 (-)	30	0.011	0.133	0.028	0.027	0.96
R (-)	30	0.023	0.236	0.050	0.048	0.96
G (-)	30	0.010	0.143	0.031	0.030	0.97
B (-)	30	0.025	0.145	0.044	0.026	0.59
Avg Temp (^{o}C)	30	20.892	25.345	22.762	1.256	0.06
Sur Temp (^{o}C)	30	26.660	28.843	27.753	0.684	0.00
Sur Cond (ms/cm)	30	2.219	3.477	3.377	0.220	0.00
Avg DO (mg/L)	30	2.606	8.537	5.231	1.527	0.29
Sur DO (mg/L)	30	7.570	14.218	10.842	1.897	0.17
Avg pH (-)	30	8.145	10.011	8.363	0.320	0.00
Sur pH (-)	30	8.440	8.758	8.559	0.088	0.01
Avg Tur (NTU)	30	6.649	33.842	12.083	5.256	0.43
Sur Tur (NTU)	30	0.675	4.600	1.493	0.771	0.00
Chl-a (0) (µg/L)	30	0.000	11.847	3.949	3.069	0.78
TSS (0) (mg/L)	30	1.000	3.000	1.920	0.560	0.00
TSS (0.5) (mg/L)	30	1.200	3.000	1.880	0.381	0.00
Secchi D (m)	30	1.200	2.500	1.825	0.356	0.00
Depth (m)	30	3.000	5.750	4.748	0.758	0.00
Wind (m/s)	30	0.000	3.800	1.903	0.940	0.00
Humidity (%)	30	11.900	57.400	25.147	12.939	0.51
PAR 0/0.5 (-)	30	3.113	27.139	5.819	4.874	0.84

In the statistical analysis of the complete dataset (Table 4-5), it was seen that band reflectance values, average DO, average turbidity, surface turbidity, chlorophyll-a, TSS, humidity, and PAR ratio had relatively higher CV values (>0.4) compared to other parameters. Among those, chlorophyll-a and PAR ratio had CV values slightly greater than 1.0. Therefore, changing conditions in different field studies resulted in deviations water clarity related parameters as well as chlorophyll-a and DO.

In the statistical analysis of complete dataset four different parameters shows a variation over the field. One is turbidity value. This situation could be explained by variation in the suspended particles in the mid and bottom layers of the lake. This variation is also supported by present variation in total suspended solid concentration and PAR ratio in water surface. Variation in the humidity did considered as this dataset is a combination of all other three datasets. Prepared correlation matrix (Appendix C) shows that blue band reflectance has a significant correlation with other bands (NIR, red and green). Beside spectral reflectance

values, chlorophyll-a has a significant correlation (R^2>0.6) with SDD. However, they are inversely related to each other. As expected increasing chlorophyll-a concentration results in decreasing SDD value with increasing TSS concentration with a significant negative correlation. Detailed correlation matrices are given in Table C-1, Table C-4 and Table C-7.

Table 4-5: Statistical Values for All Data

Variable	Obs.	Minimum	Maximum	Mean	Std. Dev.	CV
NIR (-)	85	0.001	0.133	0.021	0.018	0.86
R (-)	85	0.007	0.236	0.036	0.032	0.89
G (-)	85	0.004	0.143	0.041	0.034	0.83
B (-)	85	0.002	0.145	0.031	0.021	0.68
Avg Temp (^oC)	85	14.161	25.345	19.006	3.367	0.18
Sur Temp (^oC)	85	16.050	28.843	21.880	4.466	0.20
Sur Cond (ms/cm)	85	1.733	3.477	3.209	0.272	0.00
Avg DO (mg/L)	85	2.606	14.011	8.383	3.384	0.40
Sur DO (mg/L)	85	2.868	16.298	11.288	2.983	0.26
Avg pH (-)	85	8.145	10.011	8.605	0.321	0.00
Sur pH (-)	85	8.298	9.250	8.739	0.227	0.00
Avg Tur (NTU)	85	0.894	33.842	7.905	5.704	0.72
Sur Tur (NTU)	85	0.650	11.000	3.166	2.973	0.94
Chl-a (0) (µg/L)	85	0.000	189.547	39.944	41.654	1.04
TSS (0) (mg/L)	85	1.000	22.667	8.660	5.582	0.64
TSS (0.5) (mg/L)	85	1.200	24.667	8.562	5.655	0.66
Secchi D (m)	85	0.470	2.500	1.022	0.634	0.00
Depth (m)	85	1.780	5.750	4.328	0.923	0.00
Wind (m/s)	85	0.000	4.200	1.643	0.893	0.00
Humidity (%)	85	11.300	79.200	28.191	16.466	0.58
PAR 0/0.5 (-)	85	3.113	64.699	8.468	9.154	1.08

4.2. Analysis of the Relationships Between Chlorophyll-a and Band Reflectance Values

The correlations (given by r values) between band reflectance values and different band ratios are presented in Tables 4-6 and 4-7, respectively. The band ratios in Table 4-7 are selected based on studies in literature. In Table 4-6, correlations with reflectance values are analyzed for different representation forms (exponential, logarithmic, natural logarithmic, etc.) of reflectance values. As can be deducted, the correlations are weak. In addition, correlations for a specific band can be in contradicting directions among different field cases. This may be due to the complexity of absorption and reflectance in Case 2 waters where there a number of factors which would affect these mechanisms as discussed in Chapter 2. Moreover,

the drastic changes in the lake possess additional impacts on the results. The average depths of water were 3.36, 4.72, and 4.75 for Field 1, Field 2, and Field 3, respectively. Average TSS concentrations at the surface were about 13, 11, and 2 mg/L and SD values were about 0.6, 0.6, and 1.8 m for Field 1, Field 2, and Field 3, respectively. Moreover, respective average chlorophyll-a concentrations were about 79, 43, and 4 µg/L.

Table 4-6: Band Correlations (R values) with Chlorophyll-a Concentration

		NIR Band	**Red Band**	**Green Band**	**Blue Band**
Bands	*Field 1*	0.2150	0.0035	0.1476	-0.0718
	Field 2	-0.1169	-0.0635	0.1166	0.0004
	Field 3	0.1780	0.1632	0.1737	0.1971
1/Bands	*Field 1*	-0.2129	-0.0260	-0.1782	0.0749
	Field 2	0.0472	0.0224	-0.1976	0.0244
	Field 3	-0.0420	-0.1235	-0.1402	-0.1794
Exp(Bands)	*Field 1*	0.2150	0.0029	0.1463	-0.0718
	Field 2	-0.1166	-0.0637	0.1159	0.0005
	Field 3	0.1810	0.1621	0.1763	0.1985
1/exp(Bands)	*Field 1*	-0.2150	-0.0040	-0.1488	0.0718
	Field 2	0.1171	0.0633	-0.1173	-0.0003
	Field 3	-0.1749	-0.1639	-0.1710	-0.1957
ln(Bands)	*Field 1*	0.2136	0.0159	0.1630	-0.0730
	Field 2	-0.0978	-0.0460	0.1627	-0.0169
	Field 3	0.1041	0.1474	0.1418	0.1832
1/ln(Bands)	*Field 1*	-0.2144	-0.0081	-0.1507	0.0723
	Field 2	0.1118	0.0553	-0.1417	0.0091
	Field 3	-0.1533	-0.1593	-0.1643	-0.1948

Table 4-7: Band Ratio Correlations with Chlorophyll-a Concentration

	B1/B2	**B1/B3**	**B1/B4**	**B2/B1**	**B2/B3**	**B2/B4**
Field 1	-0.3632	-0.1322	-0.3288	0.3545	0.3427	-0.1899
Field 2	-0.2222	0.0681	0.0441	0.1430	0.3549	0.1321
Field 3	-0.0727	-0.0088	0.0781	0.0695	0.0459	0.1408
	B3/B1	**B3/B2**	**B3/B4**	**B4/B1**	**B4/B2**	**B4/B3**
Field 1	0.1385	-0.3542	-0.3040	0.3369	0.1833	0.3179
Field 2	0.0368	-0.3114	0.0481	-0.1324	-0.3381	-0.1442
Field 3	0.0495	0.0325	0.0505	0.0334	-0.1578	-0.0274

In Table 4-7, spectral bands are given in B1, B2, B3 and B4 are blue, green, red and NIR bands respectively.

In literature, chlorophyll-a concentration was analyzed with respect to spectral reflectance by Mittenzwey et al. (1992) and Han et al. (1997). They stated that the reflectance values in wavelengths smaller than 500 nm decreases with increasing chlorophyll-a concentrations. This range is covered by the blue band in

both Quickbird and Worldview 2 satellites. In that study, it was seen that blue band had a negative correlation with increasing chlorophyll-a concentration. However in Lake Eymir, there was no unique situation. As given in Table 4-6, chlorophyll-a concentrations in Lake Eymir had a weak positive correlation with the blue band reflectance in Field 3, which had the lowest average chlorophyll-a concentration. In Field 2, correlation was almost inexistent, while in Field 1, which had the highest chlorophyll-a concentration, exhibited very weak negative correlation. The differences in these situations can be caused by presence of particles like dissolved organic matters or suspended solids like yellow substances in the lake that have high spectral absorption in blue bands (Han et al., 1997) as well as the level of chlorophyll-a concentrations. Besides suspended particles, blue light tends to reach in deeper parts of water column with its lo0077er wavelength, which may also result in lower reflectance in blue band for Field 3. Mittenzwey et al. (1992) also stated that blue band may not be suitable in determination of chlorophyll-a concentrations in lakes. This statement may also support the low correlations between chlorophyll-a and blue band reflectance values. When correlations between green band reflectance values and chlorophyll-a concentrations are examined, it was observed that for all fields there was a positive relationship as expected. However, the correlation levels are weak ($r<0.2$). Yet, reflectance quantities are higher compared to red and blue bands.

Han et al. (1997) states that, wavelength around 670 nm, chlorophyll-a has its minimum absorbance. Moreover, based on the study conducted by Mittenzwey et al. (1992) and Han et al (1997) it was proved that the best wavelength range for determination of chlorophyll-a concentrations was between 690 nm and 720 nm. However, both Quickbird and Worldview 2 satellites scan the NIR band in longer wavelengths greater than 720 nm. So, the highest correlation can be expected between chlorophyll-a and NIR band. Based on results in Table 4-6, reflection in NIR and also green band had higher correlations with chlorophyll-a compared to other two bands. This condition was supported by Dall'olmo (2005) as well. Dall'olmo stated that chlorophyll-a had higher correlation with NIR and green bands. For Field 1, which had the highest average chlorophyll-a concentration, the reflectance value in NIR was the highest. Han et al. (1997) stated that due to presence of interferences like suspended particles, yellow substances and dissolved organic matters in water bodies, determination of chlorophyll-a is not likely to be successful when only reflectance values are used. Interferences should be eliminated by using hybrid input models.

As given in Table 4-7, the correlations were not improved significantly when the ratios of different band reflectance values were considered. Therefore, the correlation analyses suggest that for Lake Eymir use of just remotely sensed data for determination of chlorophyll-a concentrations may not be sufficient, and other parameters measured in-situ may be needed to improve the prediction capability of regular models that base on remotely sensed reflectance values only.

Besides analyzing the correlations between different band reflectance values and chlorophyll-a concentrations, the performances of the models given in literature are tested for prediction of chlorophyll-a concentrations in Lake Eymir. Results are presented in Table 4-8. Formulations in the table are compiled from the studies that used similar sensors and band ranges in prediction of chlorophyll-a concentrations. It must be emphasized that the coefficients in the formulations provided in Table 4-8 are determined through optimization by Microsoft EXCEL Solver. Therefore, although the band ratios are kept as similar to the ones in the original equations, the coefficients are determined for Lake Eymir through optimization. As a result, existing models are modified to fit the chlorophyll-a concentrations measured in this study. As can be seen in Table 4-8, for all models, Field 1 had the best result. This was followed by Field 3 and then Field 2. In overall, the results showed that none of the models provided in literature was able to define the chlorophyll-a concentrations in Lake Eymir. The highest R^2 value was obtained for the model provided by Zhang (2005). For Field 1, r was 0.55 at 95% confidence level with an R^2 value of 0.30. The study of Zhang (2005) was conducted in a tropical eutrophic shallow lake which is similar to Lake Eymir. Results obtained with the existing regression models indicate the need for improvement in increasing the prediction capability in modeling the chlorophyll-a concentrations in Lake Eymir using remotely sensed data.

Table 4-8: Results of Literature Equations

Equation	Field 1		Field 2		Field 3		Reference
	R^2	RMSE	R^2	RMSE	R^2	RMSE	
$Chl-a = a - b*(Band\ 2) - c*[Band\ 3/Band1]^d$	0.065	47.31	0.032	16.44	0.101	2.86	Baruah et al., 2001
$Chl-a = a^{C_0+C_1*log\frac{R(448)}{R(551)}+C_2\left[log\frac{R(448)}{R(551)}\right]^2+C_3*\left[\log\frac{R(448)}{R(551)}\right]^3}$	0.317	40.43	0.051	16.28	0.112	2.84	Cauwer et al., 2004
$Chl-a = a*\left(\frac{R(520)+R(565)}{R(490)}\right)^b$	0.213	43.42	0.028	16.47	0.103	2.86	Tanaka et al., 2000
$Chl-a = a - b*R(488) - c*R(748) - d\left(\frac{R(667)}{R(551)}\right) + e\left(\frac{R(678)}{R(531)}\right)$	0.296	41.04	0.150	15.41	0.196	2.71	Zhang, 2005
$\ln(Chl-a) = a - b*(Band\ 1/Band\ 3)$	0.042	47.88	0.005	16.67	0.003	3.01	Allan et al., 2007
$Chl-a = k*(Band\ 3*Band\ 4)+b$	0.007	48.76	0.007	16.71	0.000	3.02	Guan et al., 2009
$Chl-a = k*(Band\ 3*Band\ 4/\ln Band\ 1)+b$	0.014	48.59	0.005	16.71	0.003	3.01	
$Chl-a = k*(Band\ 3*Band\ 4/\ln(Band\ 1+Band\ 2))+b$	0.010	48.66	0.004	16.71	0.001	3.02	
$Chl-a = k*(Band\ 3*/Band\ 4)+b$	0.030	48.18	0.002	16.69	0.027	2.98	
$Chl-a = k*((Band\ 4-Band\ 3)/(Band\ 4+Band\ 3))+b$	0.001	48.28	0.013	16.60	0.044	2.95	
$Chl-a = k*(Band\ 4/Band\ 1)+b$	0.078	46.98	0.018	16.57	0.087	2.88	
$Chl-a = k*(Band\ 1/Band\ 2)+b$	0.219	43.23	0.049	16.29	0.078	2.90	
$Chl-a = k*ln(Band\ 1/Band\ 2)+b$	0.221	43.16	0.038	16.39	0.097	2.87	

*In all analysis p values are smaller than 0.05.

4.3. PCA Results

Principle component analysis (PCA) is a mathematical technique in order to reduce dimensionality in a dataset. By using PCA, the number of interrelated parameters in a dataset can be decreased. During this reduction, variation in the dataset is preserved and the most important information is extracted (Abdi and Williams, 2010). In this study, PCA is applied in order to reduce the number of variables that would be used in model development for chlorophyll-a prediction in Lake Eymir using remotely sensed and hybrid data. Therefore, by PCA, variables having less informative characteristics for the defining the changes in the system are eliminated. Parameters are classified in a way that in every class parameters are uncorrelated with each other. Here each set of parameters in classes are called as the principle components. While a number of principle components can be derived, the first principle component is expected to represent the highest variance in the dataset. Before PCA analysis is employed, a screening was applied to remove the independent parameters that exhibited multicollinearity. This application aimed at considering only independent variables in regression model development.

For the dataset of Field 1, PCA was applied on 25 observation points. 5 points were removed due to their inconsistencies (unrealistic chlorophyll-a concentrations). Then PCA was applied on the dataset comprising of near infrared band reflectance, red band reflectance, blue band reflectance, green band reflectance; average and surface water temperature, conductivity, DO concentration, pH, turbidity (at 0 m and 0.5 m depth), chlorophyll-a concentration, total suspended solid concentration, color and PAR ratio, secchi disc depth, water column depth, wind speed, humidity and air temperature. It was seen that the first 6 principle components were able to define about 82% of the system. However, the factor loadings belonging to the parameters in principle components 3 to 6 were low. Based on the factor loadings that are higher than 0.6 (Kline, 1994), parameters stated in Table 4-9 were chosen for regression and modeling purpose.

In Field 2, 30 sampling points were used in the PCA. The first 7 factors were able to define about 80% of the system. Parameters that had a factor loading higher than 0.6 were used in the upcoming modeling analysis. The following table (Table 4-10) represents the chosen factors and the parameters. Even if there is no parameter present that a factor loading greater than 0.6 in the 6^{th} and the 7^{th} principle

components, they still have intermediate effect over the system. However, only the parameters listed in 4-10 were used in modeling.

Table 4-9: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 1)

Factor #	1	2	3	4	5	6
Eigenv.	*6.904*	*4.038*	*2.125*	*1.538*	*1.401*	*1.137*
Var. %	*32.877*	*19.230*	*10.117*	*7.325*	*6.673*	*5.413*
Cum. %	*32.877*	*52.107*	*62.224*	*69.549*	*76.221*	*81.635*
Parameters	Green Ref.	Depth				
	Avg. Temp.	Sur. Chl-a				
	Red Ref.	Wind Speed				
	NIR Ref.	Avg. Turbidity				
	Blue Ref.	Sur. Conductivity				
	Humidity					
	Sur. Temp					
	TSS (0.5 m)					
	Avg. pH					

Table 4-10: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 2)

Factor #	1	2	3	4	5	6	7
Eigenv.	*4.756*	*3.031*	*2.480*	*2.199*	*1.709*	*1.304*	*1.292*
Var. %	*22.650*	*14.435*	*11.810*	*10.473*	*8.139*	*6.209*	*6.151*
Cum. %	*22.650*	*37.085*	*48.895*	*59.368*	*67.506*	*73.715*	*79.866*
Parameters	NIR Ref.	Sur. DO	Avg. pH	Avg. Temp.	Sur. Chl-a		
	Green Ref.	Depth	Sur. pH		Humidity		
	Red Ref.	Avg. Turbidity					
	Blue Ref.						

In the third and final field study, data for 30 points were used in the PCA. The first 6 factors were able to define about 82% of the system. Parameters that had a factor loading higher than 0.6 were used in the upcoming modeling studies. The

following table (Table 4-11) represents the chosen principle factors and parameters chosen for modeling for Field 3.

Table 4-11: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (Field 3)

Factor #	1	2	3	4	5	6
Eigenv.	*6.397*	*3.728*	*2.666*	*1.625*	*1.572*	*1.230*
Var. %	*30.461*	*17.755*	*12.695*	*7.736*	*7.487*	*5.859*
Cum. %	*30.461*	*48.216*	*60.911*	*68.647*	*76.133*	*81.922*
Parameters	Humidity	PAR 0m/0.5m	Sur. DO		Avg. pH	Wind
	Sur. Turbidity	Sur. Cond.	Sur. pH			
	Avg. DO					
	Red Ref.					
	Sur. Chl-a					
	Blue Ref.					
	Green Ref.					
	NIR Ref.					
	Secchi Depth					
	Sur Temp.					

After completing the PCA analysis for different field cases, PCA was applied to a dataset that was composed of the data of all three field cases. In this PCA, data of a total of 85 points were used. This time it was observed that the first 5 principle components were able to define 80% of the system. Parameters that had a factor loading higher than 0.6 in each principle factor are listed in Table 4-12. Table 4-13 summarizes the frequency of parameters in terms of appearing in a principle component with significant factor loadings. In this table, the parameters that had important contribution in defining the variations in the system can be seen.

Corresponding matrices of eigenvalues an eigenvectors are given in Appendix C for Field 1, Field 2 and Field 3 seperately.

Table 4-12: Cumulative Contribution of the Principle Components in Describing the Variations in the Dataset and Parameters that Have High Factor Loadings (All Fields)

Factor #	1	2	3	4	5
Eigenv.	*7.067*	*5.533*	*1.888*	*1.336*	*1.001*
Var. %	*33.650*	*26.348*	*8.988*	*6.363*	*4.769*
Cum. %	*33.650*	*59.999*	*68.987*	*75.350*	*80.118*
Parameters	Sur. Temperature	Green Ref.	NIR Ref.	Sur. pH	Wind Speed
	Secchi Depth	Sur. Turbidity		Sur. Conductivity	
	Avg. Temperature	Humidity			
	Sur. Chl-a	Blue Ref.			
	Avg. DO	Sur. DO			
	0.5 m TSS	Depth			
	Sur. TSS				

Table 4-13: Parameter Comparison Table

Parameters	**Field 1**	**Field 2**	**Field 3**	**All Data**	**# of Occurrence**
NIR Ref.	+	+	+	+	4
Red Ref.	+	+	+		3
Blue Ref.	+	+	+	+	4
Green Ref.	+	+	+	+	4
Avg. Temp.	+	+	+	+	4
Sur. Temp.	+		+	+	3
Avg. Cond.					0
Sur. Cond.	+		+	+	3
Avg. DO			+	+	2
Sur. DO		+	+	+	3
Avg. pH	+	+	+		3
Sur. pH		+	+	+	3
Avg. Turbidity	+	+	+		3
Sur. Turbidity			+	+	2
Surface Chl-a	+	+	+	+	4
0.5 m Chl-a					0
Sur. TSS				+	1
0.5 m TSS	+			+	2
Secchi Depth			+	+	2
Depth	+	+		+	3
Wind	+		+	+	3
Air Temp.					0
Humidity	+	+	+	+	4
PAR 0m / 0.5m			+		1

Based on the PCA analysis, it is seen that NIR reflectance, green reflectance, blue reflectance, water column average temperature, surface chlorophyll-a, humidity are the parameters that define the system variability most. Other important but partially case specific parameters are red band reflectance, average and surface temperature, surface conductivity, surface DO concentration, average and surface pH, average turbidity, water column depth and wind speed. Some of the parameters that were measured in the analysis as: average conductivity, chlorophyll-a concentration 0.5 beneath the water surface and air temperature, did not take place as important factors in defining the system.

4.4. Modeling of Chlorophyll-a

Modeling of the chlorophyll-a concentrations is applied for three different field cases (Field 1, Field 2, and Field 3). Moreover, the data for all fields are combined and a 4th set of inputs is established. Modeling is employed using linear and non-linear regression and ANNs. For each case, models are developed using hybrid inputs (both field parameters and remotely sensed data) and remotely sensed data only. Due to software limitations in the XLStat, multivariate linear regression and non-linear regression models could only be applied to using a maximum of 11 independent variables. As mentioned before, independent parameter elimination was achieved through PCA as given in Table 4-13. Further elimination was realized taking the below factors into consideration:

- Wind speed was not taken into consideration in the modeling stage due to its heterogeneous nature with respect to time. However, if sampling could have been achieved in shorter periods of time and at the image acquisition time, it could provide better modeling results since waves can impact reflectance from water surface.
- Air temperature was another heterogeneously distributed parameter like wind speed. In field, there was an average of 5-6 hour difference between the sampling of first point and the last point. In this time period, significant changes in air temperature were measured. So, air temperature values were not considered as an independent parameter similar to humidity.

4.4.1. Modeling of Chlorophyll-a for Field 1

Results for the multivariate linear and non-linear regression models and the ANN model are provided below. Measured values and predicted chlorophyll-a distributions obtained by the given models as well as error distributions are provided in Appendix-A (Figures A-1 to A-7).

4.4.1.1. Linear Regression Model

Equation 5 and Equation 6 are obtained as the linear regression models when hybrid and remotely sensed data only are used, respectively, as the independent variables.

$$Chlorophyll - a\ (\mu g/L) = 669.1558 + 3918.4151 * NIR\ Reflectance - 14692.3873 * Red\ Reflectance + 11052.6384 * Green\ Reflectance - 11290.8232 * Blue\ Reflectance - 30.9592 * Average\ Temperature + 45.0155 * Surface\ Water\ Temperature - 109.9762 * Average\ pH + 3.2581 * Average\ Turbidity + 3.5664 * TSS\ in\ 0.5m\ Depth + 15.2970 * Water\ Column\ Depth + 0.9482 * Humidity$$

(Eq-5)

$$Chlorophyll - a\ (\mu g/L) = 10.4385 + 205.0708 * NIR\ Reflectance - 11774.7265 * Red\ Reflectance + 11028.3898 * Green\ Reflectance - 9617.8901 * Blue\ Reflectance$$

(Eq-6)

The statistical analysis of the results provided by Equations 5 and 6 are summarized in Table 4-14. As can be seen, the model with the hybrid input was more successful in predicting the chlorophyll-a distribution compared to the input of band reflectance values only, based on R^2 and RMSE. However, an R^2 value of 0.731 stated that there was a room for improvement. The t-values obtained for the coefficients of the parameters in above equations indicate that not all are smaller than 0.05 (at 95% confidence level). This may be due to the small number of observations compared to the number of parameters since the parameters that exhibit multicollinearity have been removed. Therefore, not all of the coefficients are significant. This situation have been observed in other field studies as well. Yet, in modeling of chlorophyll-a, all parameters selected through PCA analysis have been included in the model, although coefficient belonging to some of them were not significant. In Figure 4-1 and Figure 4-2, measured versus predicted chlorophyll-a concentrations are

depicted. In the figures, best line is represented by a solid line, while theoretical best line is represented by a dashed line. Spatial distributions of predicted chlorophyll-a concentrations and error distributions are depicted in Appendix-A.

Table 4-14: Field 1: Comparison of the Linear Regression Models with Different Inputs

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.731	0.443
	Correlation Coefficient	0.885	0.666
	p-Value	1.595 E-30	2.705 E-82
	Root Mean Square Error	25.4	36.5

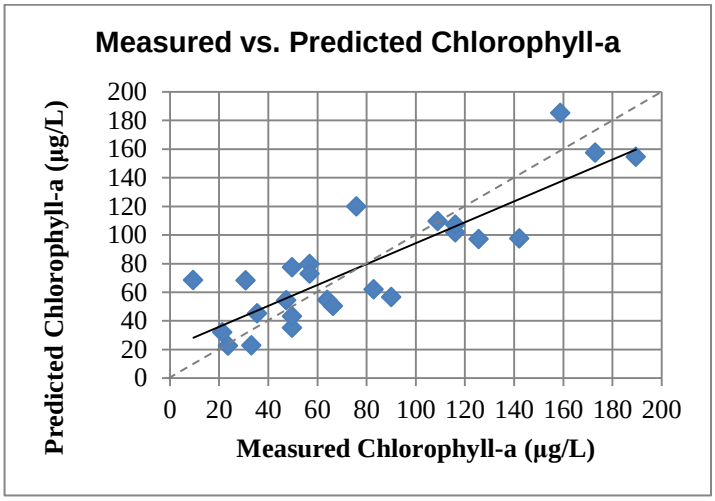

Figure 4-1: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 1)

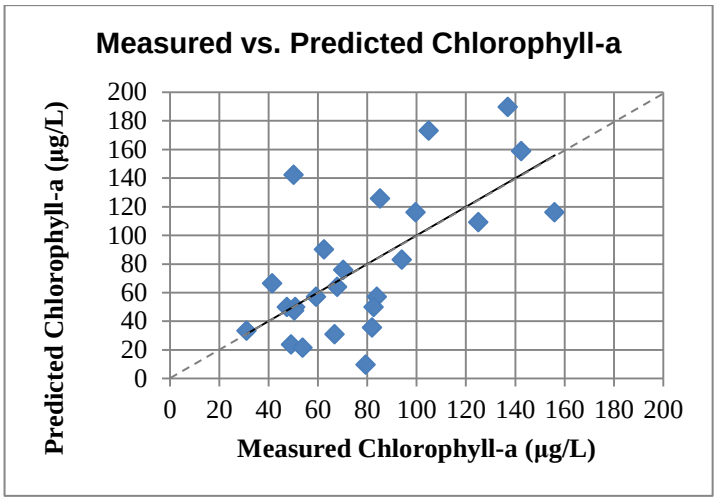

Figure 4-2: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Field 1)

The generated models (Equations 5 and 6) are applied to Fields 2 and 3 as well, in order to observe their applicability over different conditions in the same area in different times. In the following table (Table 4-15), calculated results can be seen. The N/A values in the table below represents that those values cannot be calculated, which is the case for p-value. This is mainly resulting from the presence of high residual between measured and predicted data which may result in negative chlorophyll-a concentrations in the predicted values. This shows that multivariate linear regression models derived for Field 1 are unable to model the conditions set by Field 2 and Field 3. As mentioned before, while the average chlorophyll-a concentration in Field 1 was about 79 µg/L, it was 43 µg/L and 4 µg/L in Field 2, and Field 3, respectively. Moreover, there had been drastic changes in the conditions in the lake before the study period for Field 2. Therefore, conditions were completely different and in return very poor results were obtained when the linear regression models were applied to Field 2 and Field 3 even when hybrid inputs are used.

Table 4-15: Applicability of Linear Regression Models of Field 1 to Field 2 and 3

		Hybrid Model	Band Only Model
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.001	0.037
	Correlation Coefficient	0.038	0.193
	p-Value	N/A	N/A
	Root Mean Square Error	687	213
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.036	0.025
	Correlation Coefficient	-0.189	-0.159
	p-Value	N/A	N/A
	Root Mean Square Error	1230	839

4.4.1.2. Non-Linear Regression Model

The non-Linear regression models obtained using XLStat are provided below. The model with the hybrid inputs is given by Equation 7. The model that uses only the reflectance values is stated by Equation 8.

$$Chlorophyll - a\ (\mu g/L) = -1.2008 * 10^{-10} * NIR\ Reflectance^{-5.7897} - 51545.9661 * Red\ Reflectance^{1.6402} + 10918.8476 * Green\ Reflectance^{1.3385} - 7535.1559 * Blue\ Reflectance^{0.9534} + 1.3250 * Average\ Temperature^{0.3435} + 0.7251 * Surface\ Temperature^{1.7019} + 6.9362 * Average\ pH^{0.9655} + 0.4988 * Average\ Turbidity^{1.5884} + 0.7942 * TSS\ in\ 0.5m\ Depth^{1.3992} + 7.1484 * 10^{-5} * Water\ Column\ Depth^{\ 9.9903} + 0.1168 * Humidity^{1.6060}$$

(Eq-7)

$$Chlorophyll - a\ (\mu g/L) = 137.1664 * NIR\ Reflectance^{0.9883} - 231958.5266 * Red\ Reflectance^{2.1928} + 19511.7835 * Green\ Reflectance^{1.3610} - 9315.3034 * Blue\ Reflectance^{0.9886}$$

(Eq-8)

The statistical analysis of the non-linear regression models is summarized in Table 4-16. The distributions obtained using the non-linear regression models and error distribution are provided in Appendix-A in Figures A-4 and A-5, respectively.

Application of non-linear regression models provided better results compared to linear regression models. The coefficient of determination value was calculated to be 0.806 for the model with hybrid inputs with a very small p-value. This value is also significantly higher compared to the model with the input of band data only. This proves that hybrid model has higher predictability over chlorophyll-a distribution in the lake surface. Comparisons of predicted versus measured chlorophyll-a concentrations are given in Figures 4-3 and 4-4 for different model inputs

Table 4-16: Field 1: Comparison of the Non-Linear Regression Models with Different Inputs

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.806	0.465
	Correlation Coefficient	0.898	0.682
	p-Value	3.710 E-27	4.864 E-77
	Root Mean Square Error	21.6	35.8

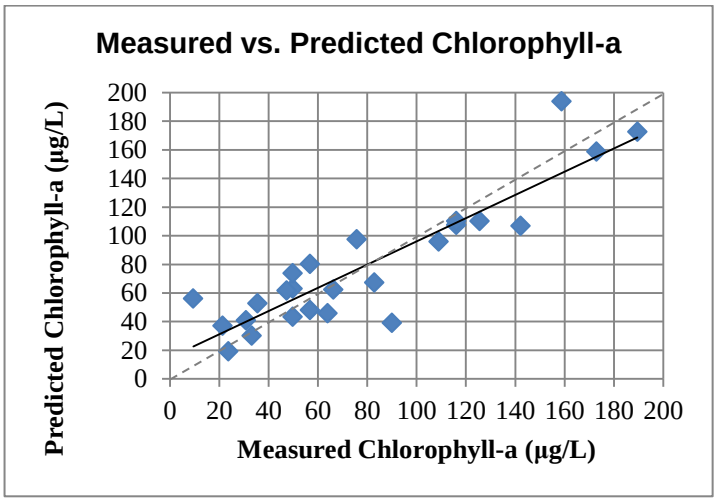

Figure 4-3: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 1)

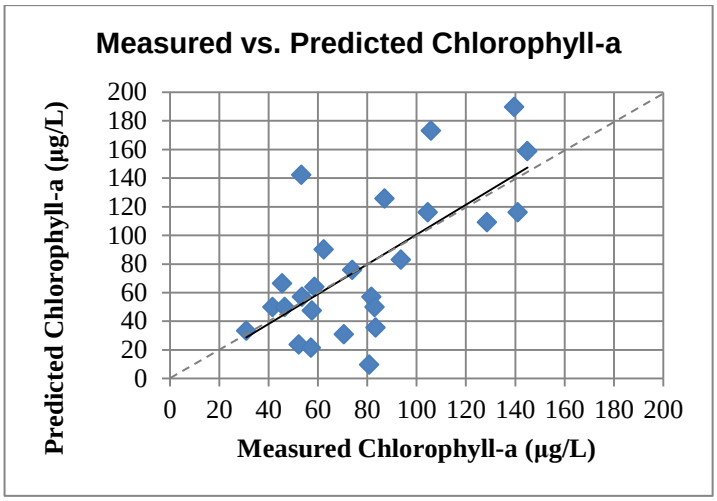

Figure 4-4: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Field 1)

Application of the non-linear regression models to Fields 2 and 3 resulted in unsatisfactory predictions as was the case for the linear regression models. The performances are summarized in Table 4-17. This was due to the conditions mentioned in the previous section. Here again the results are showing that non-linear regression models derived are incapable of modeling different conditions, same as the linear regression model.

Table 4-17: Applicability of the Non-Linear Regression Models of Field 1 to Field 2 and Field 3

		Hybrid Model	Band Only Model
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.000	0.013
	Correlation Coefficient	-0.018	0.115
	p-Value	N/A	N/A
	Root Mean Square Error	2.96 E+06	144
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.027	0.020
	Correlation Coefficient	-0.163	-0.142
	p-Value	N/A	N/A
	Root Mean Square Error	1505	2099

4.4.1.3. ANN Model

For the dataset obtained from the first field study, the best ANN model with the hybrid input was obtained with a 2-hidden layered architecture which had 7 hidden neurons each. In the best ANN model, resilient back-propagation algorithm had given the best fit by minimizing the errors between the predicted and measured chlorophyll-a concentrations. It was not possible to come up with an ANN model that exhibited an R^2 value higher than 0.6 without memorizing the training dataset. The statistical information about the best ANN model with the hybrid inputs are given in Table 4-18. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-5, 4-6, 4-7, and 4-8, respectively. Spatial distributions for the ANN predictions and errors are depicted Figure A-6 and Figure A-7, respectively.

Table 4-18: Field 1: Statistical Information for ANN Model with Hybrid Inputs

		Hybrid Model			
		All Data	**Training**	**Testing**	**Validation**
Field # 1	**# of Independent Parameters**	11			
	# of Samples	25	13	6	6
	Coefficient of Determination	0.857	0.990	0.731	0.729
	Correlation Coefficient	0.926	0.995	0.855	0.854
	p-Value	2.750 E-16	6.710 E-01	1.660 E-08	1.160 E-14
	Root Mean Square Error	18.711	6.246	20.575	31.973

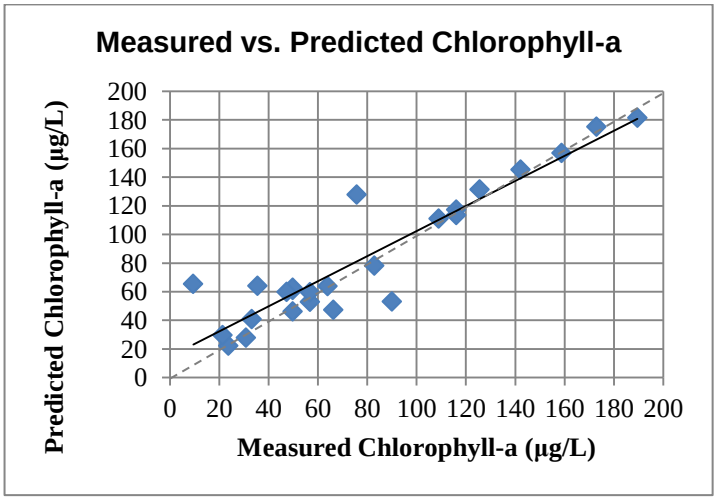

Figure 4-5: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 1)

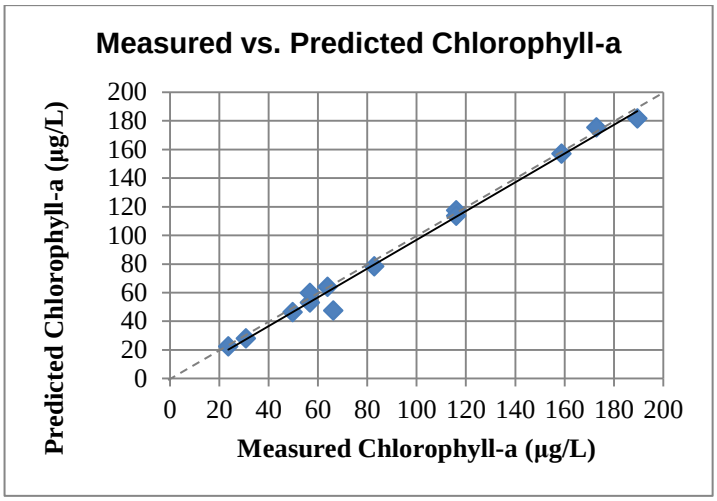

Figure 4-6: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 1)

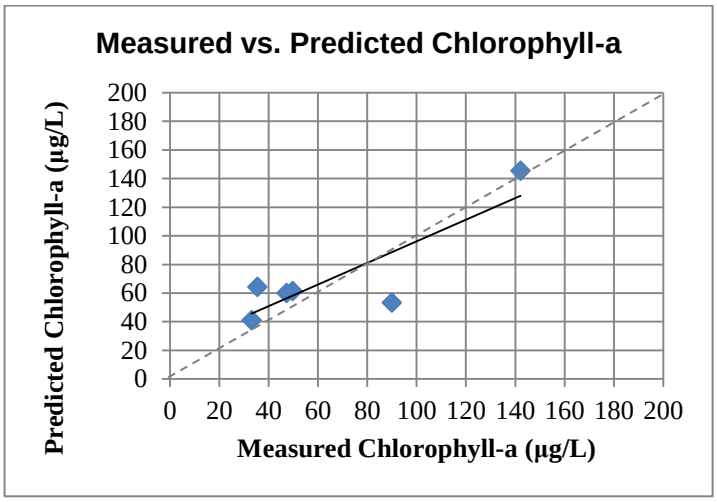

Figure 4-7: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 1)

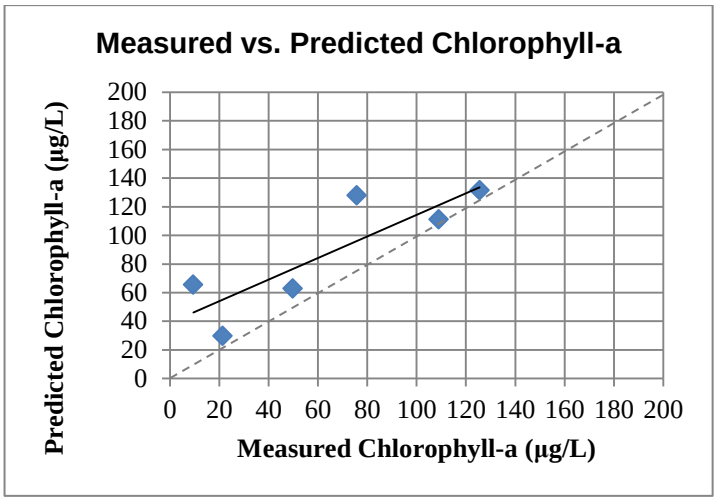

Figure 4-8: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 1)

The p-value for the validation dataset shows that the risk of obtaining an over-trained ANN model is insignificant. In ANN models, high R^2 values should be obtained for the test and validation datasets in order to prove that ANN is successful in predictions. Here, the coefficient of determination values for the test and validation datasets were reported as 0.818 and 0.839 respectively with p-values both smaller than 0.05. Consequently, ANN can model the chlorophyll-a concentrations better than other proposed algorithms.

The applicability of ANN model to Field 2 and Field 3 could not be tested due to the limitations in the MATLAB neural network software. It was not possible to extract the algorithm used in prediction of the chlorophyll-a concentrations and apply it to another dataset. Therefore, applicability of the ANN model to other conditions will be tested using the combined dataset in section 4.4.4.3.

In Table 4-19, results for multivariate linear regression, non-linear regression and ANN models are summarized. As can be seen from the table below, ANN model had the best result for Field 1. When p-values were analyzed, it was seen than all of them were close to 0, showing that predicted values were statistically significant and matched the measured values well. When RMSE values were calculated, the minimum error was achieved by the ANN model. There is a 15% improvement in the RMSE value obtained for the non-linear regression model compared to the linear regression model. Improvement in the RMSE obtained for the ANN model is 25% compared to the linear regression model. Also by

considering the significance f-values with a confidence level of 95%, the relationship between the predicted and measured chlorophyll-a concentrations is statistically significant with a f value smaller than 0.05 in all three models. In addition to this, the lower f value in ANN shows a better prediction capability. For all models, the hybrid inputs resulted in better predictability compared to using only the reflectance values for chlorophyll-a determination. This is due to the presence of other factors other than chlorophyll-a which would impact the reflectance values in Lake Eymir.

Table 4-19: First Field Study Result Summary Table for Hybrid Models

		Multivariate Linear Regression	Non-Linear Regression	Artificial Neural Network
Field # 1	**# of Independent Parameters**	11		
	# of Samples	25		
	Coefficient of Determination	0.731	0.806	0.857
	Correlation Coefficient	0.855	0.898	0.926
	p-Value	1.595 E-30	3.710 E-27	2.753 E-16
	Significance f-Value	5.316 E-08	1.189 E-09	3.402 E-11
	Root Mean Square Error	25.4	21.6	19.2

4.4.2. Modeling of Chlorophyll-a for Field 2

Results for the multivariate linear and non-linear regression models and the ANN model are provided below. Measured values and predicted chlorophyll-a distributions obtained by the given models as well as error distributions are provided in Appendix-A (Figures A-8 to A-14).

4.4.2.1. Linear Regression Model

Equation 9 and Equation 10 are obtained as the linear regression models when hybrid and only the reflectance values are used, respectively, as the independent variables.

$$Chlorophyll - a\ (\mu g/L) = 234.6159 - 2171.0666 * NIR\ Reflectance - 1528.1382 * Red\ Reflectance + 3311.7563 * Green\ Reflectance + 436.4498 * Blue\ Reflectance - 7.1966 * Average\ Temperature - 6.9197 * Surface\ DO\ Concentration - 81.7467 * Average\ pH + 95.3195 * Surface\ pH - 3.7155 * Average\ Turbidity - 6.5683 * Water\ Column\ Depth - 3.4352 * Humidity$$

(Eq-9)

$$Chlorophyll - a\ (\mu g/L) = 51.6594 - 1235.8235 * NIR\ Reflectance - 1680.2041 * Red\ Reflectance + 3002.0316 * Green\ Reflectance - 184.9398 * Blue\ Reflectance$$

(Eq-10)

The statistical analysis of the results obtained by the applications of Equations 9 and 10 for prediction of chlorophyll-a concentrations are summarized in Table 4-20. As can be seen, the model with the hybrid input was more successful in predicting the chlorophyll-a distribution compared to the input of band reflectance values only, based on R^2 and RMSE. However, the R^2 value for the model with the hybrid input was less than 0.6. This value is even less when only the reflectance values are used as the inputs. It must be noted that the average chlorophyll-a concentration in the lake was less for Field 2 compared to Field 1. Moreover, the steady state in the lake was abolished due to huge amounts of water entrance into the lake which might have changed the status and levels of other parameters in the lake that may impact the reflectance values other than chlorophyll-a. In Figure 4-9 and Figure 4-10, measured versus predicted chlorophyll-a concentrations are depicted. Spatial distributions of predicted chlorophyll-a concentrations and error distributions are given in Appendix-A.

Table 4-20: Second Field Study Linear Regression Results Comparison Table

		Hybrid Model	**Band Only Model**
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.577	0.227
	Correlation Coefficient	0.760	0.476
	p-Value	1.089 E-05	1.736 E-17
	Root Mean Square Error	10.9	14.7

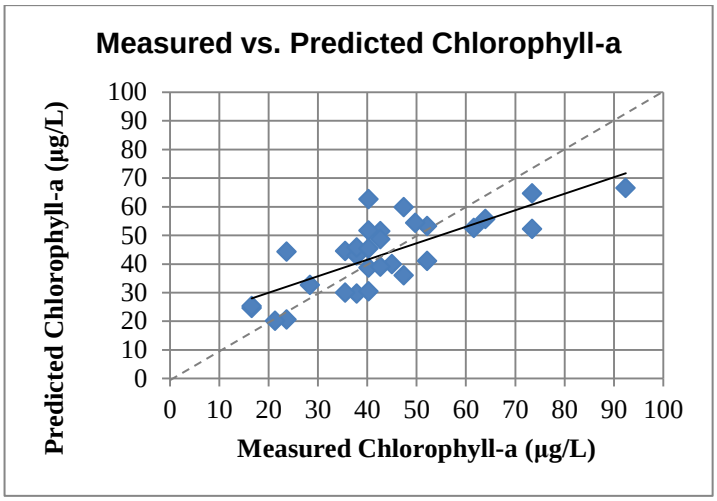

Figure 4-9: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 2)

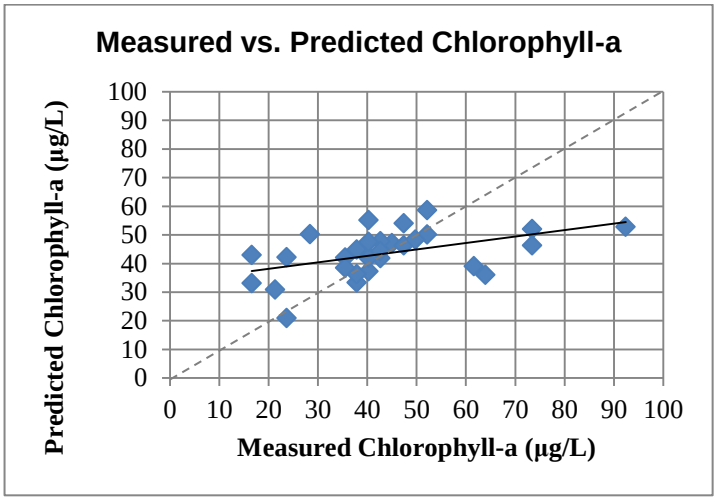

Figure 4-10: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Field 2)

The generated models (Equations 9 and 10) are applied to Fields 1 and 3 as well, in order to observe their applicability over these fields. In the following table (Table 4-21), calculated results can be seen. Results indicate that multivariate linear regression models derived for Field 2 are unable to model the conditions set by Field 1 and Field 3.

Table 4-21: Applicability of Linear Regression Model of Field 2 to Field 1 and 3

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.014	0.003
	Correlation Coefficient	0.119	0.057
	p-Value	1.171 E-271	0
	Root Mean Square Error	100	134
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.226	0.005
	Correlation Coefficient	-0.475	-0.070
	p-Value	N/A	8.819 E-16
	Root Mean Square Error	165	67.4

4.4.2.2. Non-Linear Regression Model

The non-Linear regression models obtained using XLStat are provided below. The model with the hybrid inputs is given by Equation 11. The model that uses only the reflectance values is stated by Equation 12.

$$Chlorophyll - a\ (\mu g/L) = 18.5886 * NIR\ Reflectance^{-0.2315} - 808.2183 * Red\ Reflectance^{0.6687} - 48.1262 * Green\ Reflectance^{-0.2380} + 139.8590 * Blue\ Reflectance^{0.1721} - 0.4717 * Average\ Temperature^{1.8863} + 633.3315 * Surface\ DO\ Concenrtation^{-0.8501} + 3.9685 * 10^{-5} * Average\ pH^{2.6459} + 4.4511 * Surface\ pH^{1.3957} - 1.3557 * 10^{-5} * Average\ Turbidity^{14.9053} - 1.0814 * 10^{-6} * Water\ Column\ Depth^{9.7683} + 1442.0279 * Humidity^{-1.1859}$$

(Eq-11)

$$Chlorophyll - a\ (\mu g/L) = 86.1190 * NIR\ Reflectance^{-5.7267*10^{-2}} - 1260.8070 * Red\ Reflectance^{1.0109} - 19.4418 * Green\ Reflectance^{-0.3148} + 32.2573 * Blue\ Reflectance^{0.0105}$$

(Eq-12)

The statistical analysis of the non-linear regression models is summarized in Table 4-22. The distributions obtained using the non-linear regression models and error distribution are provided in Appendix-A in Figures A-11 and A-12, respectively.

Application of non-linear regression models provided better results compared to linear regression models. The coefficient of determination value was calculated to be 0.620 for the model with hybrid inputs with p-values close to 0. This R^2 value is also significantly higher compared to the model with the input of band data only. This situation shows that the hybrid model has higher predictability over chlorophyll-a distribution in the lake surface. Comparisons of predicted versus measured chlorophyll-a concentrations are given in Figures 4-11 and 4-12 for different model inputs.

Table 4-22: Field 2: Comparison of the Non-Linear Regression Models with Different Inputs

		Hybrid Model	Band Only Model
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.620	0.171
	Correlation Coefficient	0.787	0.413
	p-Value	2.648 E-05	2.422 E-19
	Root Mean Square Error	10.3	15.2

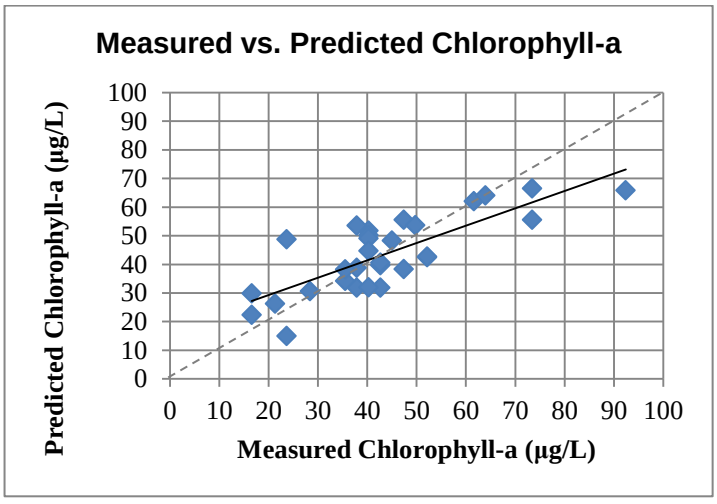

Figure 4-11: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 2)

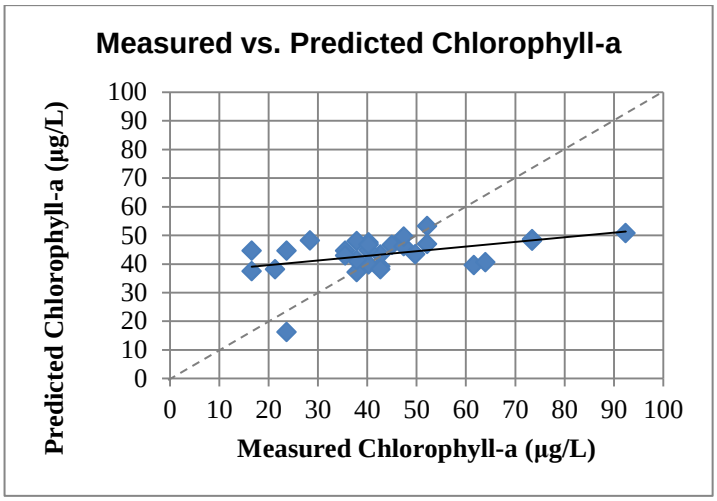

Figure 4-12: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Field 2)

Application of the non-linear regression models to Fields 1 and 3 resulted in unsatisfactory predictions similar to the linear regression models. The performances of the non-linear regression models over Field 1 and Field 3 are summarized in Table 4-23. Poor results for Fields 1 and 3 are probably due to the differences in lake characteristics as mentioned in the previous sections. Since the conditions in Lake Eymir were significantly different, results show that the non-linear model was case specific to Field 2.

Table 4-23: Applicability of the Non-Linear Regression Models of Field 2 to Field 1 and Field 3

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.010	0.027
	Correlation Coefficient	0.098	0.164
	p-Value	N/A	0
	Root Mean Square Error	1.263 E+08	58.6
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.010	0.022
	Correlation Coefficient	0.099	-0.148
	p-Value	N/A	1.152 E-22
	Root Mean Square Error	1.551 E+12	56.9

4.4.2.3. ANN Model

For the dataset obtained from the second field study, the best ANN model with the hybrid input was obtained with a 2-hidden layered architecture which had 7 hidden neurons each. In the best ANN model, Powell-Beale conjugate gradient back-propagation algorithm had given the best fit by minimizing the errors between the predicted and measured chlorophyll-a concentrations. The statistical information about the best ANN model with the hybrid inputs are given in Table 4-24. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-13, 4-14, 4-15, and 4-16, respectively. Spatial distributions for the ANN predictions and errors are depicted Figure A-13 and Figure A-14, respectively.

Table 4-24: Second Field Study ANN Results Comparison Table

		Hybrid Model			
		All Data	**Training**	**Testing**	**Validation**
Field # 2	**# of Independent Parameters**	11			
	# of Samples	30	15	8	7
	Coefficient of Determination	0.884	0.987	0.818	0.839
	Correlation Coefficient	0.919	0.994	0.905	0.916
	p-Value	4.403 E-02	9.987 E-01	7.710 E-04	3.122 E-02
	Root Mean Square Error	6.953	2.875	9.237	9.589

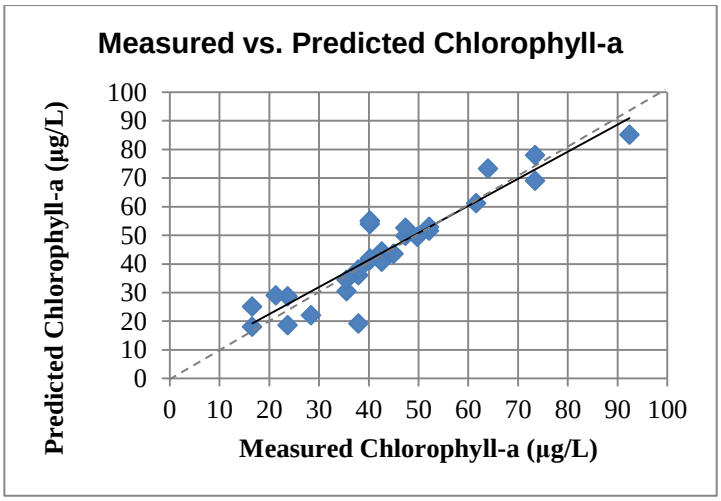

Figure 4-13: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 2)

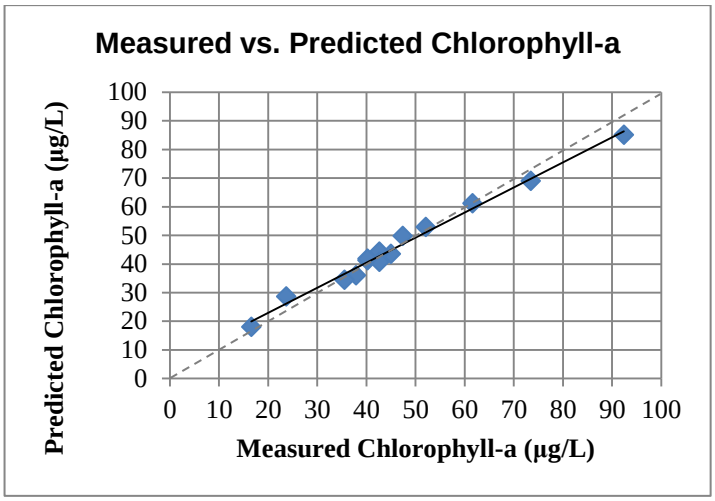

Figure 4-14: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 2)

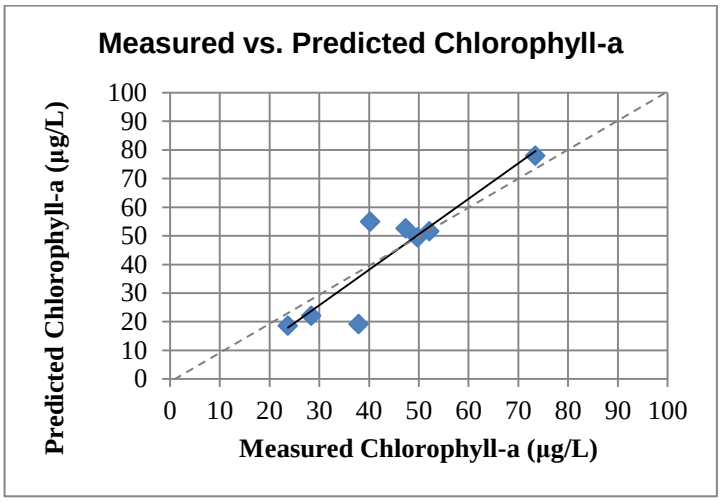

Figure 4-15: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 2)

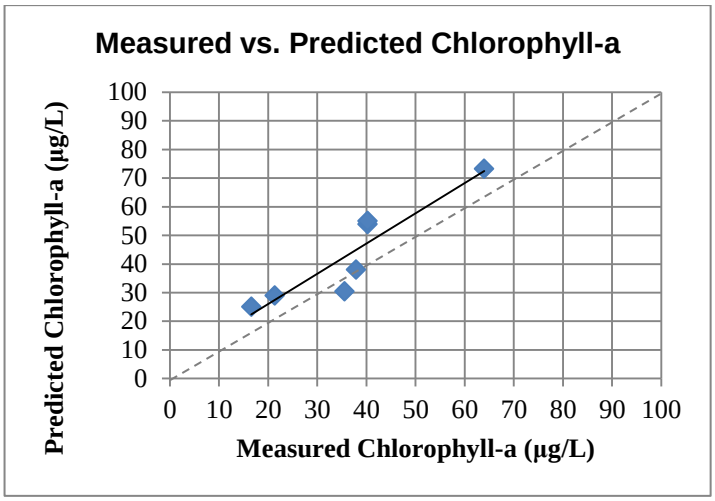

Figure 4-16: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 2)

The coefficient of determination values for the test and validation datasets were reported as 0.818 and 0.839 respectively with very small p-values. Consequently, ANN had superiority over modeling chlorophyll-a concentrations with linear and non-linear regression. As was the case for Field 1, applicability of the ANN model to Field 1 and Field 3 could not be tested due to the limitations in the MATLAB neural network software.

In Table 4-25, results for multivariate linear regression, non-linear regression and ANN models are summarized for hybrid inputs. As can be seen from the table below, ANN model had the best result for Field 2. When p-values were analyzed, it was seen than all of them were smaller than 0.05, showing that predicted values were statistically significant and matched the measured values well. When RMSE values were calculated, the minimum error was achieved by the ANN model. . There is a 6% improvement in the RMSE value obtained for the non-linear regression model compared to the linear regression model. Improvement in the RMSE obtained for the ANN model is 37% compared to the linear regression model. Also by considering the significance f-values with a confidence level of 95%, the relationship between the predicted and measured chlorophyll-a concentrations is statistically significant with a f value smaller than 0.05 in all three models. For all models, the hybrid inputs resulted in better predictability compared to using only the reflectance values for chlorophyll-a determination. This difference is due to the presence of other factors, which was discussed in Chapter 2.

Table 4-25: Second Field Study Result Summary Table for Hybrid Models

		Multivariate Linear Regression	Non-Linear Regression	Artificial Neural Network
Field # 2	**# of Independent Parameters**	11		
	# of Samples	30		
	Coefficient of Determination	0.577	0.620	0.844
	Correlation Coefficient	0.760	0.787	0.919
	p-Value	1.089 E-05	2.648 E-05	4.403 E-02
	Significance f-Value	1.124 E-06	2.476 E-07	7.899 E-13
	Root Mean Square Error	10.9	10.3	6.95

4.4.3. Modeling of Chlorophyll-a for Field 3

Results for the multivariate linear and non-linear regression models and the ANN model are provided below. Measured values and predicted chlorophyll-a distributions obtained by the given models as well as error distributions are provided in Appendix-A (Figures A-15 to A-21).

4.4.3.1. Linear Regression Model

Equation 13, Equation 14 and Equation 15 are obtained as the linear regression models when hybrid and remotely sensed data only (for 4 and 8 bands separately) are used, respectively, as the independent variables.

$$Chlorophyll - a\ (\mu g/L) = 78.2098 + 56.5244 * NIR1\ Reflectance - 8.0619 * Red\ Reflectance - 171.6369 * Green\ Reflectance + 157.3755 * Blue\ Reflectance - 2.6602 * Surface\ Temperature + 0.4944 * Surface\ DO\ Concentration - 0.7716 * Surface\ pH - 0.1713 * Surface\ Turbidity - 0.7028 * Secchi\ Depth - 0.0265 * Humidity + 1.6043 * 10^{-2} * PAR\ 0m/0.5m$$

(Eq-13)

$$Chlorophyll - a\ (\mu g/L) = 0.5904 + 11.3980 * NIR1\ Reflectance + 3.9362 * Red\ Reflectance - 132.4445 * Green\ Reflectance + 158.2747 * Blue\ Reflectance$$

(Eq-14)

$$Chlorophyll - a\ (\mu g/L) = -11.4832 - 262.1031 * (NIR2\ Reflectance) + 239.2866 * (NIR1\ Reflectance) - 228.2794 * Red\ Edge\ Reflectance + 1.4462 * Red\ Reflectance + 204.7833 * Yellow\ Reflectance - 375.1554 * Green\ Reflectance + 333.9439 * Blue\ Reflectance + 229.4724 * Coastal\ Blue\ Reflectance$$

(Eq-15)

The statistical analysis of the results provided by Equations 13, 14 and 15 are summarized in Table 4-26. As can be seen, the model with the hybrid input had higher prediction capability over the chlorophyll-a distribution compared to the input of band reflectance values only, based on R^2 and RMSE. It must be reminded that Field 3 had the lowest average chlorophyll-a concentration (4 µg/L). When only the reflectance values are utilized as the inputs, very low R^2 value (0.068) was obtained. Therefore, parameters other than chlorophyll-a were effective in reflectance values. Yet, when reflectance values on all bands were considered, although still low, the R^2 improved. In Figure 4-17, Figure 4-18 and Figure 4-19, measured versus predicted chlorophyll-a concentrations are given. Spatial distributions of predicted chlorophyll-a concentrations and error distributions are depicted in Appendix-A. In the third field study, all measured chlorophyll-a concentrations were within a range set by 6 different values according to the spectrophotometric measurements in the laboratory. These values varied from 0.000 to 0.005 in absorbance increments of 0.001. This condition resulted in overlapping values for lower chlorophyll-a concentrations.

Table 4-26: Field 3: Comparison of the Linear Regression Models with Different Inputs

		Hybrid Model	**4-Band Only Model**	**8-Band Only Model**
Field # 3	**# of Independent Parameters**	11	4	8
	# of Samples	30	30	30
	Coefficient of Determination	0.664	0.068	0.343
	Correlation Coefficient	0.802	0.261	0.585
	p-Value	5.671 E-01	1.039 E-04	2.116 E-23
	Root Mean Square Error	1.80	2.91	2.45

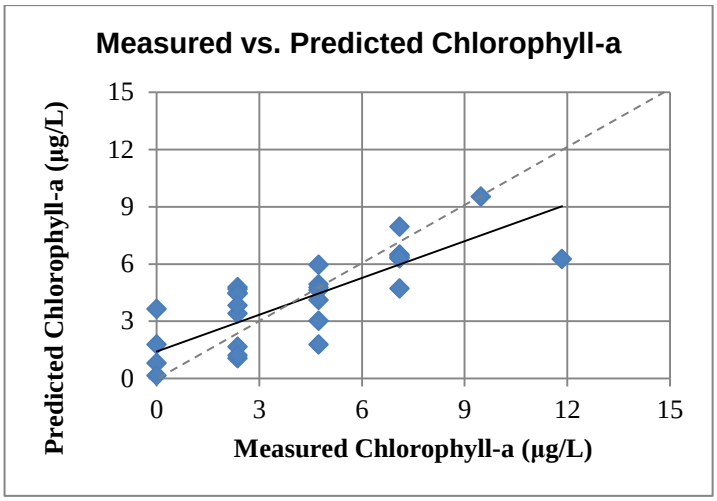

Figure 4-17: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Field 3)

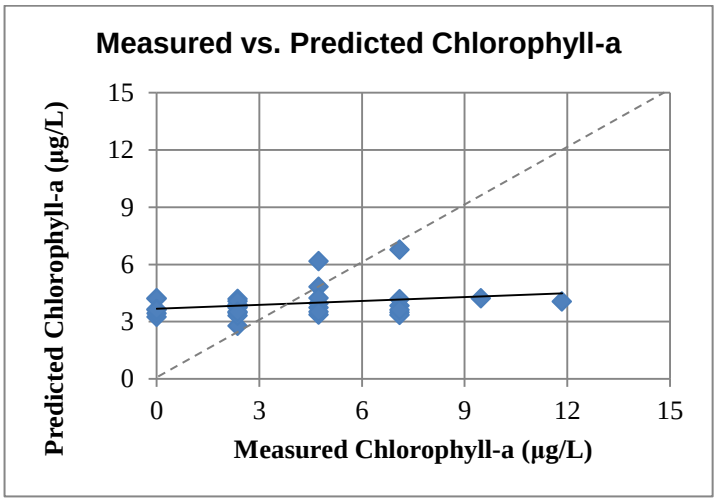

Figure 4-18: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 4 – Band Reflectance Inputs (Field 3)

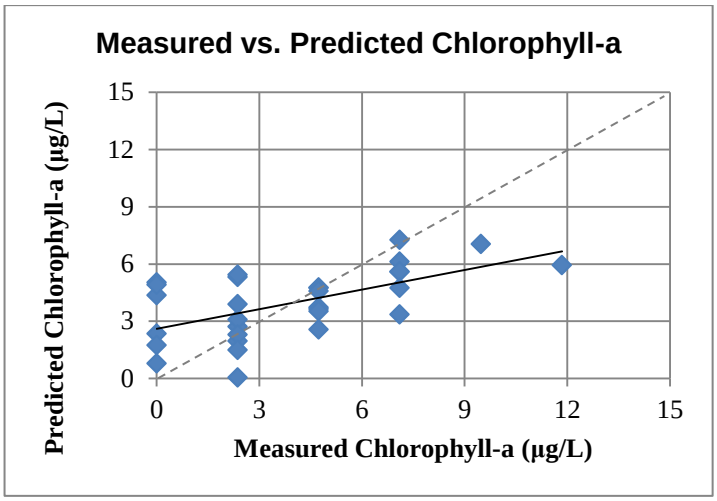

Figure 4-19: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 8 – Band Reflectance Inputs (Field 3)

The generated models (Equations 13 and 14) are also applied to Fields 1 and 2, to observe their applicability over different conditions for same area in different times. However it was not possible to apply Equation 15 to Field 1 and Field 2 due to the lower number of bands in Quickbird Satellite compared to World View 2. In the following table (Table 4-27), calculated results are shown. The N/A value in the table below represents that the p-value could not be calculated. This is mainly resulting from the presence of high residual between measured and predicted data which may result in negative chlorophyll-a concentrations in the predicted values. The multivariate linear regression models derived for Field 3 were unable to model the conditions set by Field 1 and Field 2. As mentioned before, while the average chlorophyll-a concentration in Field 3 was about 4 µg/L, it was 79 µg/L and 43 µg/L in Field 1, and Field 2, respectively. In addition there had been drastic changes in the characteristics of the lake. Therefore, conditions were completely different and in return very poor results were obtained when the linear regression models of Field 3 were applied to Field 1 and Field 2 even when hybrid inputs are used.

Table 4-27: Applicability of Linear Regression Model from Field 3 to Field 1 and 2

		Hybrid Model	4- Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.054	0.105
	Correlation Coefficient	-0.231	-0.324
	p-Value	0	N/A
	Root Mean Square Error	82.3	96.6
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.006	0.040
	Correlation Coefficient	0.078	-0.199
	p-Value	2.002 E-84	0
	Root Mean Square Error	21.7	45.8

4.4.3.2. Non-Linear Regression Model

The non-Linear regression models obtained using XLStat are provided below. The one with the hybrid inputs is given in Equation 16. Models with the reflectance values obtained from 4 bands and 8 bands are stated in Equations 17 and 18, respectively.

$$Chlorophyll - a\ (\mu g/L) = -7.582 * NIR1\ Reflectance^{6.0941*10^{-2}} + 0.6614 * Red\ Reflectance^{-0.3389} - 299130.8762 * Green\ Reflectance^{5.1106} + 1769.8220 * Blue\ Reflectance^{2.4960} - 2.8172 * Surface\ Temperature^{0.2513} + 2.5693 * Surface\ DO\ Concenrtation^{0.3782} + 29.1937 * Surface\ pH^{-3.5182} + 12.3980 * Surface\ Turbidity^{0.2620} - 7.4227 * Secchi\ Disc\ Depth^{-7.9537} - 2.5336 * Humidity^{\ 0.1448} - 204.7361 * PAR\ 0m/0.5m^{\ -3.3020}$$

(Eq-16)

$$Chlorophyll - a\ (\mu g/L) = 2.0525 * NIR1\ Reflectance^{-0.4202} - 53.2719 * Red\ Reflectance^{3.2696} - 23.0271 * Green\ Reflectance^{-5.4548*10^{-3}} 53.2094 * Blue\ Reflectance^{0.3626}$$

(Eq-17)

$$Chlorophyll - a\ (\mu g/L) = -177.2865 * NIR2\ Reflectance^{0.8270} + 1.1521 * 10^{-2} * NIR1\ Reflectance^{-1.0436} - 292.2789 * Red\ Edge\ Reflectance^{1.3695} - 4457.3227 * Red\ Reflectance^{5.7317} + 123.7252 * Yellow\ Reflectance^{0.6637} - 44.5671 * Green\ Reflectance^{0.2768} + 531.1430 * Blue\ Reflectance^{1.4566} + 82.5562 * Coastal\ Blue\ Reflectance^{0.6753}$$

(Eq-18)

The statistical analysis of the non-linear regression models is summarized in Table 4-28. The distributions obtained using the non-linear regression models and error distribution are provided in Appendix-A in Figures A-18 and A-19, respectively. Application of non-linear regression models provided better results compared to linear regression models. The coefficient of determination value was calculated to be 0.675 for the model with hybrid inputs. Similar to the linear regression models, increased bands resulted in better prediction of chlorophyll-a. Comparisons of predicted versus measured chlorophyll-a concentrations are given in Figures 4-20, 4-21 and Figure 4-22 for different model inputs.

Table 4-28: Field 3: Comparison of the Non-Linear Regression Models with Different Inputs

		Hybrid Model	**4-Band Only Model**	**8-Band Only Model**
Field # 3	**# of Independent Parameters**	11	4	8
	# of Samples	30	30	30
	Coefficient of Determination	0.675	0.112	0.330
	Correlation Coefficient	0.822	0.334	0.574
	p-Value	7.868 E-01	9.329 E-05	3.565 E-04
	Root Mean Square Error	1.72	2.84	2.47

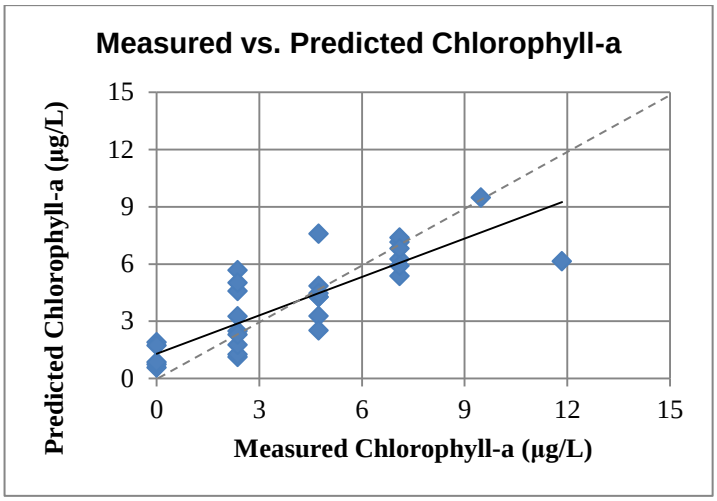

Figure 4-20: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Field 3)

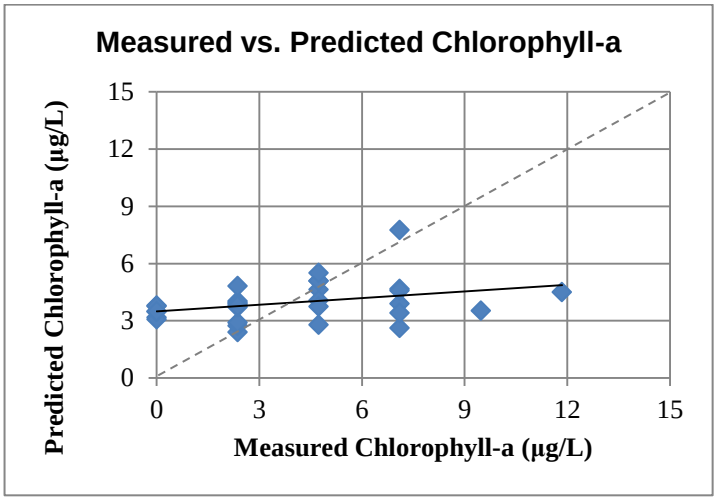

Figure 4-21: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with 4 – Band Only Reflectance Inputs (Field 3)

Application of the non-linear regression models to Fields 1 and 2 resulted in unsatisfactory predictions as was the case for the linear regression models. The performances are summarized in Table 4-29. This was due to the conditions mentioned in the previous section. Here again the results are showing that non-linear regression models derived are incapable of modeling different conditions, same as the linear regression model.

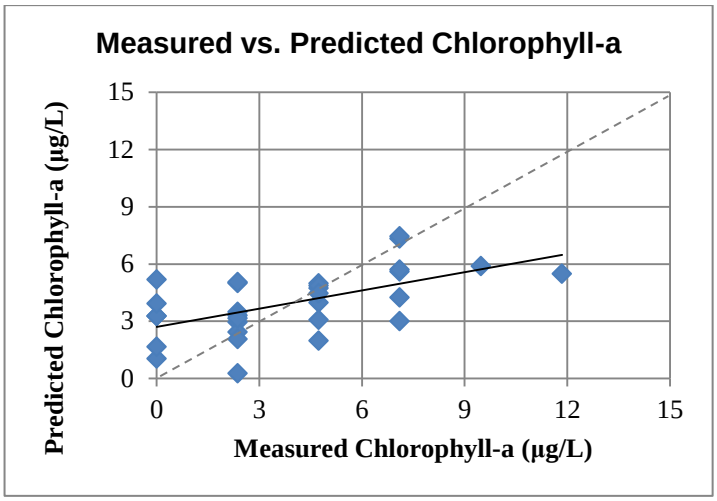

Figure 4-22: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with 8 – Band Only Reflectance Inputs (Field 3)

Table 4-29: Applicability of Non-Linear Regression Model of Field 3 to Field 1 and 2

		Hybrid Model	4- Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.082	0.092
	Correlation Coefficient	0.286	-0.304
	p-Value	N/A	0
	Root Mean Square Error	1271	89.5
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.032	0.007
	Correlation Coefficient	0.180	0.085
	p-Value	N/A	N/A
	Root Mean Square Error	686	47.1

4.4.3.3. ANN Model

For the dataset obtained from the third field study, the best ANN model with the hybrid input was obtained with a 2-hidden layered architecture which had 8 hidden neurons each. In the best ANN model, Fletcher-Powell conjugate gradient back-propagation had given the best fit by minimizing the errors between the

predicted and measured chlorophyll-a concentrations. The statistical information about the best ANN model with the hybrid inputs are given in Table 4-30. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-23, 4-24, 4-25, and 4-26, respectively. Spatial distributions for the ANN predictions and errors are depicted Figure A-20 and Figure A-21, respectively.

Table 4-30: Field 3: Statistical Information for ANN Model with Hybrid Inputs

		Hybrid Model			
		All Data	**Training**	**Testing**	**Validation**
Field # 3	**# of Independent Parameters**	11			
	# of Samples	30	15	7	8
	Coefficient of Determination	0.726	0.922	0.609	0.495
	Correlation Coefficient	0.852	0.960	0.780	0.704
	p-Value	3.221 E-01	9.800 E-01	9.247 E-03	2.082 E-01
	Root Mean Square Error	1.689	1.103	2.519	1.692

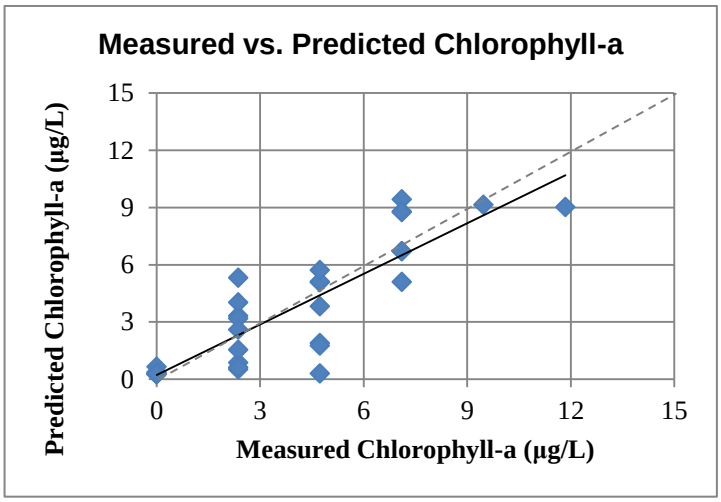

Figure 4-23: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (All Data Set - Field 3)

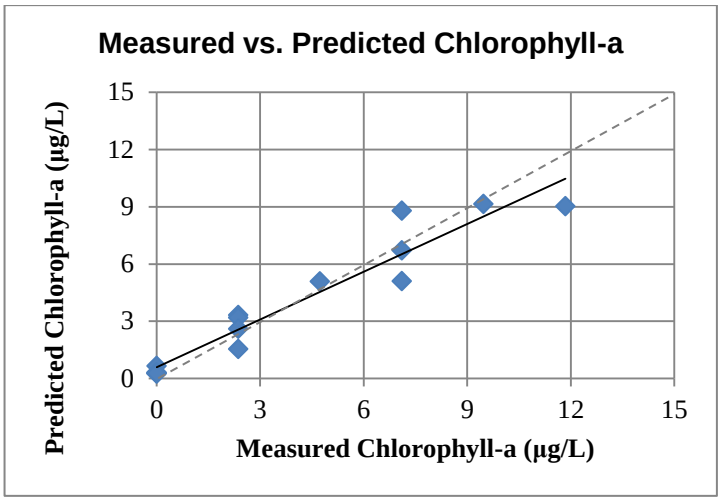

Figure 4-24: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set - Field 3)

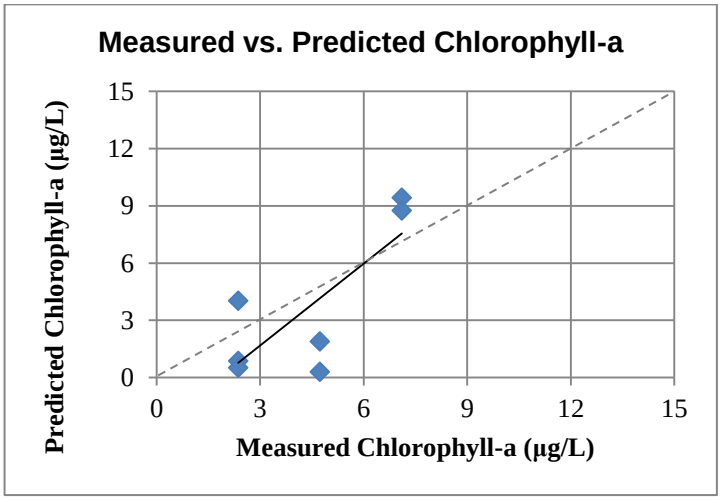

Figure 4-25: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set - Field 3)

In ANN models, the R^2 value obtained for the validation data set was much higher compared to the ones obtained for regression models. However, in overall, prediction performance of ANN was lower for Field 3 compared to Fields 1 and 2. This is probably due to low chlorophyll-a concentrations in Field 3. The coefficient of determination values for the test and validation datasets were reported as 0.609 and 0.495, respectively with relatively high p-values. Yet, still the ANN was more successful compared to other proposed algorithms.

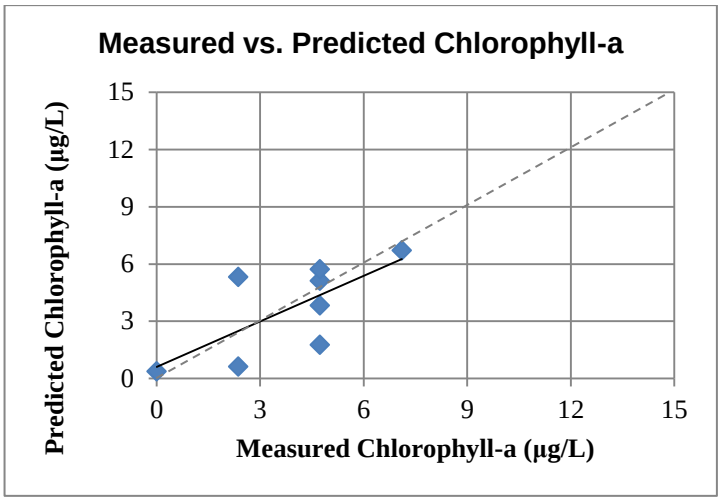

Figure 4-26: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set - Field 3)

In Table 4-31, results for multivariate linear regression, non-linear regression and ANN models are summarized. As can be seen from the table below, ANN model had the best result for Field 3. When the RMSE values for the linear and non-linear regression models were compared, it was observed that the non-linear regression model ended up with a 4% less RMSE. The RMSE for the ANN model was 6% less compared to the one for the linear regression model. When p-values were analyzed, it was seen than all of them were higher than 0.05, showing that predicted values may not be statistically significant. Higher p-values can result due to the repetitive values in the model dataset as a result of low chlorophyll-a concentrations. Significance f-values were smaller than 0.05 in a confidence level of 95%, indicating that the regression parameters are nonzero and that the regression equation have some validity in fitting the data (i.e., the independent variables are not purely random with respect to the dependent variable). With lower f value, ANN showed better prediction capability.

Table 4-31: Third Field Study Result Summary Table for Hybrid Models

		Multivariate Linear Regression	Non-Linear Regression	Artificial Neural Network
Field # 3	**# of Independent Parameters**	11		
	# of Samples	30		
	Coefficient of Determination	0.644	0.675	0.726
	Correlation Coefficient	0.802	0.822	0.852
	p-Value	5.671 E-01	7.868 E-01	3.221 E-01
	Significance f-Value	9.777 E-08	2.581 E-08	2.279 E-09
	Root Mean Square Error	1.80	1.72	1.69

4.4.4. Modeling of Chlorophyll-a for Combined Data

Results for the multivariate linear and non-linear regression models and the ANN model are provided below. Based on the results given in the sections 4.4.1., 4.4.2. and 4.4.3., none of the models were able to model the other cases. Therefore all data representing different field were combined to generate models that would represent the changing conditions in the lake. Results are provided below.

4.4.4.1. Linear Regression Model

Equation 19 and Equation 20 are obtained as the linear regression models when hybrid and remotely sensed data only are used, respectively, as the independent variables.

$$Chlorophyll-a\ (\mu g/L) = 12.1139 + 523.9834 * NIR\ Reflectance + 1758.3758 * Green\ Reflectance - 2575.5769 * Blue\ Reflectance - 0.6193 * Surface\ Temperature - 2.5167 * Average\ DO\ Concentration + 4.7740 * Surface\ DO\ Concentration + 0.0553 * Surface\ Turbidity - 2.9515 * Surface\ TSS + 1.8951 * TSS\ in\ 0.5m\ Depth + 2.2507 * Secchi\ Disc\ Depth + 2.7703 * Water\ Column\ Depth$$

(Eq-19)

$$Chlorophyll-a\ (\mu g/L) = 40.6360 + 503.5351 * NIR\ Reflectance + 3.4571 * Red\ Reflectance + 1252.0397 * Green\ Reflectance - 2028.7702 * Blue\ Reflectance$$

(Eq-20)

The statistical analysis of the results provided by Equations 19 and 20 are summarized in Table 4-32. As can be seen, the model with the hybrid input was more successful in predicting the chlorophyll-a distribution compared to the input of band reflectance values only, based on R^2 and RMSE. However, an R^2 value of 0.630 stated that there was room for improvement. In Figure 4-27 and Figure 4-28, measured versus predicted chlorophyll-a concentrations are depicted.

Table 4-32: Combined Dataset: Comparison of the Linear Regression Models with Different Inputs

		Hybrid Model	Band Only Model
All Data	**# of Independent Parameters**	11	4
	# of Samples	85	85
	Coefficient of Determination	0.630	0.550
	Correlation Coefficient	0.794	0.742
	p-Value	1.789 E-102	5.392 E-94
	Root Mean Square Error	25.2	27.783

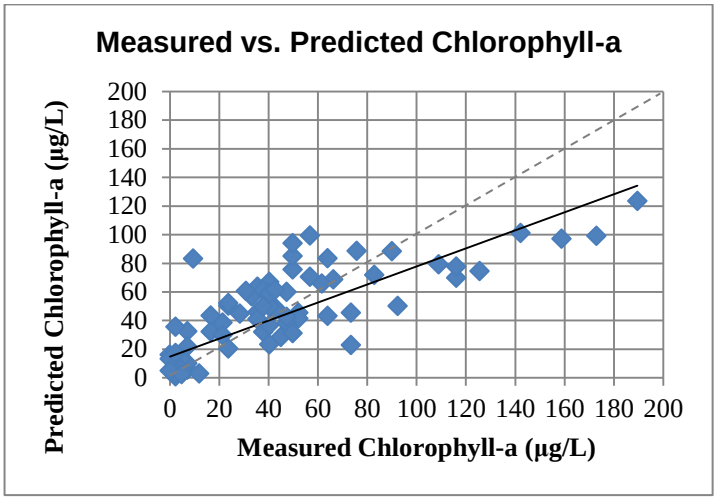

Figure 4-27: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Hybrid Inputs (Combined Dataset)

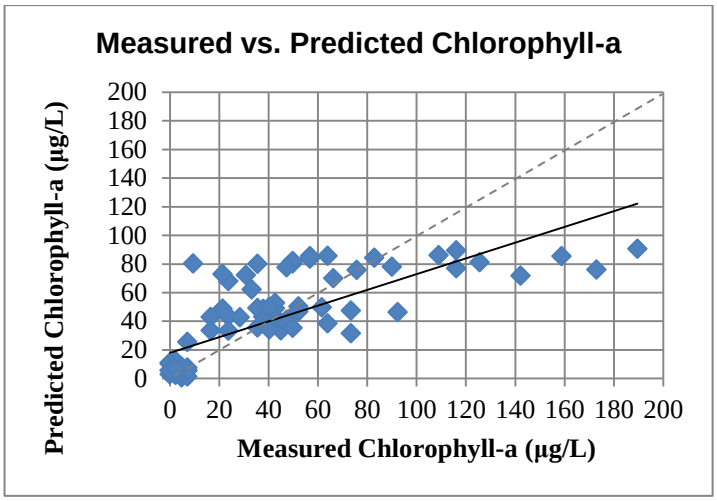

Figure 4-28: Measured versus Predicted Chlorophyll-a for the Linear Regression Model with Only Reflectance Inputs (Combined Data)

The generated models (Equations 19 and 20) are applied to Fields 1, 2 and 3 as well, in order to observe their applicability over different conditions in the same area in different times. In the following table (Table 4-33), calculated results can be seen. Again, the multivariate linear regression models derived from the combined dataset were not successful in modeling the chlorophyll-a concentrations for Field 1, Field 2 and Field 3 separately. The best result was obtained for Field 1, which had the highest chlorophyll-a concentrations.

Table 4-33: Applicability of Linear Regression Model obtained using the combined dataset to Fields 1, 2 and 3

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.464	0.153
	Correlation Coefficient	0.681	0.391
	p-Value	1.915 E-78	1.940 E-129
	Root Mean Square Error	38.9	46.691
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.015	0.006
	Correlation Coefficient	0.124	0.076
	p-Value	2.256 E-48	4.991 E-32
	Root Mean Square Error	19.4	17.418
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.001	0.037
	Correlation Coefficient	-0.030	-0.193
	p-Value	N/A	N/A
	Root Mean Square Error	12.7	8.181

4.4.4.2. Non-Linear Regression Model

The non-Linear regression models obtained using XLStat are provided below. The model with the hybrid inputs is given by Equation 21. The model that uses only the reflectance values is stated by Equation 22.

$$Chlorophyll - a\ (\mu g/L) = 11653.2809 * NIR\ Reflectance^{2.2890} + 1133.0913 * Green\ Reflectance^{0.7525} - 2531.3781 * Blue\ Reflectance^{0.9358} - 1.8932 * Surface\ Temperature^{1.0369} - 7.5677 * 10^{-4} * Average\ DO\ Concentration^{4.1881} + 23.6271 * Surface\ DO\ Concentration^{0.5611} - 3.2580 * 10^{-11} * Surface\ Turbidity^{11.4676} - 3.8393 * Surface\ TSS^{1.0378} + 7.7372 * TSS\ in\ 0.5m\ Depth^{0.5902} + 9.5426 * 10^{-2} * Secchi\ Disc\ Depth^{5.8485} + 4.3950 * Water\ Column\ Depth^{0.8109}$$

(Eq-21)

$$Chlorophyll - a\ (\mu g/L) = -77933445015.5919 * NIR\ Reflectance^{9.7598} + 7318927.4312 * Red\ Reflectance^{9.8598} + 246283747073.34 * Green\ Reflectance^{9.2317} - 866136195747.829 * Blue\ Reflectance^{9.9810}$$

(Eq-22)

The statistical analysis of the non-linear regression models is summarized in Table 4-34. Application of non-linear regression models provided better results compared to linear regression models. The coefficient of determination value was calculated to be 0.674 for the model with hybrid inputs with a very small p-value. This value is also significantly higher compared to the model with the input of band data only. The model with the hybrid inputs had higher predictability over chlorophyll-a distribution in the lake surface. Comparisons of the predicted versus measured chlorophyll-a concentrations are given in Figures 4-29 and 4-30 for different model inputs.

Table 4-34: Non-Linear Regression Results Comparison Table for All Data

		Hybrid Model	Band Only Model
All Data	**# of Independent Parameters**	11	4
	# of Samples	85	85
	Coefficient of Determination	0.674	0.130
	Correlation Coefficient	0.821	0.361
	p-Value	1.051 E-92	0
	Root Mean Square Error	23.6	49.2

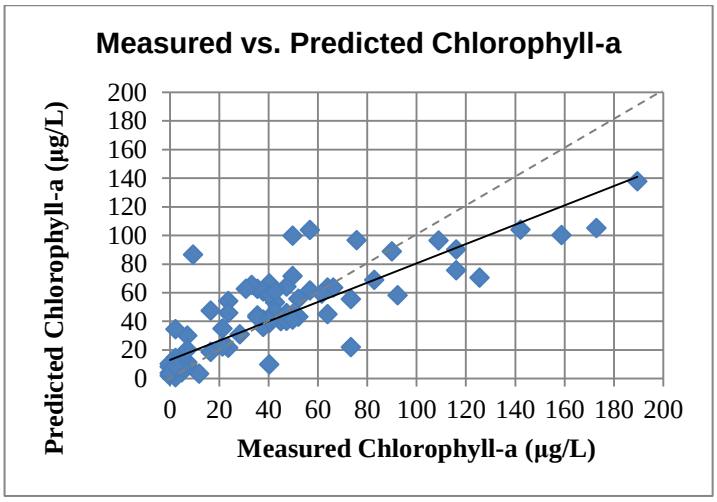

Figure 4-29: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Hybrid Inputs (Combined Data)

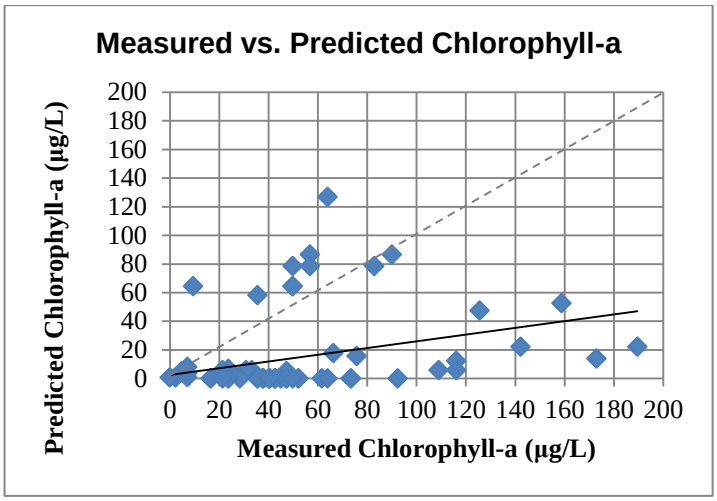

Figure 4-30: Measured versus Predicted Chlorophyll-a for the Non-Linear Regression Model with Only Reflectance Inputs (Combined Data)

Application of the non-linear regression models to Fields 1, 2 and 3 resulted in unsatisfactory predictions as was the case for the linear regression models. However, again, the performance of the hybrid model was better compared to the model that used only the band reflectance values. The performances of the non-linear regression models with different inputs over Field 1, 2 and 3 separately are summarized in Table 4-35. This was due to the conditions mentioned in the previous section. Similar to previous cases, poor results were obtained and the best result was for Field 1.

Table 4-35: Applicability of Linear Regression Model for the Combined Data to Fields 1, 2 and 3

		Hybrid Model	Band Only Model
Field # 1	**# of Independent Parameters**	11	4
	# of Samples	25	25
	Coefficient of Determination	0.495	0.024
	Correlation Coefficient	0.704	-0.156
	p-Value	5.248 E-67	0
	Root Mean Square Error	36.7	74.443
Field # 2	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.119	0.000
	Correlation Coefficient	0.345	-0.005
	p-Value	7.330 E-52	0
	Root Mean Square Error	17.6	46.542
Field # 3	**# of Independent Parameters**	11	4
	# of Samples	30	30
	Coefficient of Determination	0.001	0.037
	Correlation Coefficient	0.026	-0.193
	p-Value	N/A	0
	Root Mean Square Error	12.3	8.181

4.4.4.3. ANN Model

The dataset is generated by combining the data for three different field cases. For training, 12 data were chosen from the dataset of each field. Care was given to cover the whole chlorophyll-a concentration range for a given field. For the test part, 3 data were chosen randomly from each 3 field datasets. Finally, for the validation purposes remaining data were used, 10 from Field 1, 15 each from Field 2 and Field 3. The best ANN model with the hybrid input was obtained with a 2-hidden layered architecture which had 7 hidden neurons each. In the best ANN model, Powell-Beale conjugate gradient back-propagation algorithm had given the best fit by minimizing the errors between the predicted and measured chlorophyll-a concentrations. The statistical information about the best ANN model with the hybrid

inputs are given in Table 4-36. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-31, 4-32, 4-33, and 4-34, respectively.

Table 4-36: ANN Results Comparison Table for All Dataset

		Hybrid Model			
		All Data	**Training**	**Testing**	**Validation**
All Data	**# of Independent Parameters**	11			
	# of Samples	85	36	9	40
	Coefficient of Determination	0.798	0.835	0.651	0.846
	Correlation Coefficient	0.893	0.914	0.807	0.920
	p-Value	2.225 E-168	3.208 E-84	2.494 E-30	1.027 E-60
	Root Mean Square Error	11.472	18.201	36.133	19.832

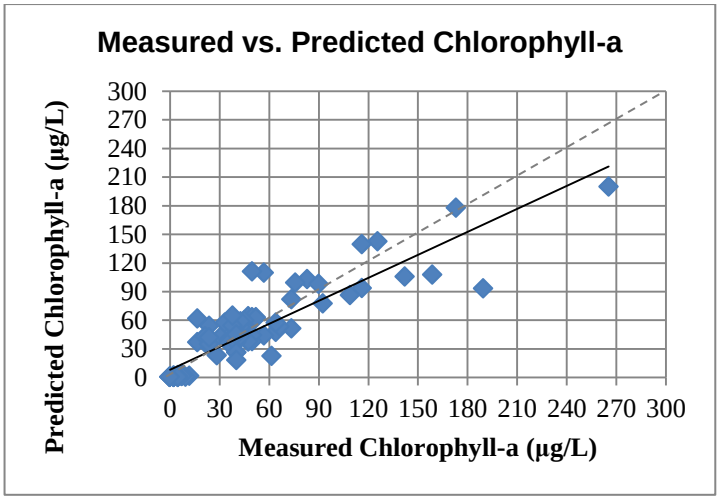

Figure 4-31: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Combined Data)

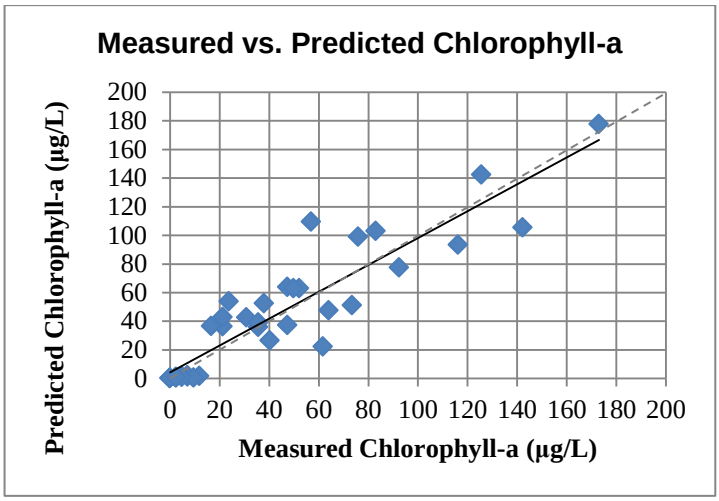

Figure 4-32: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Training Data Set – Combined Data)

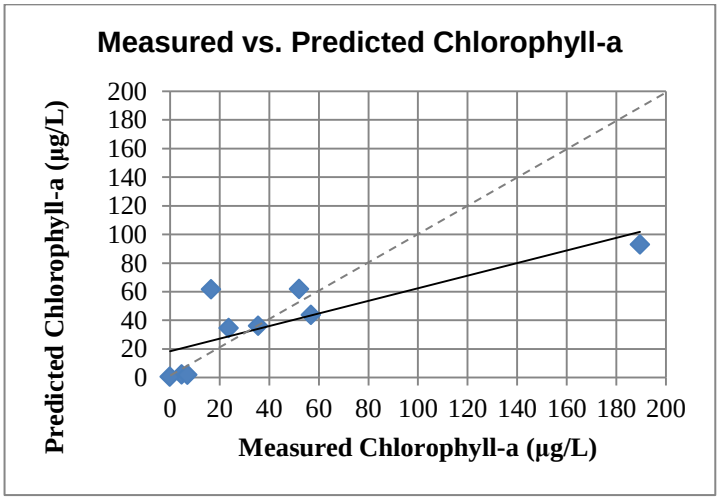

Figure 4-33: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Testing Data Set – Combined Data)

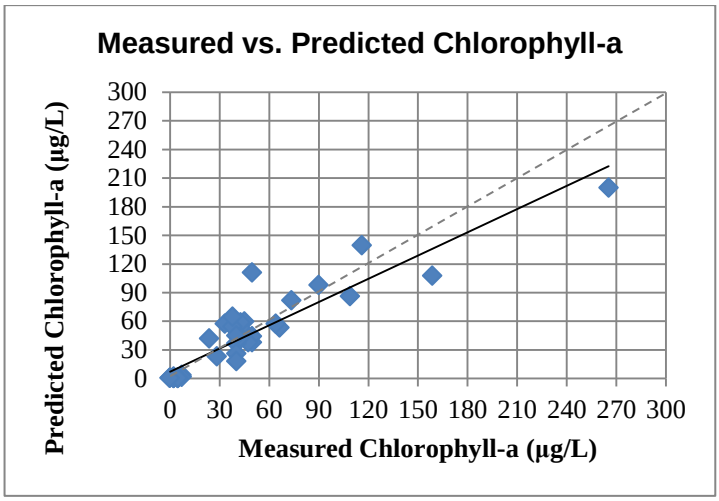

Figure 4-34: Measured versus Predicted Chlorophyll-a for the ANN Model with Hybrid Inputs (Validation Data Set – Combined Data)

As seen from Table 4-36, the coefficient of determination values for the test and validation datasets were 0.651 and 0.846, respectively, with very small p-values. Compared to regression models, ANN was significantly superior in handling the changes in the conditions of the lake. In order to analyze the prediction performance of the ANN model for different fields, the data belonging to each field were isolated and statistical analysis was conducted. Results for Field 1 are provided in Table 4-37. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-35, 4-36, 4-37, and 4-38, respectively. As can be seen, when the ANN model was trained with the dataset representing various conditions, applicability of the model to different conditions was possible compared to regression models. The coefficient of determination values for the testing and validation datasets were 0.998 and 0.742 respectively. The R^2 value of 0.998 can be ignored as it is achieved from a dataset containing 3 points only, while R^2 value of 0.742 can be classified as significant with a p-value close to 0.

Table 4-37: Application of the ANN model with Hybrid Inputs of the Combined Dataset to Field 1.

		Hybrid Model			
		All Data	Training	Testing	Validation
Field 1	# of Independent Parameters	11			
	# of Samples	25	12	3	10
	Coefficient of Determination	0.675	0.761	0.998	0.742
	Correlation Coefficient	0.821	0.872	0.999	0.861
	p-Value	3.288 E-47	3.211 E-11	4.771 E-24	3.493 E-19
	Root Mean Square Error	34.033	23.498	56.568	35.680

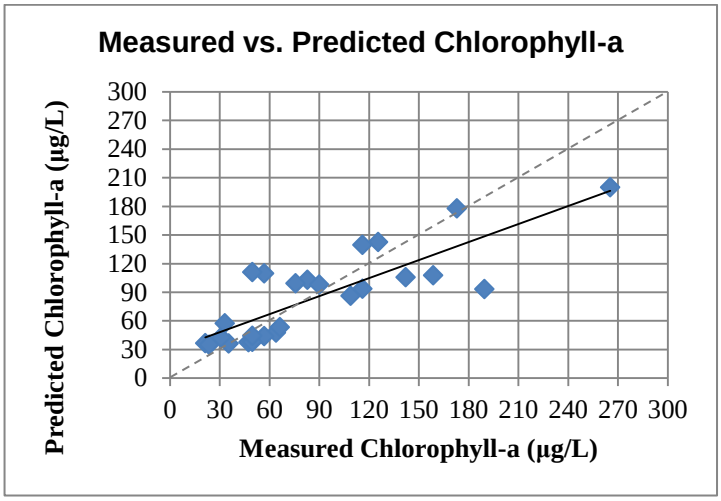

Figure 4-35: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 1)

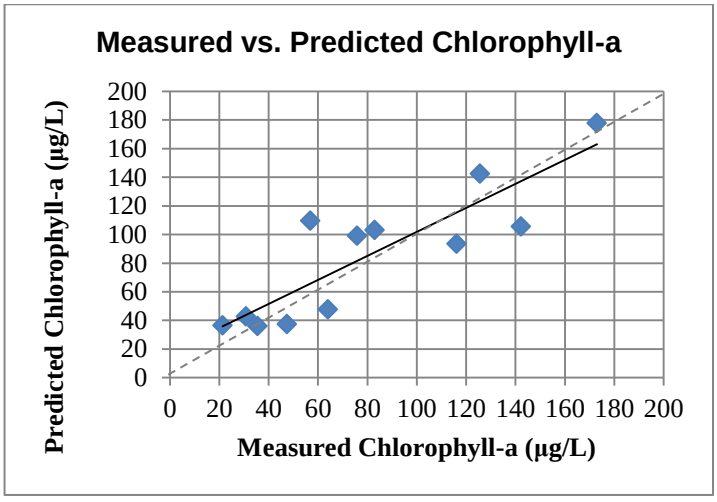

Figure 4-36: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 1)

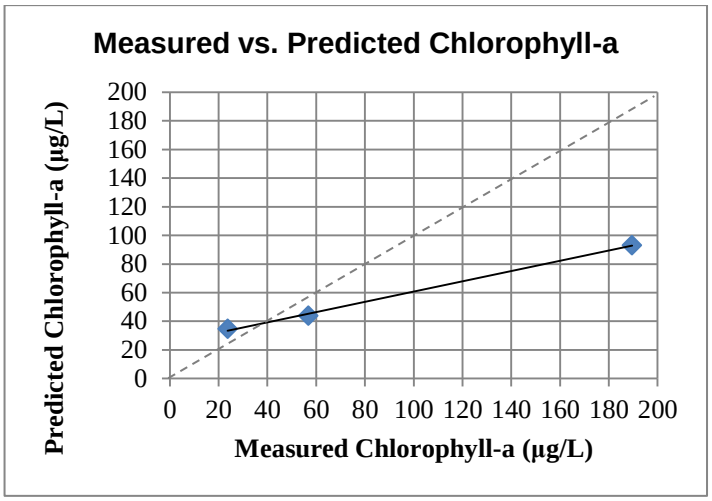

Figure 4-37: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 1)

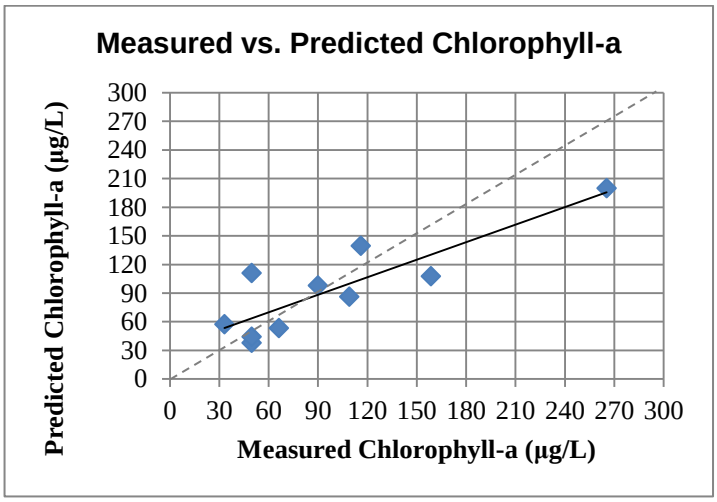

Figure 4-38: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 1)

Information about the performance of combined model over Field 2 are given in Table 4-38. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-39, 4-40, 4-41, and 4-42, respectively. The prediction performance for Field 2 was 0.365 for the validation set, however, unsatisfactory result was obtained for the test set. Therefore, the outcome was poor compared to the results for Field 1.

Table 4-38: Results Comparison Table for Field 2 (combined model)

		Hybrid Model			
		All Data	Training	Testing	Validation
Field 2	# of Independent Parameters	11			
	# of Samples	30	12	3	15
	Coefficient of Determination	0.179	0.187	0.001	0.365
	Correlation Coefficient	0.423	0.433	-0.029	0.604
	p-Value	4.199 E-37	1.007 E-24	3.399 E-08	2.354 E-10
	Root Mean Square Error	18.248	20.467	26.566	13.774

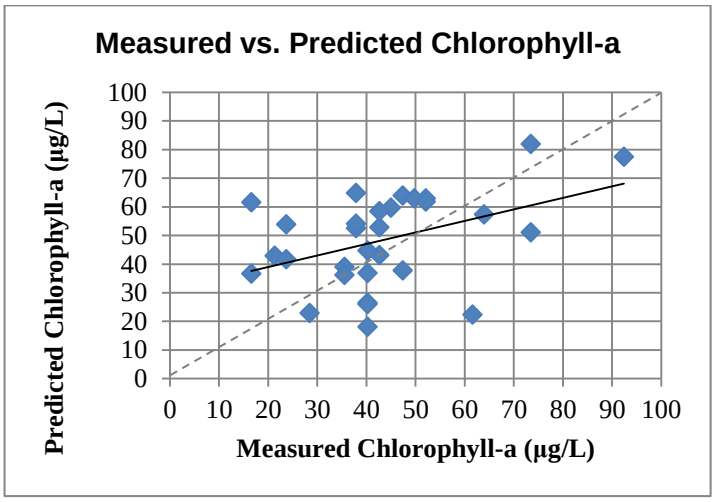

Figure 4-39: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 2)

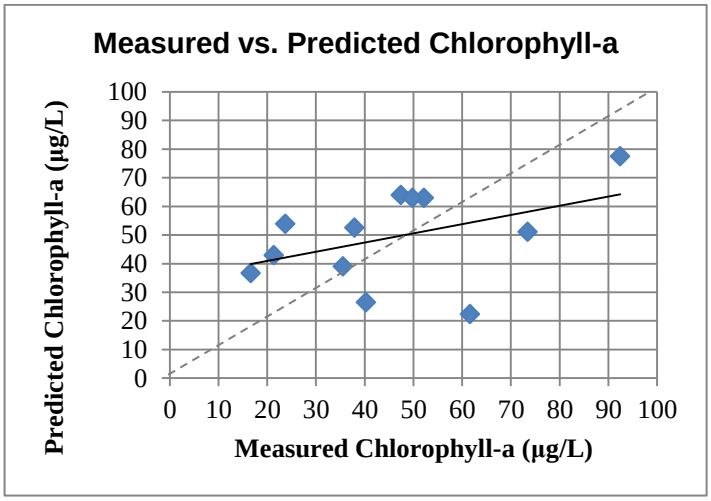

Figure 4-40: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 2)

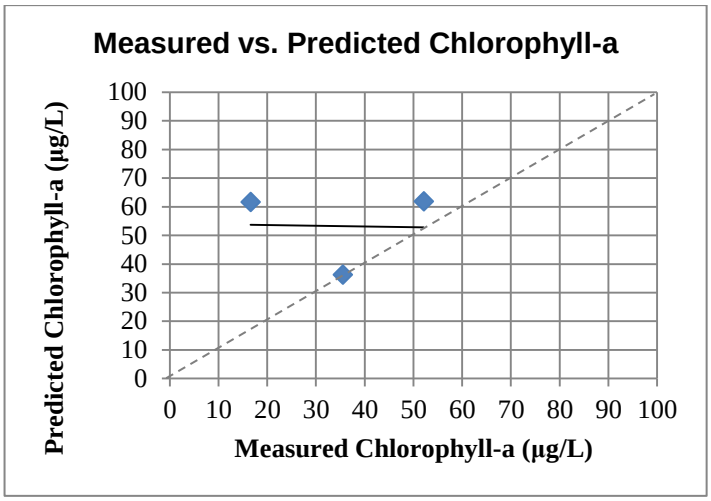

Figure 4-41: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 2)

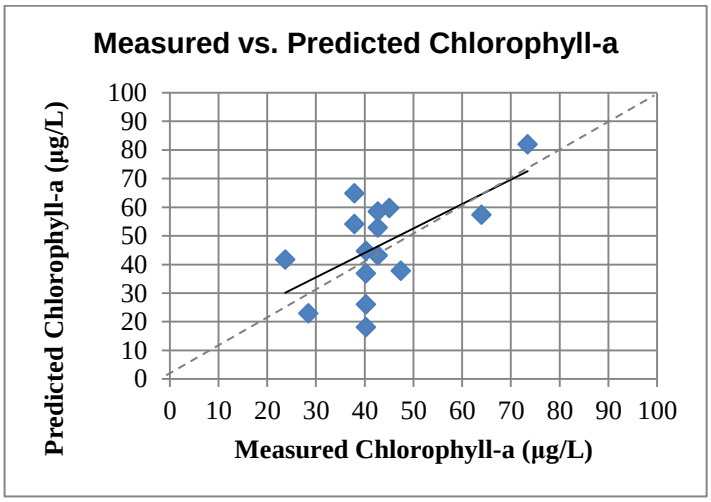

Figure 4-42: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 2)

When the data belonging to Field 3 in the combined model outputs are analyzed, it is seen that it was not possible to obtain high predictability for the chlorophyll-a concentrations in Field 3. The results are given Table 4-39. Measured versus predicted chlorophyll-a concentrations for all dataset, training set, testing set, and validation set are depicted in Figures 4-43, 4-44, 4-45, and 4-46, respectively. The coefficient of determination values for the testing and validation datasets are calculated as 0.758 and 0.327, respectively. High R^2 value achieved for the test data is due to the number of points (3 points) in the dataset.

Table 4-39: ANN Results Comparison Table for All Dataset, Third Field Data

		Hybrid Model			
		All Data	Training	Testing	Validation
Field 3	# of Independent Parameters	11			
	# of Samples	30	12	3	15
	Coefficient of Determination	0.307	0.328	0.758	0.327
	Correlation Coefficient	0.554	0.618	0.871	0.572
	p-Value	3.004 E-91	1.996 E-55	1.765 E-4	1.057 E-37
	Root Mean Square Error	3.925	4.770	3.337	3.222

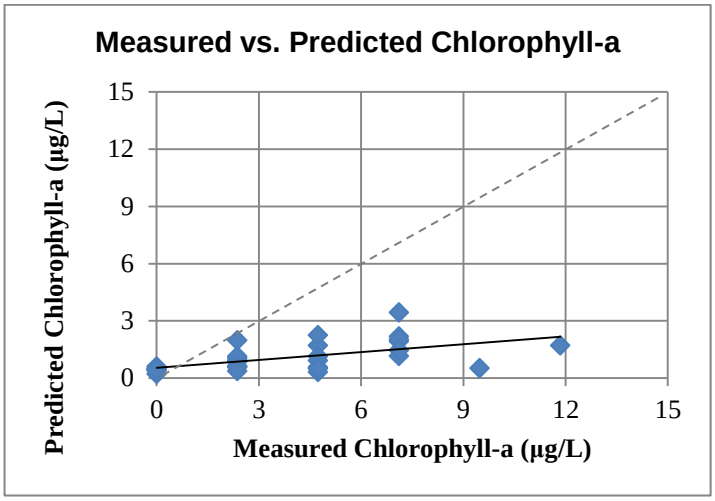

Figure 4-43: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (All Data Set - Field 3)

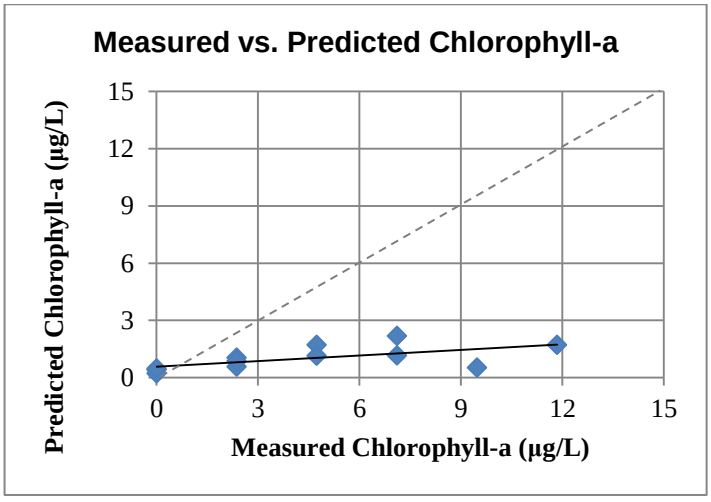

Figure 4-44: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Training Data Set - Field 3)

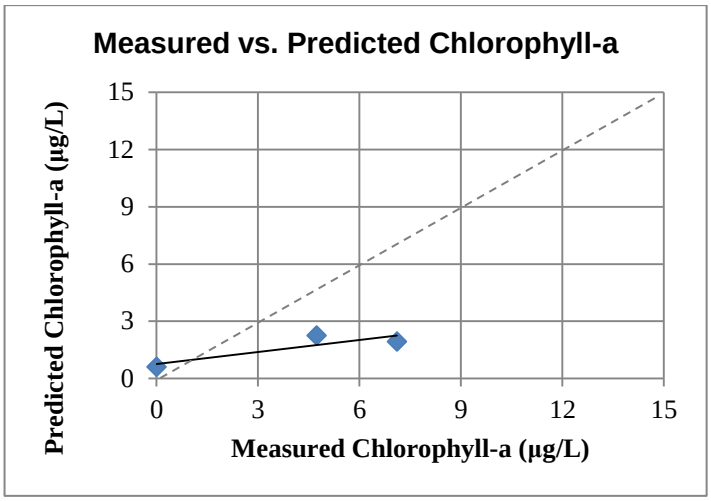

Figure 4-45: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Testing Data Set - Field 3)

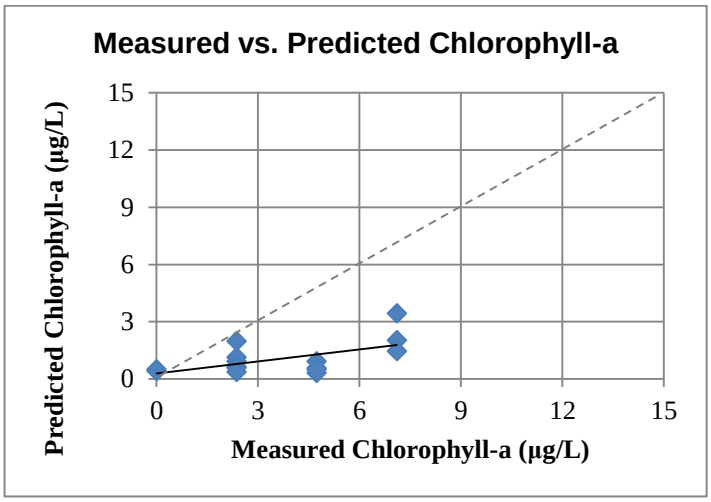

Figure 4-46: Measured versus Predicted Chlorophyll-a for the ANN Model of Combined Data with Hybrid Inputs (Validation Data Set - Field 3)

In Table 4-40, results for multivariate linear regression, non-linear regression and ANN models are summarized for the hybrid inputs. As can be seen from the table below, ANN model had the best result for the combined data. When p-values were analyzed, it was seen than all of them were close to 0, showing that predicted values were statistically significant and matched the measured values well. When RMSE values were calculated, the minimum error was achieved by the ANN model. With respect to the RMSE of the linear regression model, the RMSE values were 6% and 55% less for the non-linear regression and ANN models, respectively. Calculated significance f values were all found to be smaller than 0.05 in 95 % confidence level. In addition to this, the smallest f value was calculated for the ANN

model, which shows its better prediction capability. For all models, the hybrid inputs resulted in better predictability compared to using only the reflectance values for chlorophyll-a determination. This is due to the presence of other factors other than chlorophyll-a which would impact the reflectance values in Lake Eymir. It must also be stated that application of the models to each field separately did not result in satisfactory results, although some improvement was observed compared to previous cases.

Table 4-40: Summary Table for the Models with Hybrid Inputs

		Multivariate Linear Regression	Non-Linear Regression	Artificial Neural Network
All Data	**# of Independent Parameters**	11		
	# of Samples	85		
	Coefficient of Determination	0.630	0.674	0.798
	Correlation Coefficient	0.794	0.821	0.893
	p-Value	1.789 E-102	1.051 E-92	2.255 E-168
	Significance f-Value	1.326 E-19	6.742 E-22	1.493 E-30
	Root Mean Square Error	25.2	23.6	11.472

4.4.5. Overall Analysis

In the Table 4-41, results for the models with hybrid inputs are summarized for all three fields and the combined dataset.

Table 4-41: Hybrid Model Results Comparison Table

		Field # 1 Models			Field # 2 Models			Field # 3 Models			Combined Data Models		
		# of Ind. Par.		11	# of Ind. Par.		11	# of Ind. Par.		11	# of Ind. Par.		11
		LR	*NLR*	*ANN*	*LR*	*NLR*	*ANN*	*LR*	*NLR*	*ANN*	*LR*	*NLR*	*ANN*
Field # 1	# Sample	25			25			25			25		
	R^2	0.731	0.806	0.857	0.014	0.010		0.054	0.082		0.464	0.495	0.675
	RMSE	25.4	21.6	19.2	1.00E+2	1.26E+8		82.3	1.27E+3		38.9	36.7	34.0
	Correl. Cf.	0.855	0.898	0.926	0.119	0.098		-0.231	0.286		0.681	0.704	0.821
	p-Value	<0.05	<0.05	<0.05	<0.05	N/A		0.00	N/A		<0.05	<0.05	<0.05
Field # 2	# Sample	30			30			30			30		
	R^2	0.001	0.000		0.577	0.620	0.844	0.006	0.032		0.015	0.119	0.179
	RMSE	6.87E+2	2.96E+6		10.90	10.30	6.95	2.17E+1	6.86E+2		19.4	17.6	18.2
	Correl. Cf.	0.038	-0.018		0.760	0.787	0.919	0.078	0.180		0.124	0.345	0.423
	p-Value	N/A	N/A		<0.05	<0.05	<0.05	<0.05	N/A		<0.05	<0.05	<0.05
Field # 3	# Sample	30			30			30			30		
	R^2	0.036	0.027		0.226	0.010		0.644	0.675	0.726	0.001	0.001	0.307
	RMSE	1.23E+3	1.50E+3		1.65E+2	1.55E+12		1.80	1.72	1.69	12.7	12.3	3.92
	Correl. Cf.	-0.189	-0.163		-0.475	0.099		0.802	0.822	0.852	-0.030	0.026	0.554
	p-Value	N/A	N/A		N/A	N/A		>0.05	>0.05	>0.05	N/A	N/A	<0.05
Combined Data	# Sample										85		
	R^2										0.630	0.674	0.798
	RMSE										25.2	23.6	11.5
	Correl. Cf.										0.794	0.821	0.893
	p-Value										<0.05	<0.05	<0.05

Surface chlorophyll-a concentrations in the first field study has an average value of 78.946 µg/L. In other fields, this value decreased where in Field 3 the lowest value was observed. It is seen that the performances of the models in predicting the chlorophyll-a concentrations decreased as the average chlorophyll-a concentrations decreased, especially for the regression models. Table 4-41 shows that models that are derived based on the data for Field 1, Field 2 and Field 3 show lower correlations in other fields. This is possibly a reason of differences in the ranges of the model parameters and significantly different conditions in the fields. Therefore, the models generated for Lake Eymir in this study are case and time specific. Based on the results, it can be said that higher chlorophyll-a concentrations can be modeled easier. As chlorophyll-a concentrations in the lake surface decreases, predictive capacity of the models also decreased. This decrease in the generalized applicability of the models can be explained by following conditions,

- In Field 1 and Field 2, there was a wider range of chlorophyll-a concentration distributions. In Field 3, there were only 6 different concentrations, which resulted in overlapping and lead to statistically insignificant results in the linear, non-linear regression and artificial neural network models. In the first dataset, the standard deviation in chlorophyll-a concentration was 49.928 µg/L, while this value was 16.998 µg/L and 3.069 µg/L in the second and third field study data, respectively.

- As mentioned before there was time difference between the dates the satellite images were acquired and the dates of sampling and measurements. Although up to 10 day difference had been reported in studies that employ remotely sensed data for chlorophyll-a determination, surface blooms may travel to different locations as a result of wind and wave actions in the time gap. This may have an adverse effect on the results. In addition, surface reflectance changes due to wind action, resuspension of bottom sediments, morning haze including high levels of humidity may also impact the results when delay is realized in sampling. Basically, the data ground or water truth data used in remote sensing applications are critical. Even though the wind speed was included in the hybrid inputs, it had a random distribution over the lake surface since readings at different points were realized at different times . In the stereo image present in Figure 4-47, possible wind directions that can generate the waves are shown.

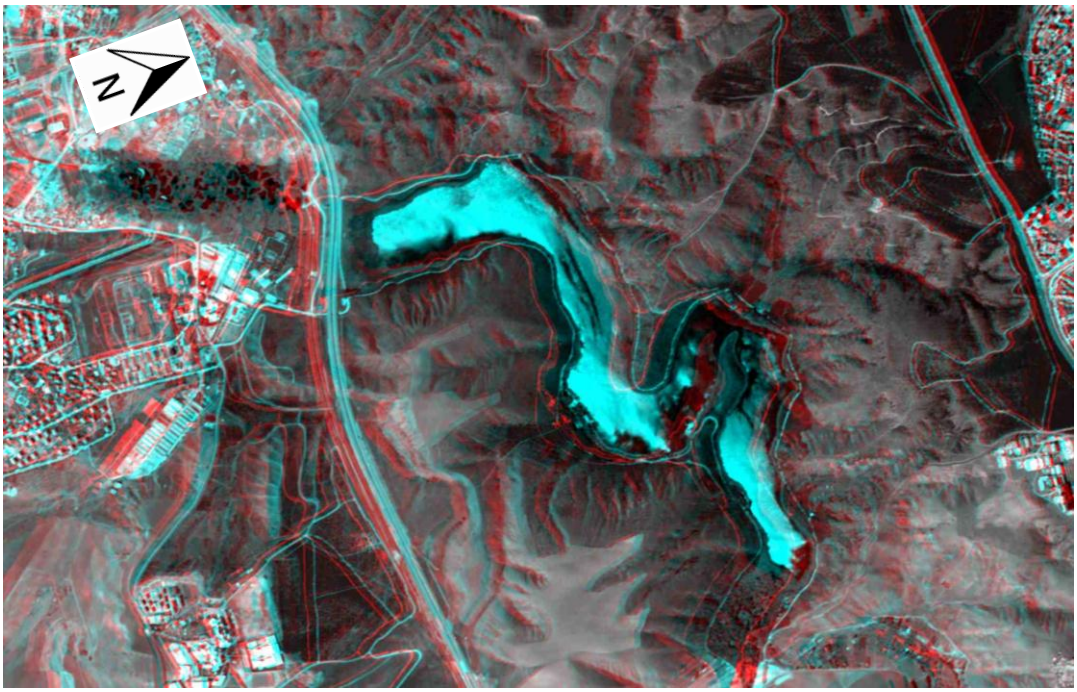

Figure 4-47: Stereo Image of Lake Eymir

3D version of the image can be viewed by using the Red-Cyan stereo glasses provided at the back page of this thesis. As can be seen, the lake is surrounded by small hills with an average height of 1050 m, while the water surface has an elevation of 950 m. This 50 m difference can prevent wind direction throughout the lake surface. Because of this height difference, the wind directions that can generate waves are limited to winds blowing from south and east when the wind speed is low. Waves can easily affect the reflectance values in the surface as they disturb the water surface and changes the angle of reflected light.

The yellow arrows in Figure 4-48, represent the path of satellite, while the red line shows the potential wave direction in the lake. As all of the images are taken in the morning hours at between 9:00 AM and 10:00 AM, sunlight should be coming from an inclined angle from the east direction. So, presence of waves can easily alter the reflectance values in different parts of the lakes. The impacts of haze and humidity are minimized with atmospheric correction, which is applied when the pixel values were converted to reflectance values.

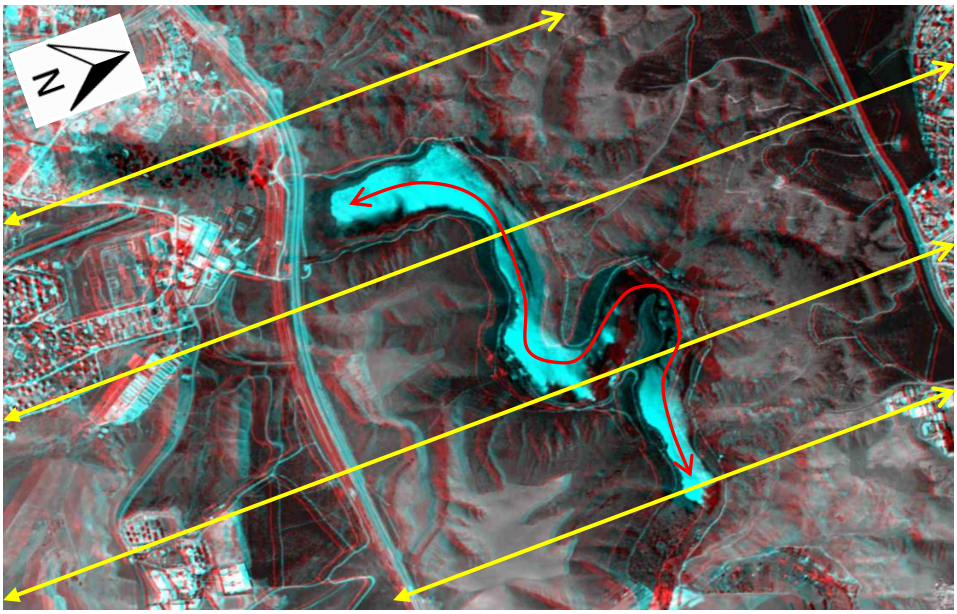

Figure 4-48: Satellite Pass Path and Wave Directions in Lake Eymir

The reflectance from the bottom of the lake and resuspension of bottom sediment are other interferences that might have impacted the results.

Differences in the performances of the models are also originating from different conditions in the lake. Yet, it is clearly seen that in general the hybrid inputs provided better predictability since chlorophyll-a was not the only parameter impacting the reflectance values. Better results for higher chlorophyll-a concentrations indicate that as well.

CHAPTER 5

CONCLUSIONS and RECOMMENDATIONS FOR FUTURE STUDIES

For three different field cases considered, the ranges of chlorophyll-a concentrations were different. The highest range was measured in Field 1, followed by Field 2 and Field 3. Similar to this rank, models with the best fits were achieved for Field 1. Field 2 and Field 3 had the 2nd and 3rd ranks. These results indicate that use of remotely sensed data for chlorophyll-a predictions can be easier when concentrations are higher.

Results obtained in this study state that use of inputs that may impact the reflectance values together with reflectance values may increase the performance of remote sensing applications for monitoring of chlorophyll-a concentrations in inland waters. By including other parameters in model development, better results can be achieved. When different models are examined that ANN models had superiority over regression models in general in predicting chlorophyll-a concentrations. This is due to the fact that ANNs can handle noisy data and capture the pattern between inputs and outputs.

In this study, we have applied concentrated sampling 30 different points, which is equal to 24 sampling point for km^2 and higher than similar studies conducted before. Increase in the number of sampling points may result in better prediction of the spatial distribution based on the t-values obtained for the coefficients in the linear and non-linear regression models. However, model results might also be affected from errors in laboratory and field analyses.

As a result of the study, it has been once more seen that remote sensing is an effective tool for the determination of water quality parameters, although problems and interferences can be observed. In this study it is shown that for

shallow lakes, such as Lake Eymir, inclusion of easily measurable water quality and atmospheric parameters may significantly improve the prediction performances of the models. The use of hybrid models in remote sensing of water quality parameters can be an effective approach in modeling of the chlorophyll-a concentrations. Development of such models can be important for sites requiring frequent monitoring.

Use of hybrid inputs brought improvements to the models. The models generated with hybrid inputs can be used in lakes where chlorophyll-a has a heterogeneous distribution using smaller number of sampling locations. Since the independent parameters in the generated models are easily measurable, spatial distribution of chlorophyll-a throughout the lake surface can be obtained quickly with a satellite imagery and field measurements from discrete locations. Monitoring of the spatial distribution of chlorophyll-a in the lake with remote sensing and modeling can be advantageous compared to traditional sampling procedures which employs limited number of samples due to sampling costs. If there had been automatic stations present in the lake that transmits real-time data, then it would have been much more convenient to apply remote sensing and in-situ measurements in modeling of the distribution of chlorophyll-a in the lake. As a result, agencies responsible from monitoring of the status in the lake can assess the present conditions of chlorophyll-a concentrations with much less operating costs due to lower field measurement needs. Moreover, they also can observe the occurrence of algal blooms and their locations faster.

Use of remotely sensed data proved its applicability in water quality management studies. This technology can also be applied easily in determination of total suspended solid concentration, turbidity, surface water temperature, colored dissolved organic matter, chlorophyll-b and chlorophyll-c concentration, total phosphorus concentration, depth of water column and light attenuation coefficient in Lake Eymir. Models for these parameters can also be generated and applied for assessment of water quality. Though, some points are still needed to be considered for better predictability. Firstly, date of field study and the date satellite imagery taken should be close to each other as much as possible, although this may not be controllable. As time passes, inconsistencies may occur in the ground or water truth data. Therefore, especially for lakes of complex nature, such as Lake Eymir, field spectrophotometer can be used in the analysis as well, which can provide reflectance values in a defined wavelength simultaneously with other

measurements. This can also eliminate the interferences from atmosphere and errors originating from GPS in determination of the sampling point coordinates.

As another recommendation for future studies, more parameters can be included in the analysis. By considering new satellite systems, higher number of spectral bands can be used in the model developments. In addition, band ratios can be used in regression and ANN models. As an example, the coastal blue band in World View 2 satellite can be used in bathymetric analysis, which then can be included in model development for chlorophyll-a determination in shallow lakes along with blue band to consider bottom vegetation and sediment properties.

REFERENCES

Abdi, Hervé, and Lynne J Williams. 2010. Principle Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics In Press. doi:10.1002/wics.101.

Altınbilek, D., Kutoğlu, Y., Soyupak S., Yazıcıgil, H., Usul, N., Doyuran, V., Göğüş¸, M., Gökçay C.F., Günyaktı, A., Özsan E., Sürücü G., Merzi,N. (1995) Gölbaşı Mogan- Eymir Gölleri için su kaynakları ve çevre yönetim planı projesi. Teknik Rapor. No. 93-03-03-04-01, Orta Doğu Teknik Üniversitesi, Ankara.

Ankara, Turkey Forecast : Weather Underground. (n.d.). Welcome to Weather Underground : Weather Underground. from http://www.wunderground.com/cgi-bin/findweather/getForecast?query=ankara&wuSelect=WEATHER, Retrieved August 19, 2010

Baruah, Pranab J, Masayuki Tamura, Kazuo Oki, and Hitoshi Nashimura. (2001). Neural Network Modeling of Lake Surface Chlorophyll And Sedıment Content from Landsat TM Imagery. In 22nd Asian Conference on Remote Sensing, 5 - 9. Singapore.

Barrett, E.C., and Curtis L.F., (1992), Introduction To Environmental Remote Sensing, Chapman and Hall, Inc., Singapore.

Beisl, U.; Telaar, J.; von Schönemark, M. Atmospheric correction, reflectance calibration and BRDF correction for ADS40 image data. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Proceedings of the XXI ISPRS Congress, Commission VII, Beijing, China, July 3-11, 2008; 37(B7)

Beklioglu, (2000). Benthi-planktivorous fish-induced low water quality of Lake Eymir before biomanipulation. Turkish journal of zoology, 24(3), 315.

Beklioglu, (2003). Restoration of the eutrophic Lake Eymir, Turkey, by biomanipulation after a major external nutrient control I. Hydrobiologia, 490(1), 93.

Beklioglu, M, O Ince, and I Tuzun. (2003). Restoration of Eutrophic Lake Eymir, Turkey, by biomanipulation undertaken after a major external nutrient control I. Hydrobiologia 489: 93-105.

Beklioglu, M., Akkas S. B., Ozcan E. H., Bezirci G., and Togan I.. (2010). Effects of 4-nonylphenol, fish predation and food availability on survival and life history traits of Daphnia magna straus. Ecotoxicology (London, England).

Beklioglu, Meryem. (2000). Does manipulation work for warm-temperate lakes? First case study in an eutrophic Turkish lake. In Lake Issyk-Kul: Its Nautal Environment, ed. Jean Klerkx and Beishen Imanackunov, 207-215. Kluwer Academic Publishers.Bernstein, Robert L, Larry Breaker, and Robert Whritner. (1977). California Current Eddy Formation: Ship, Air, and Satellite Resuls. Advancement Of Science 195, no. 4276: 353-359.

Bilge, F., B. Yazıcı, T. Dogeroglu, and C. Ayday. (2003). Statistical evaluation of remotely sensed data for water quality monitoring. International Journal of Remote Sensing 24, no. 24: 5317-5326.

Blanco, Alfonso, and William E. Roper. (2007). Remote Sensing Techniques to detect Surface Water Quality Constituents in Coastal and Inland Water Bodies from Point or Non Point Pollution Sources.pdf. In WEF TEC.

Bricaud, A, E Bosc, and D Antoine. (2002). Algal biomass and sea surface temperature in the Mediterranean Basin Intercomparison of data from various satellite sensors, and implications for primary production estimates. Remote Sensing of Environment 81, no. 2-3 (August): 163-178.

Camur, M Zeki, Hasan Yazicigil, and Doğan Altınbilek. (1997). Hydrogeochemical Modeling of Waters in Mogan and Eymir Lakes Special Environmental. Water Environment Research 69, no. 6: 1144-1153.

Cannizzaro, J, and K Carder. (2006). Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters. Remote Sensing of Environment 101, no. 1: 13-24.

Cauwer, Vera De, Kevin Ruddick, Youngje Park, Bouchra Nechad, and Michael Kyramarios. (2004). Optical remote sensing in support of eutrophication monitoring in the southern north sea. EARSeL eProceedings, no. 3: 208-221.

Cui, T, J Zhang, L Sun, and W Zhao. (2008). Neural network model for chlorophyll-a concentration retrieval in the bohai sea, no. April.

Dall'olmo, G, A Gitelson, D Rundquist, B Leavitt, T Barrow, and J Holz. (2005). Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. Remote Sensing of Environment 96, no. 2: 176-187.

Demirci, G., Aslan, M., Elahdab, T., & Aksoy, A. (2005). Eymir Gölü'ndeki Su Kalitesinin İzlenmesi ve Değerlendirilmesi. Proceedings of 6th National Environmental Engineering Congress (pp. 72-76). Istanbul: CMO.

Demuth, Howard, Mark Beale, and Martin Hagan. Neural Network Toolbox™ 6. MATLAB.

Dien, Tran Van, Dan Ling Tang, and Hiroshi Kawamura. (2002). Validation of SeaWiFS-derived Ocean Color Data and Using for Study Distribution and Seasonal Variation of Chlorophyll-a Concentration in the Vietnam Waters. Remote Sensing and Geographic Information Systems (GIS) Applications for Sustainable Development: 32-37.

DigitalGlobe | DigitalGlobe: QuickBird Satellite - 60cm Resolution. (n.d.). DigitalGlobe | An Imagery and Information Company, 2010, from http://www.digitalglobe.com/index.php/85/QuickBird, Retrieved December 8

DigitalGlobe | DigitalGlobe: Worldview-2 Satellite - 46cm Resolution. (n.d.). DigitalGlobe | An Imagery and Information Company. from http://www.digitalglobe.com/index.php/88/WorldView-2, Retrieved December 20, 2010

Diker, Z. (1992). A Hydrobiological and ecological study in Lake Eymir, MSc thesis, METU, Ankara

Ekercin, Semih. (2007). Water Quality Retrievals from High Resolution Ikonos Multispectral Imagery: A Case Study in Istanbul, Turkey. Water, Air, and Soil Pollution 183, no. 1-4: 239-251.

Elahdab, Tarek. (2006). Investigation of algae distribution in Eymir Lake using site measurements and remotely sensed data, MSc thesis, METU, Ankara

EPASA. 2006. Envionmental Protection Agency for Special Areas, Golbasi. http://www.ockkb.gov.tr/tr/Icerik.ASP?ID=132, Last visited on: January 11, 2011

Fish, G R. (1969). Lakes - The Value of Recent Research to Measure Eutrophication and to Indicate Possible Causes.

Garcia, V. M. T., S. Signorini, C. a. E. Garcia, and C. R. McClain. 2006. Empirical and semi- analytical chlorophyll algorithms in the southwestern Atlantic coastal region (25–40°S and 60–45°W). International Journal of Remote Sensing 27, no. 8 (April): 1539-1562.

Geldiay, R., (1949). Çubuk barajı ve Eymir Gölü'nün makro ve mikro faunasının mukayeseli incelenmesi. Ankara Üniversitesi Fen Fakültesi Mecmuası, Cilt 2.

S., L., Greenberg, A., & Eaton, A. (1998). Standard Methods for the Examination of Water and Wastewater. Washington: APHA-AWWA-WEF.

Gross, L, Sylvie Thiria, and Robert Frouin. (1999). Applying artificial neural network methodology to ocean color remote sensing. Ecological Modelling 120, no. 2-3 (August): 237-246.

Gross, Lidwine, Sylvie Thiria, Robert Frouin, and B. Greg Mitchell. (2000). Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. Paris. Paris.

GPS Review – Magellan SporTrak. (2004, June 6). Maps and GPS Info - Useful Information on Maps and GPS. from http://www.maps-gps-info.com/gp-rvw-mag-sptrk.html, Retrieved December 21, 2010

Han, Luoheng, and Donald Rundquist. (1997). Comparison of NIR/RED ratio and first derivative of reflectance in estimating algal-chlorophyll concentration: A case study in a turbid reservoir. Remote Sensing of Environment 62, no. 3 (December): 253-261.

Heege, Thomas, and Jürgen Fischer. (2004). Mapping of water constituents in Lake Constance using multispectral airborne scanner data and a physically based processing scheme. Canadian Journal of Remote Sensing 30, no. 1: 77-86.

Hough, H. (1991). Fundamentals of satellite imaging. Satellite surveillance (pp. 23-38). Port Townsend, Wash.: Loompanics Unlimited.

How Does GPS Work?. (1998). Smithsonian National Air and Space Museum. from http://www.nasm.si.edu/gps/work.html, Retrieved December 21, 2010

Image Processing and Analysis Solutions. (n.d.). ENVI Software. Retrieved September 9, 2010, from www.ittvis.com/ProductServices/ENVI.aspx.

Jonge, Victor N De, M Elliott, and E Orive. (2002). Causes, historical development, effects and future challenges of a common environmental problem: eutrophication. Hydrobiologia 475/476: 1-19.

Karakoç, Gamze, Figen Unlü Erkoç, and Hikmet Katircioğlu. (2003). Water quality and impacts of pollution sources for Eymir and Mogan Lakes (Turkey). Environment international 29, no. 1 (April): 21-7.

Karul, C., Soyupak S., Çilesiz A. F., Akbay N., and Germen E.. (2000). Case studies on the use of neural networks in eutrophication modeling. Ecological Modelling 134, no. 2-3 (October): 145-152.

Keiner, L. (1998). A Neural Network Model for Estimating Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery. Remote Sensing of Environment 66, no. 2 (November): 153-165.

Keiner, L. (1998). A Neural Network Model for Estimating Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery. Remote Sensing of Environment 66, no. 2 (November): 153-165.

Kishino, M, a Tanaka, and J Ishizaka. (2005). Retrieval of Chlorophyll , suspended solids, and colored dissolved organic matter in Tokyo Bay using ASTER data. Remote Sensing of Environment 99, no. 1-2: 66-74.

Kline, Paul. (1994). An Easy Guide to Factor Analysis. London: Routledge.

Kneale, Pauline E. (2004). Neural Networks for Hydrological Modelling. -.

Larsson, Ulf, Ragnar Elmgren, and Fredrik Wulff. (1985). Eutrophication and the Baltic Sea: Causes and Consequences. Ambio 14, no. 1: 9-14.

Le, Chengfeng, Yunmei Li, Yong Zha, Deyong Sun, Changchun Huang, and Heng Lu. (2009). A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: The case of Taihu Lake, China. Remote Sensing of Environment 113, no. 6: 1175-1182.

Lin, Q, Y Zhang, Y Nie, and Y Guan. (2003). Detection of harmful algal blooms over the Gulf of Bohai Sea in China at visible and near infrared (NIR) wavelengths of remote sensing. Journal of Electromagnetic Waves and Applications 17, no. 6: 861-871.

Liu, Yansui, Md Anisul Islam, and Jay Gao. (2003). Quantification of shallow water quality parameters by means of remote sensing. Progress in Physical Geography 27, no. 1: 24-43.

Marullo, Salvatore, Maria Ragni, and Rosalia Santoleri. (2002). Validation of empirical SeaWiFS algorithms for chlorophyll-a retrieval in the Mediterranean Sea A case study for oligotrophic seas `. Remote Sensing of Environment 82: 79 - 94.

Mittenzwey, K.-H., S. Ullrich, a. a. Gitelson, and K. Y. Kondratiev. (1992). Determination of chlorophyll a of inland waters on the basis of spectral reflectance. Limnology and Oceanography 37, no. 1: 147-149

Mogan ve Eymir Arasındaki Kanal Temizleniyor. (2010, October 1). Ankara Gölbaşı Gazetesi., from www.golbasigazetesi.com/haber_oku.asp?puan=4&haber=2643#, Retrieved October 17, 2010

Moisan, John R, Arthur J Miller, Emanuele Di Lorenzo, and John Wilkin. (2005). Modeling and Data Assimilation. In Remote Sensing and Digital Image Processing, 229-257. 7th ed. Springer Netherlands.

Muluk, Cagri B., and Meryem Beklioglu. (2005). Absence of typical diel vertical migration in Daphnia: varying role of water clarity, food, and dissolved oxygen in Lake Eymir, Turkey. Hydrobiologia 537, no. 1-3 (March): 125-133.

Musavi, Mohamad T., Richard L. Miller, Habtom Ressom, and Padma Natarajan. (2002). Neural network-based estimation of chlorophyll-a concentration in coastal waters. Proceedings of SPIE 4488, no. 207: 176-183.

Ormeci, C., Sertel, E. and Sarikaya, O., (2009). Determination of chlorophyll-a amount in Golden Horn, Istanbul, Turkey using IKONOS and in situ data. Environmental monitoring and assessment 155, no. 1-4: 83-90.

Ozen, A. (2006). Role of hydrology, nutrients and fish predation in determining the ecology of a system of shallow lakes, MSc thesis, METU, Ankara

Özaydın, Vehbi, Uygur Şendil, and Doğan Altınbilek. (2001). Stable Isotope Mass Balance Method to Find the Water Budget of a Lake ¨. Environmental Science 25: 329 - 344.

Panda, S. S., Vijay Garg, and Indrajeet Chaubey. (2004). Artificial Neural Networks Application in Lake Water Quality Estimation Using Satellite Imagery. Journal of Environmental Informatics 4, no. 2 (December): 65-74.

Pozdnyakov, D, R Shuchman, a Korosov, and C Hatt. (2005). Operational algorithm for the retrieval of water quality in the Great Lakes. Remote Sensing of Environment 97, no. 3: 352-370.

Ryther, J. H., and C. S. Yentsch. (1957). The Estimation of Phytoplankton Production in the Ocean from Chlorophyll and Light. Limnology and oceanography 2, no. 3: 281-286.

Slade, W.H., R.L. Miller, H. Ressom, and P. Natarajan. (2002). Ensemble neural network methods for satellite-derived estimation of chlorophyll α. Proceedings of the International Joint Conference on Neural Networks, 2003.. Ieee.

Smith, Val H, and David W Schindler. (2009). Eutrophication science: where do we go from here? Trends in ecology & evolution (Personal edition). Vol. 24.

Smith, Val H, Samantha B Joye, and Robert W Howarth. (2006). Eutrophication of freshwater and marine ecosystems. Limnology and oceanography 51: 351-355.

Spears, Bryan M, Laurence Carvalho, Rupert Perkins, and David M Paterson. 2008. Effects of light on sediment nutrient flux and water column nutrient stoichiometry in a shallow lake. Water research 42, no. 4-5: 977-86.

Stergiou, C., & Siganos, D. (n.d.). Neural Networks. Computing., from http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
Retrieved December 20, 2010

Sudheer, K.P., Indrajeet Chaubey, and Vijay Garg. (2006). Lake Water Quality Assessment From Landsat Thematic Mapper Data Using Neural Network: an Approach To Optimal Band Combination Selection 1. Journal of the American Water Resources Association 42, no. 6 (December): 1683-1695.

Tan, C, and M Beklioglu. (2005). Catastrophic-like shifts in shallow Turkish lakes: a modeling approach. Ecological Modelling 183, no. 4: 425-434.

Tan, C. O., & Orta Doğu Teknik Üniversitesi (Ankara, Turkey). (2002). The roles of hydrology and nutrients in alternative equlibria of two shallow lakes of anatolia, lake Eymir and lake Mogan: Using monitoring and modeling approaches. Ankara.

Tanaka, Akihiko, Motoaki Kishino, Roland Doerffer, Helmut Schiller, Tomohiko Oishi, and Tadashi Kubota. (2004). Development of a Neural Network Algorithm for Retrieving Concentrations of Chlorophyll, Suspended Matter and Yellow Substance from Radiance Data of the Ocean Color and Temperature Scanner. Journal of Oceanography 60, no. 3: 519-530.

Wang, Hongqing, C. M. Hladik, Wenrui Huang, K. Milla, L. Edmiston, M. a. Harwell, and J. F. Schalles. (2010). Detecting the spatial and temporal variability of chlorophyll-a concentration and total suspended solids in Apalachicola Bay, Florida using MODIS imagery. International Journal of Remote Sensing 31, no. 2: 439-453.

Wang, Tai-Sheng, Chih-Hung Tan, Li Chen, and Yu-Chu Tsai. (2008). Applying Artificial Neural Networks and Remote Sensing to Estimate Chlorophyll-a Concentration in Water Body. 2008 Second International Symposium on Intelligent Information Technology Application: 540-544.

Wang, Tai-Sheng, Chih-Hung Tan, Li Chen, and Yu-Chu Tsai. (2008). Applying Artificial Neural Networks and Remote Sensing to Estimate Chlorophyll-a Concentration in Water Body. 2008 Second International Symposium on Intelligent Information Technology Application: 540-544.

White, Paul A., Jacob Kalff, Joseph B. Rasmussen, and Josep M. Gasol. 1991. The effect of temperature and algal biomass on bacterial production and specific growth rate in freshwater and marine habitats. Microbial Ecology 21, no. 1 (December): 99-118.

Wu, Guangyu. (1998). Seasonal Change Detection of Water Quality in Texas Gulf Coast Using MODIS Remote Sensing Data. Development, no. 1: 1-22.

Wu, Min, Wei Zhang, Xuejun Wang, and Dinggui Luo. (2009). Application of MODIS satellite data in monitoring water quality parameters of Chaohu Lake in China. Environmental monitoring and assessment 148, no. 1-4: 255-64.

Yagbasan, Ozlem, and Hasan Yazicigil. 2008. Sustainable management of Mogan and Eymir Lakes in Central Turkey. Environmental Geology 56, no. 6: 1029-1040.

Yang, He, and Qian Du. (2009). Neural network approach for mobile bay water quality mapping with spaceborne measurements. Proceedings of SPIE 7343: 73430I-73430I-11.

Yang, Xiao-e, Xiang Wu, Hu-lin Hao, and Zhen-li He. (2008). Mechanisms and assessment of water eutrophication. Journal of Zhejiang University. Science. B. Vol. 9.

Zhang, Y, J Pulliainen, S Koponen, and M Hallikainen. (2003). Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland. IEEE Trans. Geosci. Remote Sens 41: 622-629.

Zhang, Yuanzhi, Jouni Pulliainen, Sampsa Koponen, and Martti Hallikainen. (2002). Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. Remote Sensing of Environment 81: 327 - 336.

APPENDICES

SPATIAL CHLOROPHYLL-A DISTRIBUTIONS

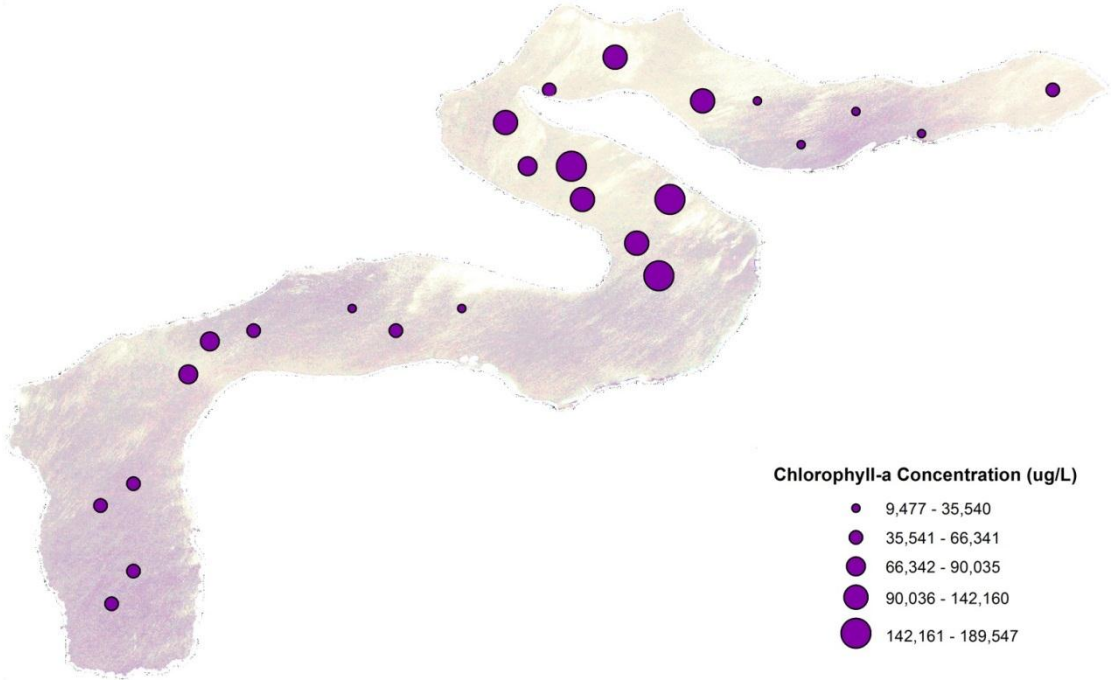

Figure A-1: First Field Study Measured Chl-a Distribution

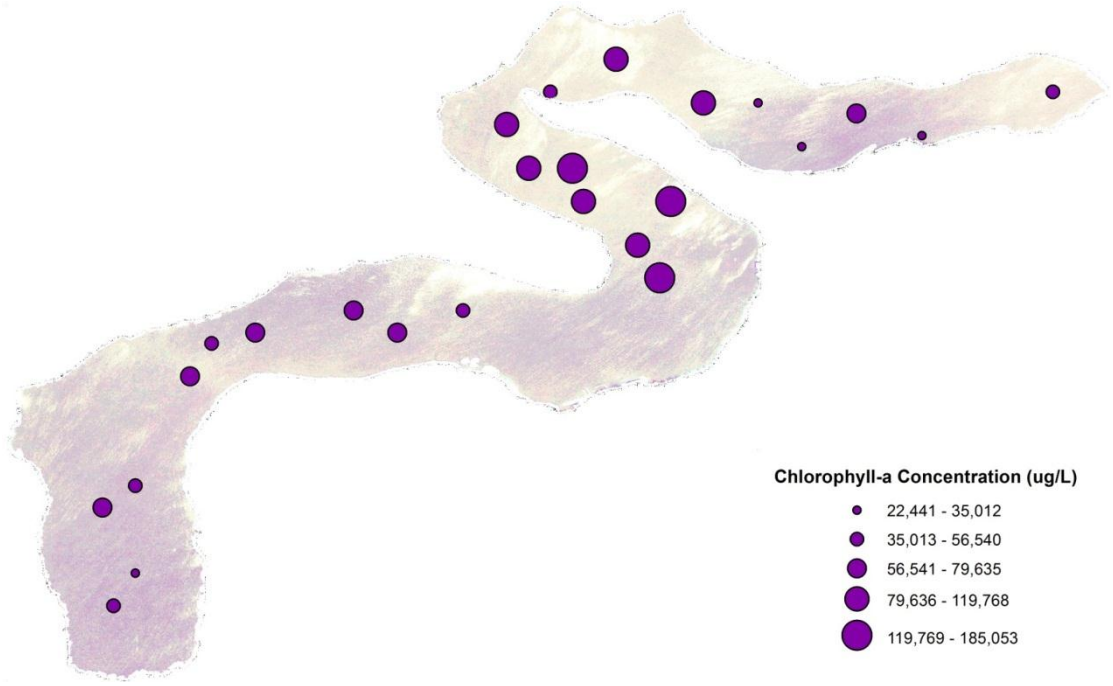

Figure A-2: First Field Study Linear Regression Predicted Chl-a Distribution

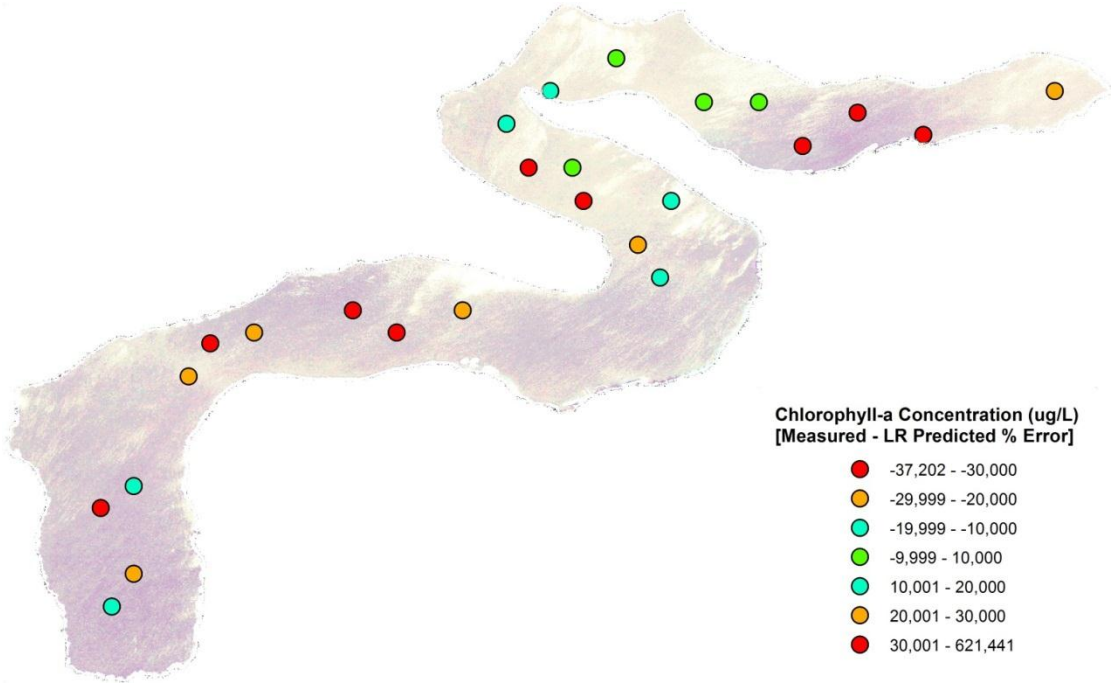

Figure A-3: First Field Study % Deviation of Linear Regression Predicted from Measured Chl-a Distribution

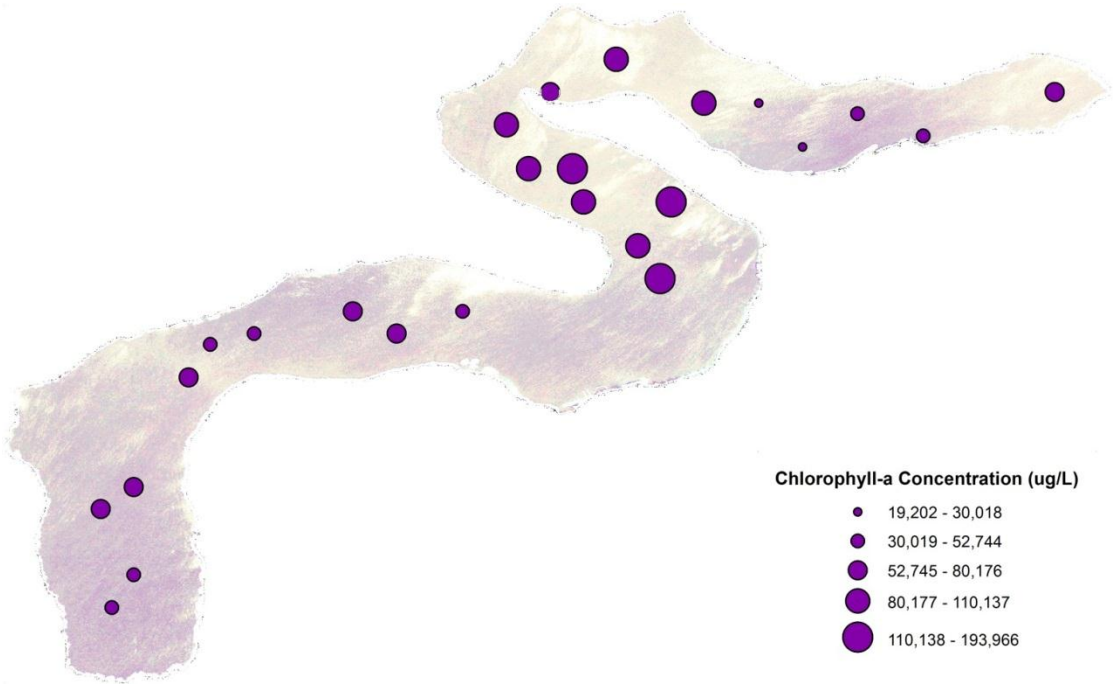

Figure A-4: First Field Study Non-Linear Regression Predicted Chl-a Distribution

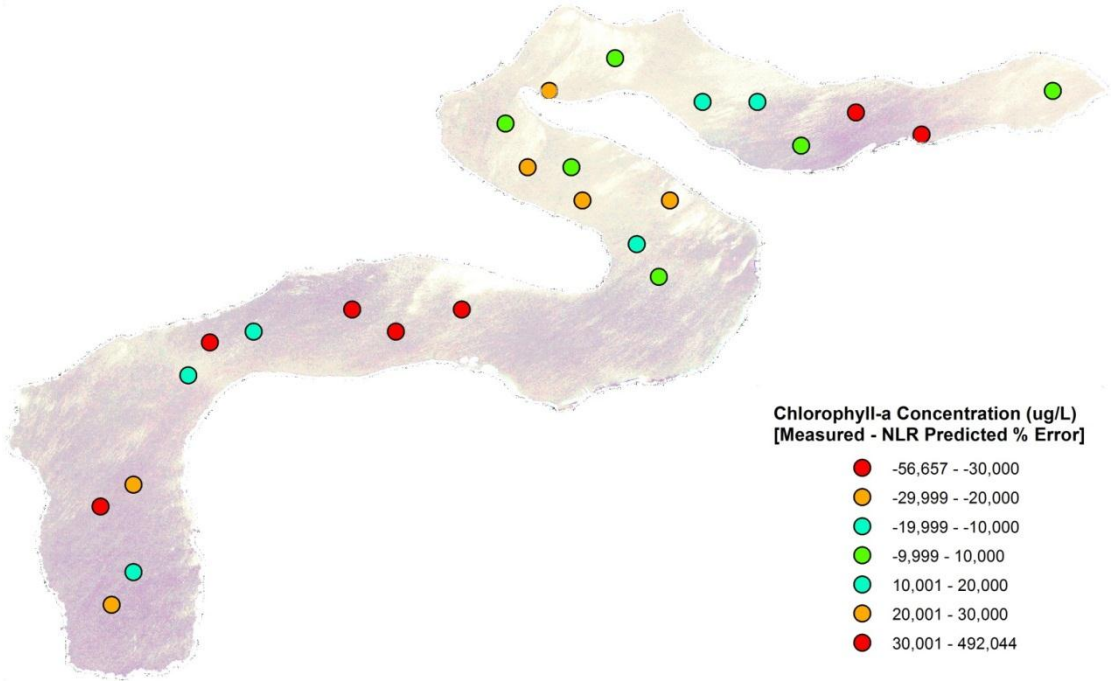

Figure A-5: First Field Study % Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution

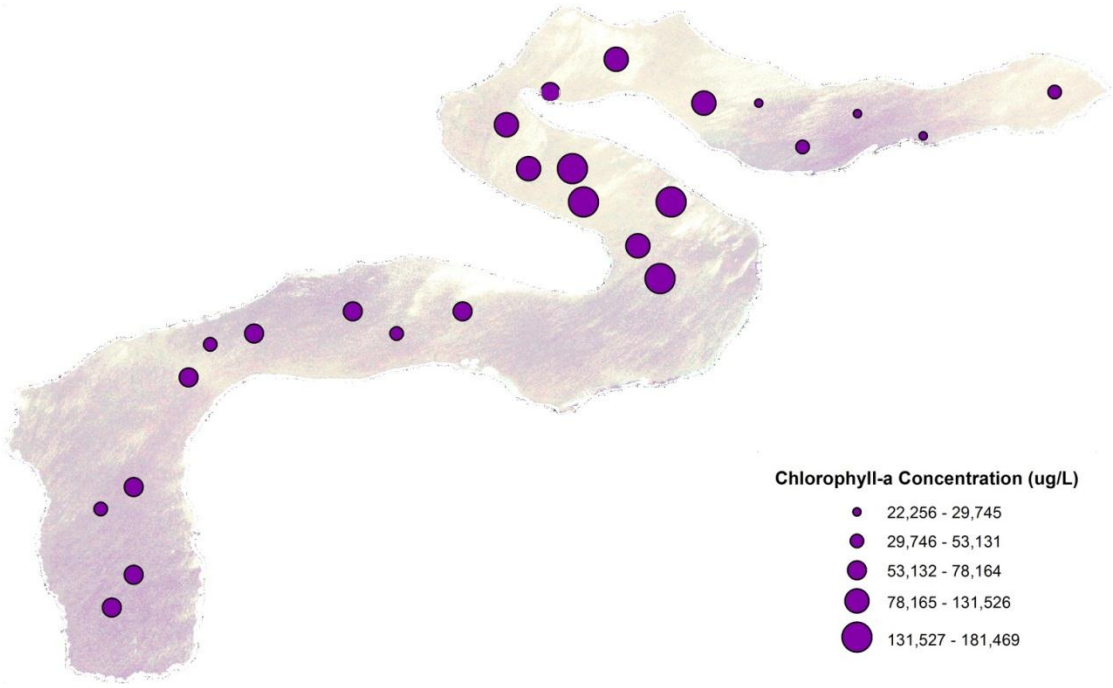

Figure A-6: First Field Study ANN Predicted Chl-a Distribution

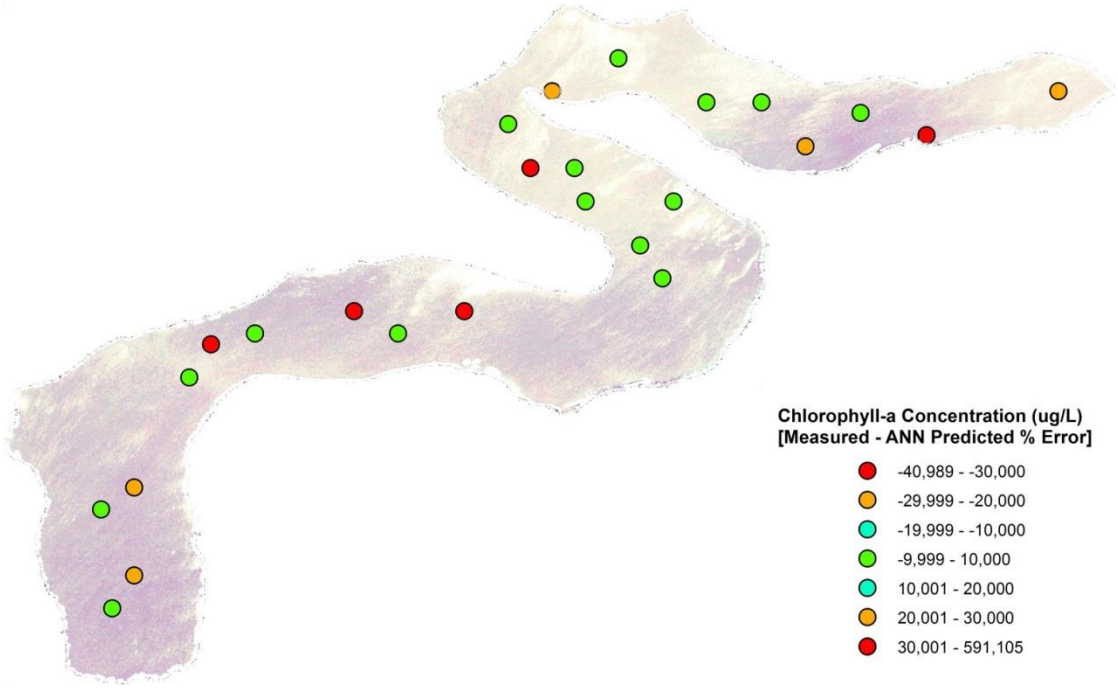

Figure A-7: First Field Study % Deviation of ANN Predicted from Measured Chl-a Distribution

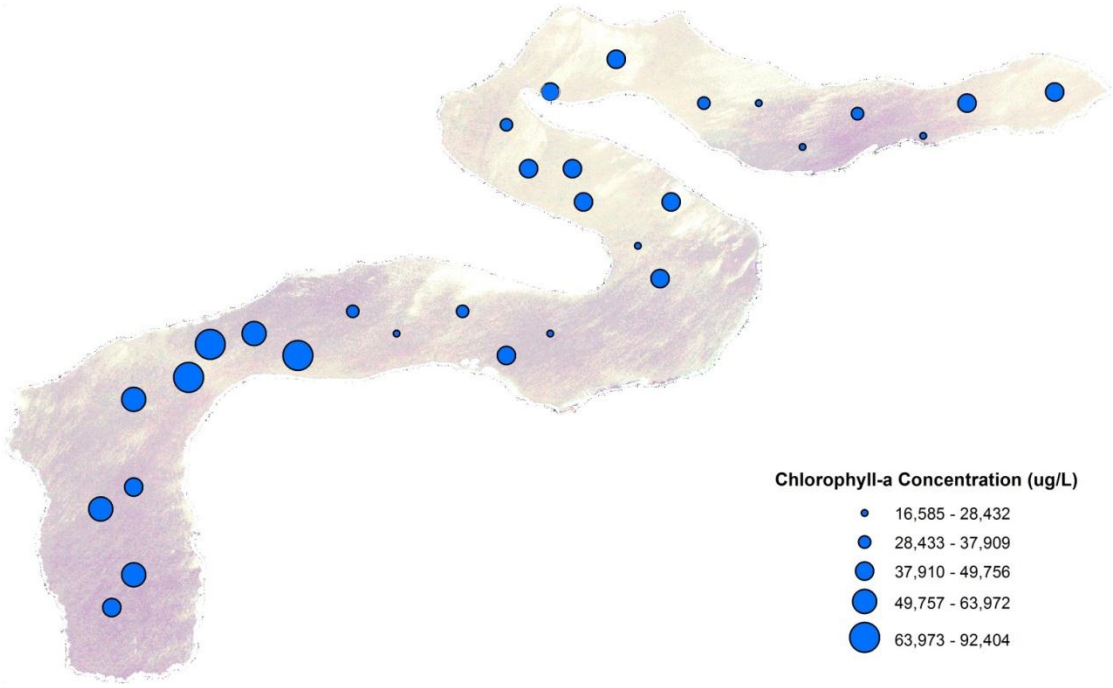

Figure A-8: Second Field Study Measured Chl-a Distribution

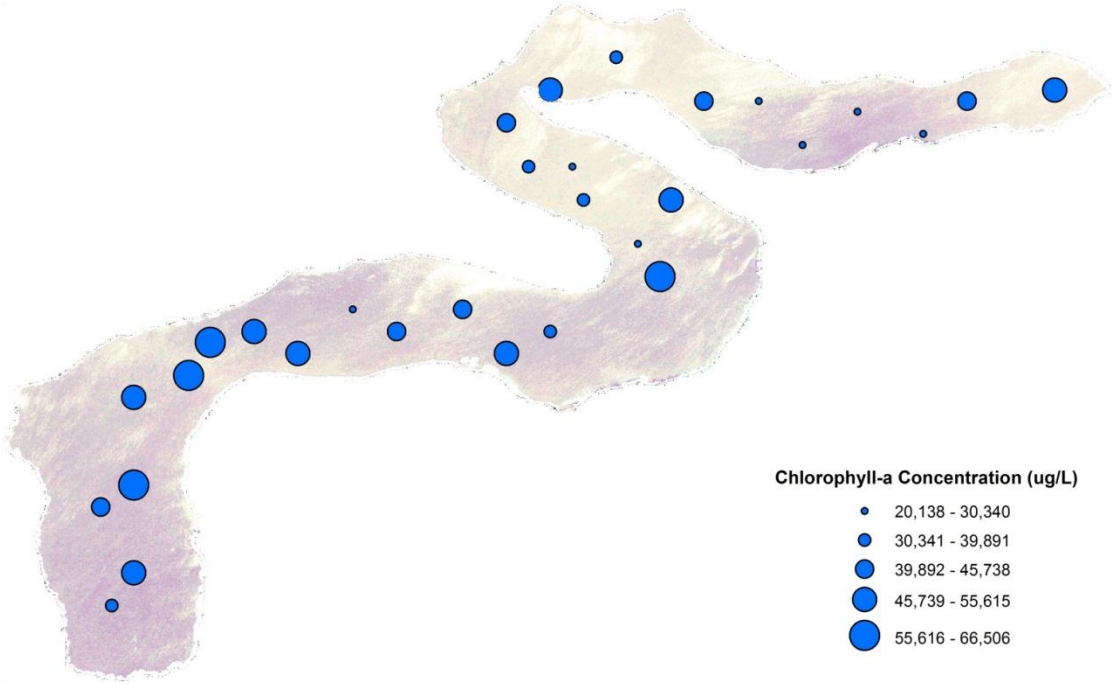

Figure A-9: Second Field Study Linear Regression Predicted Chl-a Distribution

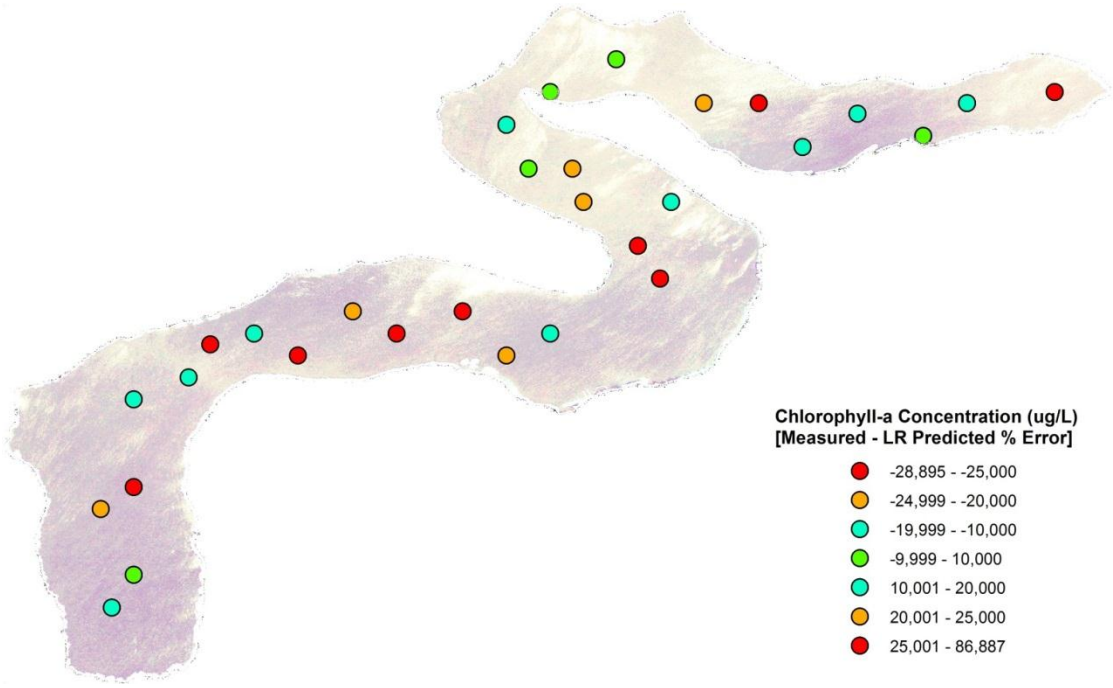

Figure A-10: Second Field Study Deviation of Linear Regression Predicted from Measured Chl-a Distribution

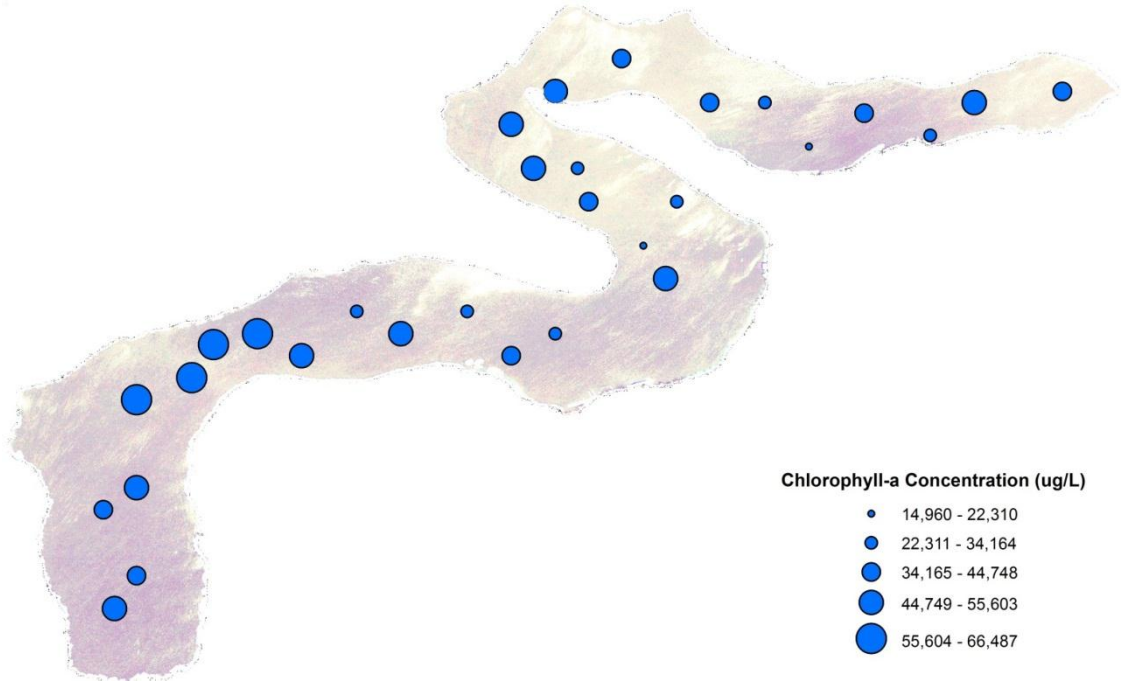

Figure A-11: Second Field Study Non-Linear Regression Predicted Chl-a Distribution

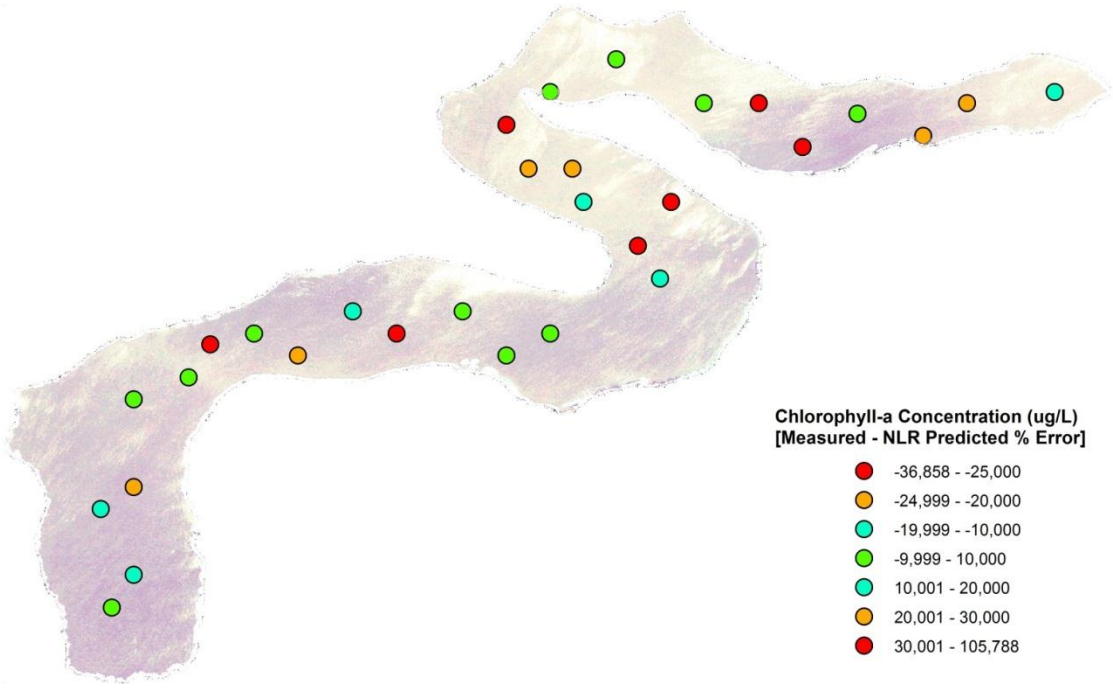

Figure A-12: Second Field Study Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution

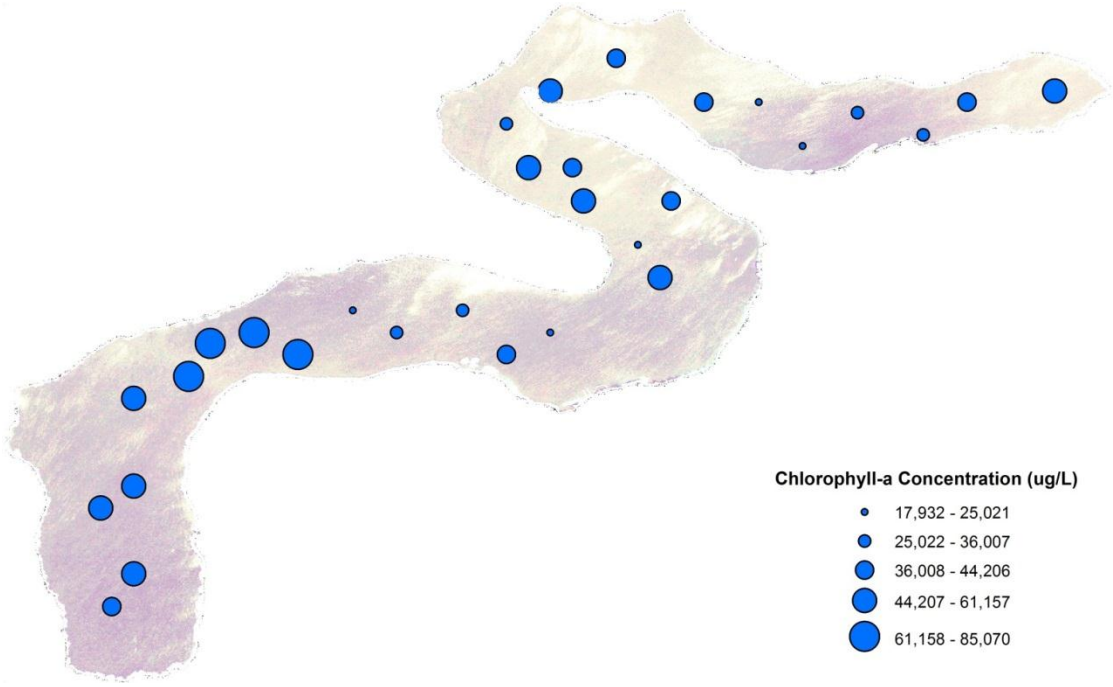

Figure A-13: Second Field Study ANN Predicted Chl-a Distribution

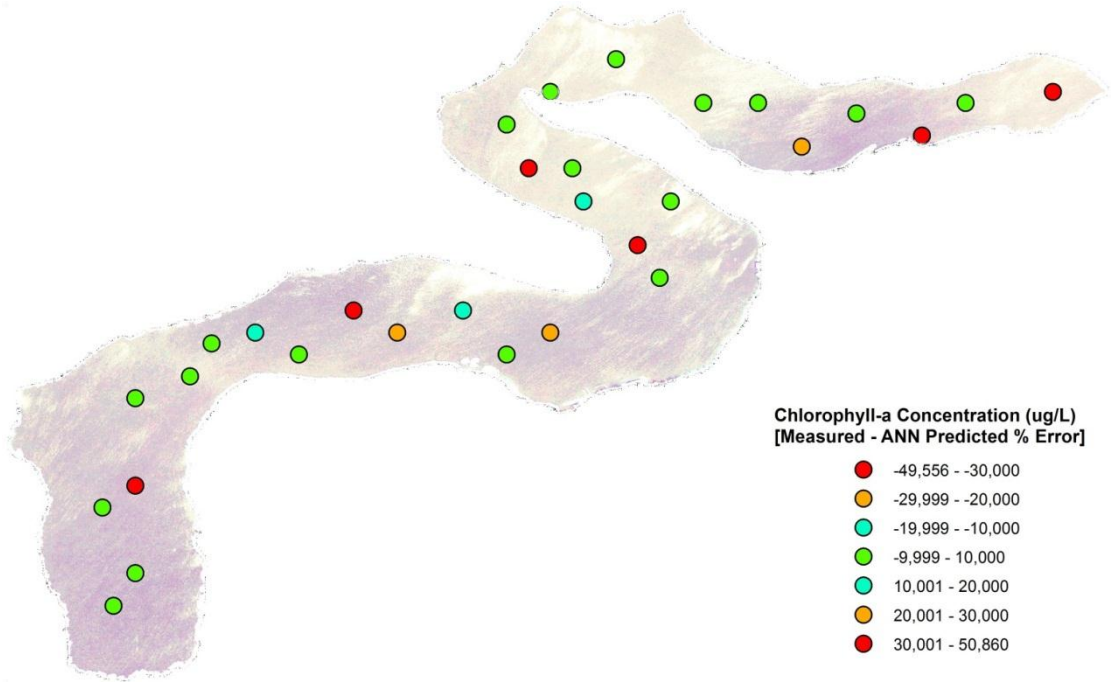

Figure A-14: Second Field Study Deviation of ANN Predicted from Measured Chl-a Distribution

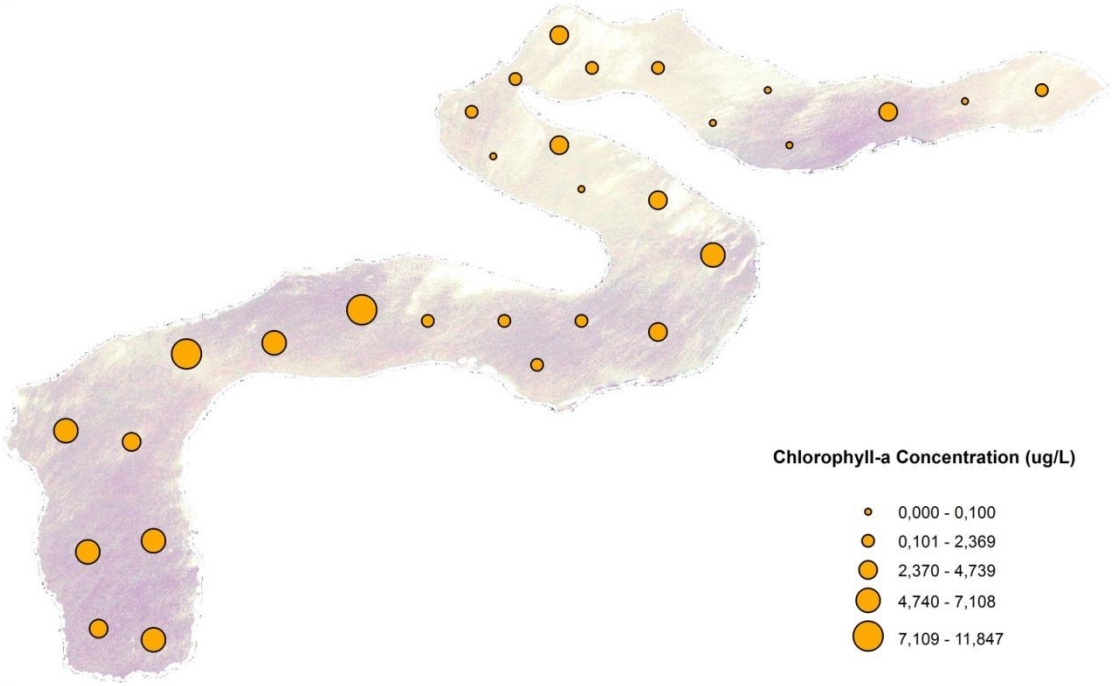

Figure A-15: Third Field Study Measured Chl-a Distribution

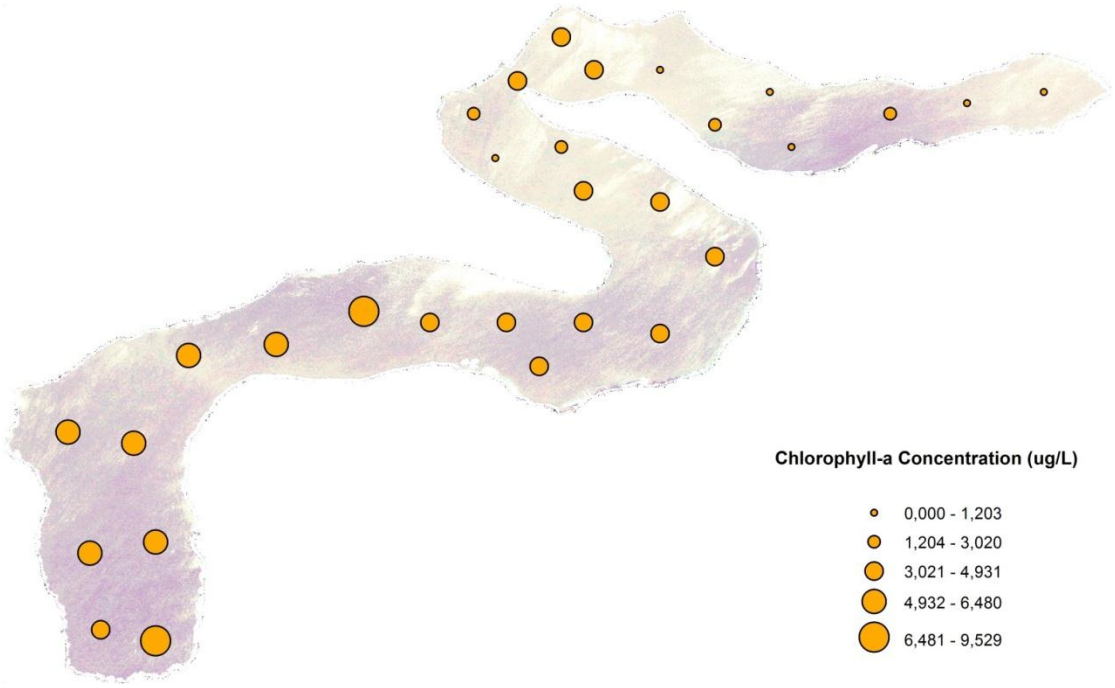

Figure A-16: Third Field Study Linear Regression Predicted Chl-a Distribution

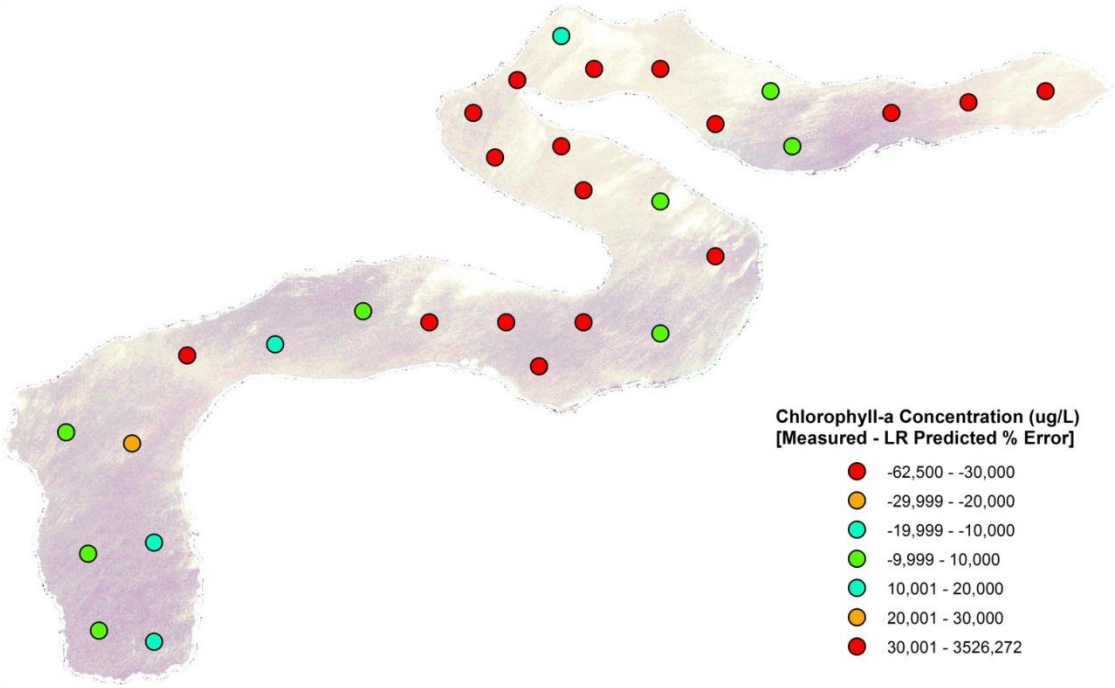

Figure A-17: Third Field Study Deviation of Linear Regression Predicted from Measured Chl-a Distribution

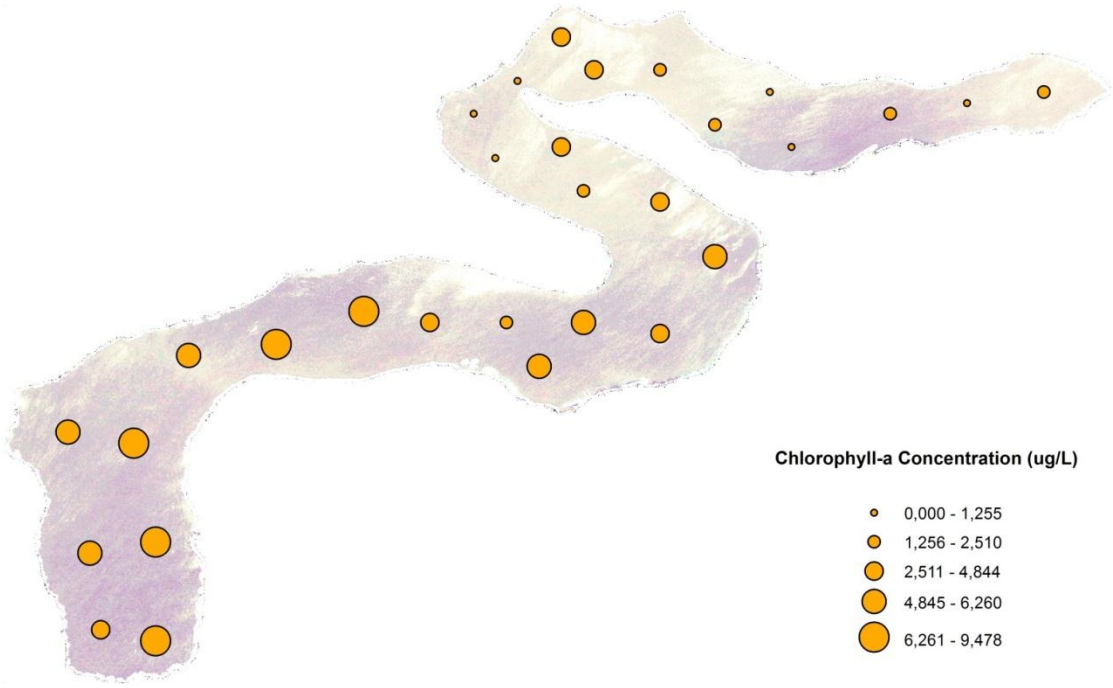

Figure A-18: Third Field Study Non-Linear Regression Predicted Chl-a Distribution

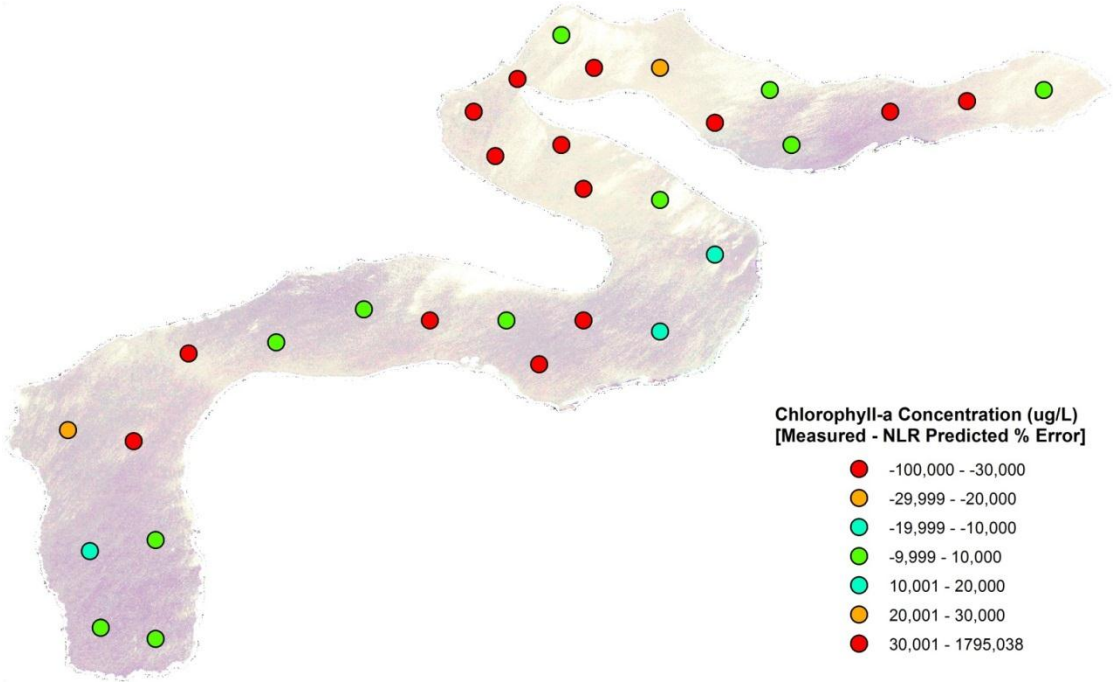

Figure A-19: Third Field Study Deviation of Non-Linear Regression Predicted from Measured Chl-a Distribution

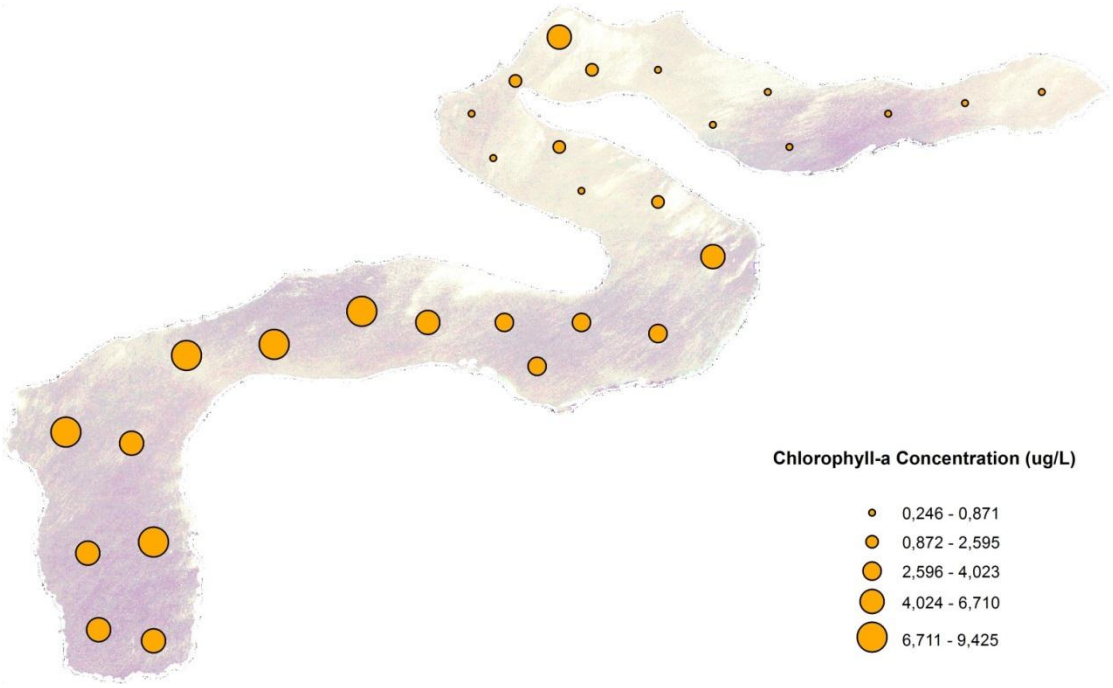

Figure A-20: Third Field Study ANN Predicted Chl-a Distribution

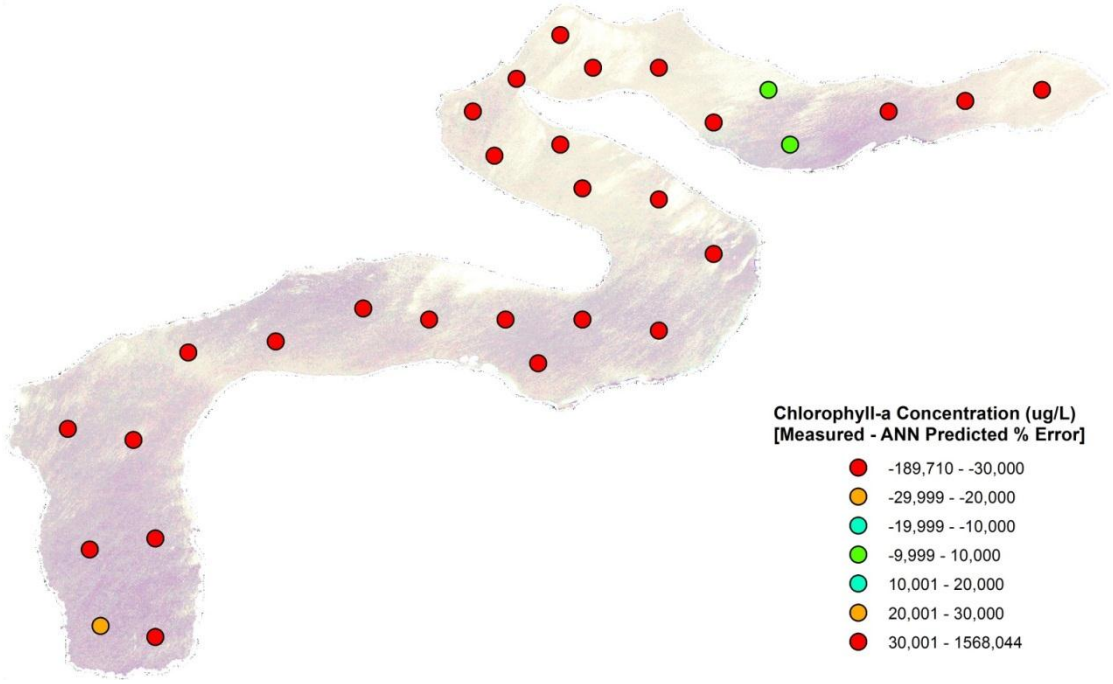

Figure A-21: Third Field Study Deviation of ANN Predicted from Measured Chl-a Distribution

IMAGES of LAKE EYMIR

Figure B-1: Lake Eymir First Study Satellite Imagery

Figure B-2: Lake Eymir Second Study Satellite Imagery

Figure B-3: Lake Eymir Third Study Satellite Imagery

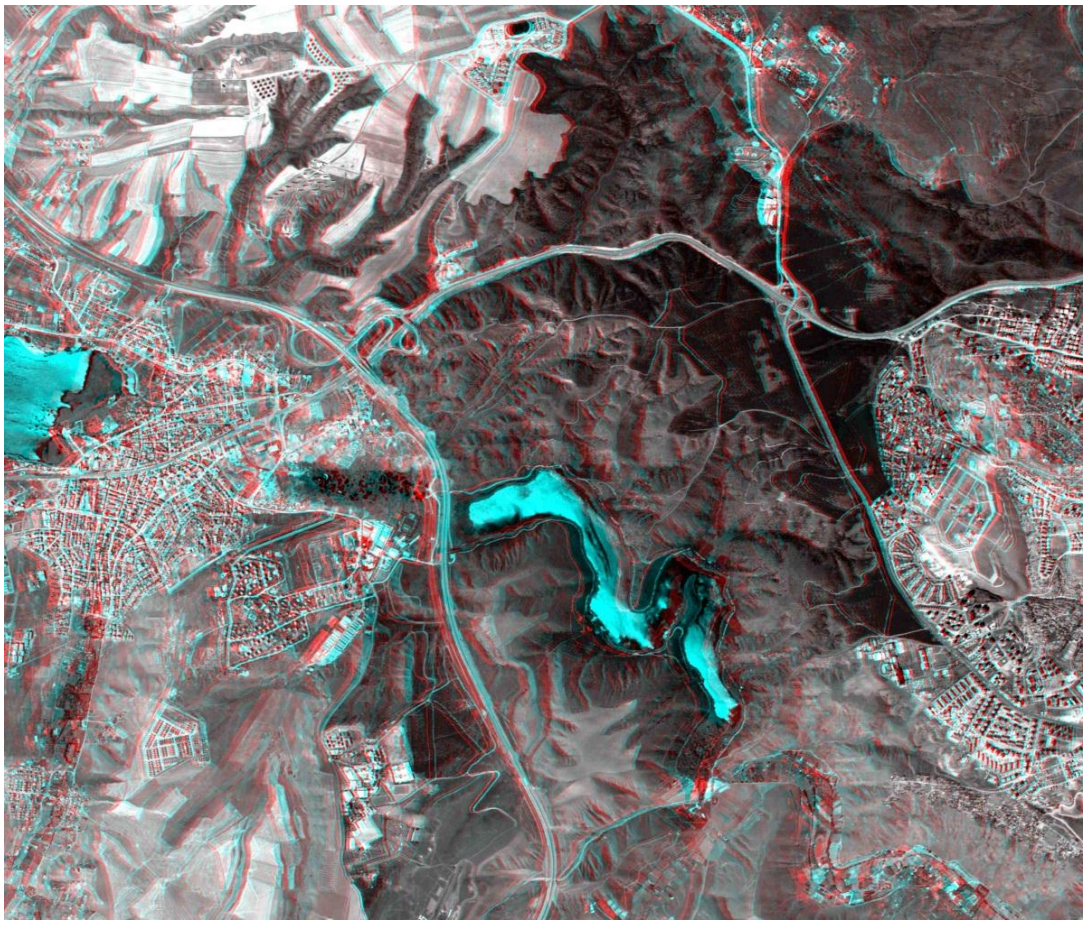

Figure B-4: Stereo Image of Lake Eymir

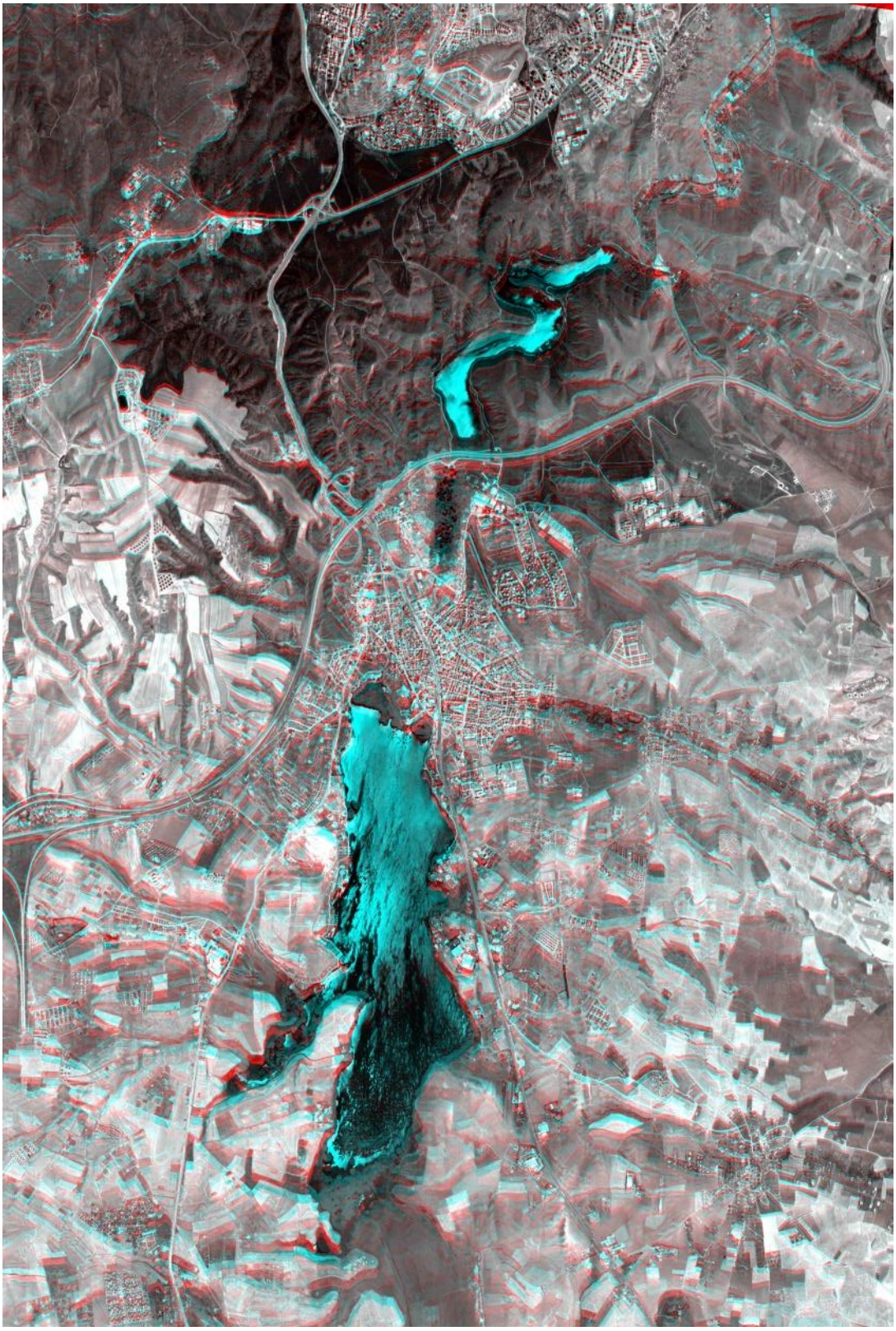

Figure B-5: Stereo Image of Lake Mogan and Lake Eymir

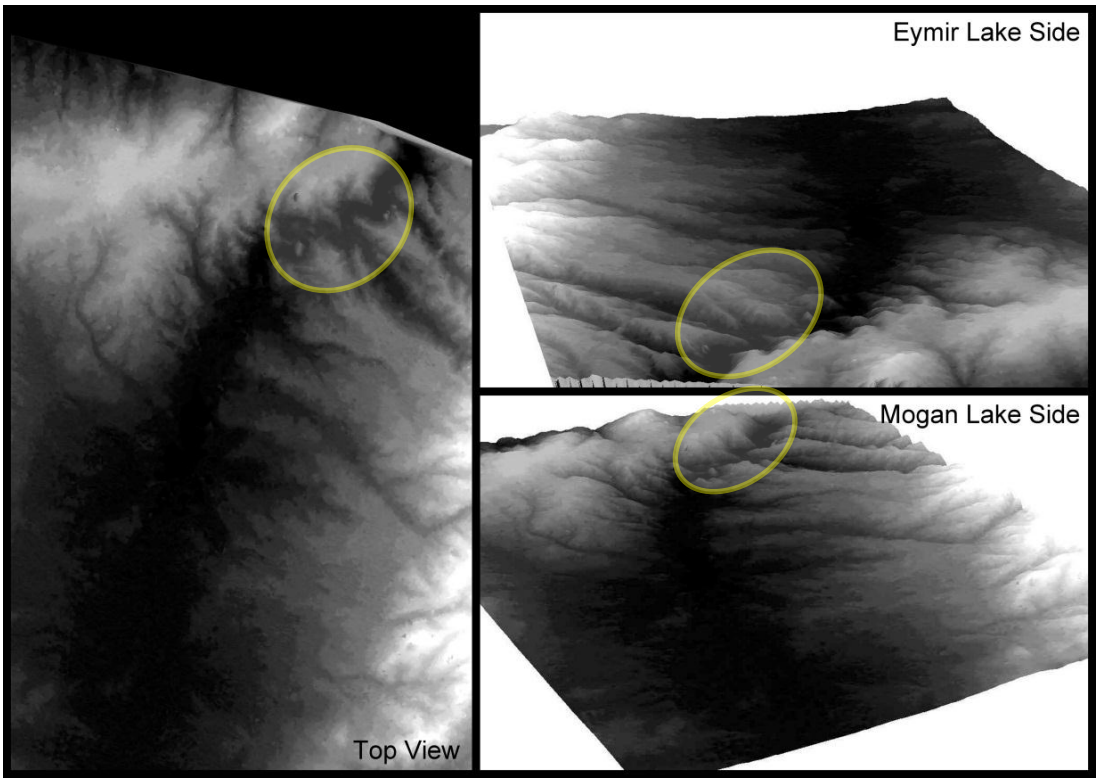

Figure B-6: Digital Elevation Model of Lake Eymir

CORRELATION MATRICES and EIGEN VECTORS

Table C-1: Correlation Matrix of Parameters Measured in First Field Study

Variables	NIR	R	G	B	Avg Temp	Sur Temp	Sur Cond	Avg DO	Sur DO	Avg pH	Sur pH	Avg Tur	Sur Tur	Chl-a (0)	TSS (0)	TSS (0.5)	Secchi D	Depth	Wind	Humidity	PAR 0/0.5
NIR	1	0.780	0.806	0.668	0.546	0.343	0.189	-0.178	0.109	-0.541	-0.224	0.066	0.346	-0.005	-0.424	-0.426	0.131	0.184	-0.174	0.673	-0.157
R	0.780	1	0.940	0.804	0.596	0.414	0.352	-0.100	0.013	-0.493	-0.450	0.275	0.418	-0.160	-0.295	-0.508	0.066	-0.116	0.091	0.871	-0.205
G	0.806	0.940	1	0.877	0.516	0.495	0.170	-0.243	0.173	-0.544	-0.395	0.214	0.360	-0.028	-0.368	-0.504	0.153	0.016	-0.089	0.852	-0.128
B	0.668	0.804	0.877	1	0.583	0.459	0.184	-0.234	0.257	-0.534	-0.513	0.296	0.344	-0.259	-0.187	-0.413	0.085	-0.149	-0.015	0.757	0.092
Avg Temp	0.546	0.596	0.516	0.583	1	0.531	0.244	-0.457	0.364	-0.499	-0.325	-0.157	0.124	0.331	-0.521	-0.375	0.335	0.422	-0.245	0.493	-0.159
Sur Temp	0.343	0.414	0.495	0.459	0.531	1	0.416	-0.283	0.368	-0.599	-0.204	-0.197	0.034	0.422	-0.446	-0.108	0.248	0.469	-0.174	0.273	-0.090
Sur Cond	0.189	0.352	0.170	0.184	0.244	0.416	1	0.445	0.178	-0.083	-0.190	-0.018	-0.006	-0.062	0.105	0.220	-0.255	-0.075	0.334	0.195	0.139
Avg DO	-0.178	-0.100	-0.243	-0.234	-0.457	-0.283	0.445	1	0.080	0.583	0.281	0.094	0.037	-0.101	0.382	0.285	-0.505	-0.195	0.397	-0.213	0.409
Sur DO	0.109	0.013	0.173	0.257	0.364	0.368	0.178	0.080	1	-0.087	-0.032	-0.116	-0.039	0.412	-0.030	-0.041	0.254	0.127	-0.330	0.129	0.210
Avg pH	-0.541	-0.493	-0.544	-0.534	-0.499	-0.599	-0.083	0.583	-0.087	1	0.460	0.028	-0.126	-0.113	0.661	0.404	-0.366	-0.370	0.250	-0.352	0.202
Sur pH	-0.224	-0.450	-0.395	-0.513	-0.325	-0.204	-0.190	0.281	-0.032	0.460	1	-0.399	-0.409	0.448	0.095	0.118	0.053	0.428	-0.333	-0.406	0.076
Avg Tur	0.066	0.275	0.214	0.296	-0.157	-0.197	-0.018	0.094	-0.116	0.028	-0.399	1	0.549	-0.226	0.171	0.199	-0.275	-0.539	0.190	0.214	-0.028
Sur Tur	0.346	0.418	0.360	0.344	0.124	0.034	-0.006	0.037	-0.039	-0.126	-0.409	0.549	1	-0.045	-0.106	-0.082	-0.010	-0.314	0.051	0.358	-0.117
Chl-a (0)	-0.005	-0.160	-0.028	-0.259	0.331	0.422	-0.062	-0.101	0.412	-0.113	0.448	-0.226	-0.045	1	-0.339	-0.004	0.348	0.562	-0.494	-0.191	-0.203
TSS (0)	-0.424	-0.295	-0.368	-0.187	-0.521	-0.446	0.105	0.382	-0.030	0.661	0.095	0.171	-0.106	-0.339	1	0.488	-0.384	-0.555	0.223	-0.259	0.109
TSS (0.5)	-0.426	-0.508	-0.504	-0.413	-0.375	-0.108	0.220	0.285	-0.041	0.404	0.118	0.199	-0.082	-0.004	0.488	1	-0.362	-0.330	0.224	-0.592	-0.002
Secchi D	0.131	0.066	0.153	0.085	0.335	0.248	-0.255	-0.505	0.254	-0.366	0.053	-0.275	-0.010	0.348	-0.384	-0.362	1	0.310	-0.433	0.260	-0.242
Depth	0.184	-0.116	0.016	-0.149	0.422	0.469	-0.075	-0.195	0.127	-0.370	0.428	-0.539	-0.314	0.562	-0.555	-0.330	0.310	1	-0.425	-0.214	0.031
Wind	-0.174	0.091	-0.089	-0.015	-0.245	-0.174	0.334	0.397	-0.330	0.250	-0.333	0.190	0.051	-0.494	0.223	0.224	-0.433	-0.425	1	-0.020	0.181
Humidity	0.673	0.871	0.852	0.757	0.493	0.273	0.195	-0.213	0.129	-0.352	-0.406	0.214	0.358	-0.191	-0.259	-0.592	0.260	-0.214	-0.020	1	-0.093
PAR 0/0.5	-0.157	-0.205	-0.128	0.092	-0.159	-0.090	0.139	0.409	0.210	0.202	0.076	-0.028	-0.117	-0.203	0.109	-0.002	-0.242	0.031	0.181	-0.093	1

Table C-2: Eigen Values of Parameters Measured in First Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
Eigenvalue	6.904	4.038	2.125	1.538	1.401	1.137	0.954	0.688	0.482	0.427	0.408	0.247	0.223	0.160	0.120	0.071	0.042	0.018	0.012	0.005	0.001
Variability (%)	32.877	19.230	10.117	7.325	6.673	5.413	4.541	3.278	2.296	2.032	1.944	1.176	1.060	0.761	0.569	0.337	0.200	0.088	0.056	0.022	0.005
Cumulative %	32.877	52.107	62.224	69.549	76.221	81.635	86.176	89.454	91.750	93.783	95.726	96.902	97.961	98.723	99.292	99.629	99.829	99.916	99.973	99.995	100.000

Table C-3: Eigen Vectors of Parameters Measured in First Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
NIR	0.310	-0.057	-0.007	-0.280	0.061	-0.216	-0.012	-0.119	0.367	-0.030	0.027	-0.249	-0.082	0.513	0.222	-0.133	0.112	-0.193	0.404	-0.073	0.063
R	0.319	-0.209	0.043	-0.101	0.003	-0.196	-0.166	-0.001	-0.101	0.069	-0.057	0.001	0.080	-0.232	-0.198	0.128	0.201	-0.224	-0.121	-0.652	-0.334
G	0.347	-0.125	0.024	-0.118	0.093	-0.051	-0.085	-0.149	-0.194	-0.092	-0.089	-0.026	0.075	-0.154	-0.170	0.036	0.477	-0.170	-0.067	0.658	-0.041
B	0.308	-0.193	0.067	-0.032	0.074	0.246	0.003	-0.254	-0.158	0.041	0.085	-0.237	0.309	0.097	-0.067	-0.127	-0.052	0.710	0.028	-0.093	-0.006
Avg Temp	0.327	0.125	0.184	0.187	-0.076	0.044	-0.035	-0.108	-0.206	-0.206	0.127	0.048	-0.056	-0.055	0.193	-0.135	0.041	-0.179	-0.242	-0.214	0.697
Sur Temp	0.254	0.157	0.333	0.292	-0.117	-0.029	-0.005	-0.080	-0.248	-0.055	0.162	0.018	-0.067	-0.059	0.315	-0.236	-0.317	-0.149	0.169	0.138	-0.519
Sur Cond	0.061	-0.145	0.536	0.089	-0.150	-0.174	-0.214	0.182	0.283	0.402	0.042	0.131	-0.263	0.151	0.036	0.083	0.133	0.257	-0.301	0.119	0.017
Avg DO	-0.173	-0.174	0.399	-0.255	0.183	-0.259	0.099	0.184	0.101	0.031	-0.047	-0.157	0.280	-0.510	0.067	-0.275	-0.136	-0.018	0.228	0.039	0.208
Sur DO	0.092	0.122	0.329	0.107	0.470	0.398	-0.014	0.141	0.307	-0.287	-0.261	-0.037	0.243	0.124	0.043	0.257	-0.072	-0.141	-0.176	-0.021	-0.110
Avg pH	-0.295	-0.102	0.063	-0.180	0.270	-0.063	-0.205	0.078	-0.311	-0.302	0.100	0.281	-0.077	0.093	0.491	0.041	0.363	0.221	0.067	-0.110	-0.097
Sur pH	-0.171	0.273	0.074	-0.318	0.225	-0.279	-0.146	-0.192	-0.338	0.261	0.023	-0.142	0.229	0.321	-0.038	0.006	-0.263	-0.151	-0.384	0.052	0.000
Avg Tur	0.023	-0.336	-0.150	0.262	0.287	-0.114	0.286	-0.177	-0.140	0.275	-0.401	-0.238	-0.265	-0.084	0.381	0.186	-0.082	-0.014	-0.090	0.006	0.048
Sur Tur	0.128	-0.225	-0.138	0.175	0.317	-0.298	0.385	0.240	0.065	-0.130	0.617	0.070	0.076	0.098	-0.105	0.092	-0.078	0.000	-0.186	0.066	-0.035
Chl-a (0)	0.042	0.357	0.133	0.186	0.336	-0.305	0.076	0.080	-0.158	-0.155	-0.249	-0.021	-0.378	0.070	-0.451	-0.157	0.072	0.241	0.204	-0.088	0.029
TSS (0)	-0.223	-0.210	0.082	0.043	0.193	0.179	-0.436	-0.256	0.040	-0.119	0.374	-0.394	-0.395	-0.144	-0.189	0.117	-0.134	-0.114	0.063	0.017	0.031
TSS (0.5)	-0.244	-0.066	0.182	0.503	0.004	-0.047	-0.048	-0.133	-0.141	0.253	0.080	0.039	0.409	0.205	-0.121	0.193	0.250	-0.166	0.406	-0.050	0.127
Secchi D	0.144	0.258	-0.212	0.052	0.147	0.223	-0.186	0.604	-0.159	0.366	0.145	-0.373	-0.003	-0.065	0.149	-0.078	0.191	-0.022	0.069	0.001	0.020
Depth	0.085	0.413	0.133	-0.172	-0.168	-0.152	0.202	-0.105	0.011	-0.044	0.163	-0.199	-0.021	-0.225	0.141	0.690	0.049	0.180	0.147	0.000	0.026
Wind	-0.087	-0.308	0.151	-0.012	-0.401	-0.069	0.053	0.401	-0.322	-0.365	-0.187	-0.382	0.010	0.285	-0.099	0.149	-0.103	-0.035	-0.038	0.037	0.017
Humidity	0.297	-0.183	-0.049	-0.181	0.133	0.060	-0.238	0.187	-0.204	0.133	-0.062	0.432	-0.066	0.037	-0.089	0.314	-0.450	-0.011	0.345	0.090	0.198
PAR 0/0.5	-0.075	-0.065	0.311	-0.323	0.036	0.448	0.530	-0.030	-0.228	0.230	0.140	0.083	-0.258	0.109	-0.149	-0.053	0.147	-0.185	0.068	-0.094	-0.025

Table C-4: Correlation Matrix of Parameters Measured in Second Field Study

Variables	NIRr	Rr	Gr	Br	Avg Temp	Sur Temp	Sur Cond	Avg DO	Sur DO	Avg pH	Sur pH	Avg Tur	Sur Tur	Chl-a (0)	TSS (0)	TSS (0.5)	Secchi D	Depth	Wind	Humidity	PAR 0/0.5
NIRr	1	0.812	0.829	0.781	-0.316	-0.230	-0.166	-0.064	-0.025	-0.209	-0.246	-0.131	0.085	-0.117	0.142	-0.112	-0.265	0.215	-0.479	-0.065	0.547
Rr	0.812	1	0.886	0.815	-0.262	-0.295	-0.178	-0.128	-0.027	-0.200	-0.198	0.063	0.091	-0.063	0.239	-0.087	-0.033	0.133	-0.434	0.049	0.412
Gr	0.829	0.886	1	0.846	-0.282	-0.259	-0.232	-0.129	0.088	-0.293	-0.254	-0.048	-0.021	0.117	0.181	-0.140	0.032	0.146	-0.361	0.033	0.344
Br	0.781	0.815	0.846	1	-0.277	-0.246	-0.096	-0.117	0.031	-0.220	-0.202	-0.110	-0.006	0.000	-0.047	-0.255	-0.159	0.188	-0.425	0.120	0.502
Avg Temp	-0.316	-0.262	-0.282	-0.277	1	0.357	0.194	0.322	-0.370	0.404	0.199	0.128	-0.159	-0.008	-0.053	0.136	0.138	-0.518	0.112	0.071	0.225
Sur Temp	-0.230	-0.295	-0.259	-0.246	0.357	1	-0.032	0.208	-0.152	0.171	0.104	-0.068	-0.250	-0.124	0.095	0.075	0.106	-0.021	0.255	-0.066	-0.002
Sur Cond	-0.166	-0.178	-0.232	-0.096	0.194	-0.032	1	-0.236	-0.135	0.403	0.454	0.081	0.167	-0.161	-0.028	-0.025	-0.035	0.075	0.111	0.204	-0.013
Avg DO	-0.064	-0.128	-0.129	-0.117	0.322	0.208	-0.236	1	0.401	0.386	0.216	-0.137	-0.157	-0.245	-0.453	0.381	-0.105	-0.072	0.096	-0.278	0.135
Sur DO	-0.025	-0.027	0.088	0.031	-0.370	-0.152	-0.135	0.401	1	0.134	0.393	-0.474	-0.236	0.084	-0.384	-0.041	-0.134	0.485	0.301	-0.314	-0.431
Avg pH	-0.209	-0.200	-0.293	-0.220	0.404	0.171	0.403	0.386	0.134	1	0.490	0.161	0.211	-0.068	-0.188	0.149	-0.131	0.067	0.214	-0.352	0.076
Sur pH	-0.246	-0.198	-0.254	-0.202	0.199	0.104	0.454	0.216	0.393	0.490	1	0.049	0.200	0.014	-0.138	-0.052	-0.077	0.255	0.334	-0.358	-0.162
Avg Tur	-0.131	0.063	-0.048	-0.110	0.128	-0.068	0.081	-0.137	-0.474	0.161	0.049	1	0.613	0.015	0.357	-0.050	0.296	-0.186	-0.240	-0.046	0.174
Sur Tur	0.085	0.091	-0.021	-0.006	-0.159	-0.250	0.167	-0.157	-0.236	0.211	0.200	0.613	1	-0.076	0.116	-0.100	0.243	0.026	-0.259	-0.207	0.315
Chl-a (0)	-0.117	-0.063	0.117	0.000	-0.008	-0.124	-0.161	-0.245	0.084	-0.068	0.014	0.015	-0.076	1	-0.161	-0.233	0.093	-0.096	-0.203	-0.414	-0.195
TSS (0)	0.142	0.239	0.181	-0.047	-0.053	0.095	-0.028	-0.453	-0.384	-0.188	-0.138	0.357	0.116	-0.161	1	-0.142	0.253	-0.083	0.013	0.300	0.089
TSS (0.5)	-0.112	-0.087	-0.140	-0.255	0.136	0.075	-0.025	0.381	-0.041	0.149	-0.052	-0.050	-0.100	-0.233	-0.142	1	-0.098	-0.175	0.036	-0.150	-0.012
Secchi D	-0.265	-0.033	0.032	-0.159	0.138	0.106	-0.035	-0.105	-0.134	-0.131	-0.077	0.296	0.243	0.093	0.253	-0.098	1	-0.120	0.217	0.086	-0.192
Depth	0.215	0.133	0.146	0.188	-0.518	-0.021	0.075	-0.072	0.485	0.067	0.255	-0.186	0.026	-0.096	-0.083	-0.175	-0.120	1	0.007	-0.197	-0.400
Wind	-0.479	-0.434	-0.361	-0.425	0.112	0.255	0.111	0.096	0.301	0.214	0.334	-0.240	-0.259	-0.203	0.013	0.036	0.217	0.007	1	0.140	-0.383
Humidity	-0.065	0.049	0.033	0.120	0.071	-0.066	0.204	-0.278	-0.314	-0.352	-0.358	-0.046	-0.207	-0.414	0.300	-0.150	0.086	-0.197	0.140	1	0.090
PAR 0/0.5	0.547	0.412	0.344	0.502	0.225	-0.002	-0.013	0.135	-0.431	0.076	-0.162	0.174	0.315	-0.195	0.089	-0.012	-0.192	-0.400	-0.383	0.090	1

Table C-5: Eigen Values of Parameters Measured in Second Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
Eigenvalue	4.756	3.031	2.480	2.199	1.709	1.304	1.292	0.976	0.804	0.669	0.515	0.343	0.248	0.222	0.179	0.115	0.080	0.035	0.020	0.015	0.007
Variability (%)	22.650	14.435	11.810	10.473	8.139	6.209	6.151	4.650	3.830	3.184	2.452	1.634	1.181	1.059	0.852	0.547	0.380	0.167	0.095	0.071	0.034
Cumulative %	22.650	37.085	48.895	59.368	67.506	73.715	79.866	84.515	88.345	91.529	93.981	95.616	96.797	97.856	98.708	99.254	99.634	99.801	99.896	99.966	100.000

Table C-6: Eigen Vectors of Parameters Measured in Second Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
NIRr	0.398	0.104	0.166	0.082	-0.096	0.050	-0.013	0.130	0.039	0.034	-0.185	-0.246	-0.183	0.186	0.263	0.388	-0.055	-0.221	-0.376	-0.318	-0.296
Rr	0.397	0.044	0.142	0.010	-0.084	0.159	-0.125	-0.103	0.218	0.009	0.123	0.059	-0.198	-0.032	-0.307	-0.180	0.717	0.035	-0.068	0.088	-0.022
Gr	0.398	0.097	0.062	0.030	-0.023	0.145	-0.261	-0.133	0.200	-0.059	0.014	0.086	0.091	0.119	0.003	0.412	-0.234	-0.048	0.450	0.157	0.439
Br	0.390	0.112	0.126	0.062	-0.121	-0.081	-0.182	-0.085	-0.097	-0.185	0.041	0.325	0.022	-0.071	-0.103	-0.375	-0.432	0.354	0.003	-0.172	-0.325
Avg Temp	-0.199	-0.275	0.213	0.311	0.014	-0.102	-0.350	-0.106	0.065	-0.028	0.177	-0.220	-0.286	0.069	-0.100	0.171	-0.176	0.258	-0.274	0.457	-0.079
Sur Temp	-0.173	-0.074	0.001	0.198	-0.108	0.293	-0.282	0.623	-0.156	-0.346	-0.055	0.144	0.216	0.050	-0.310	0.141	0.087	-0.084	0.029	-0.076	-0.078
Sur Cond	-0.123	-0.082	0.189	-0.227	-0.398	-0.393	-0.001	-0.078	0.229	-0.403	-0.056	-0.196	0.258	0.441	0.080	-0.077	0.147	0.065	0.119	-0.036	-0.039
Avg DO	-0.140	0.191	0.257	0.386	0.084	0.294	0.086	-0.183	-0.229	-0.008	0.299	-0.153	0.221	0.117	0.299	0.042	0.207	0.372	0.130	-0.273	0.046
Sur DO	-0.067	0.504	-0.032	-0.042	-0.007	0.129	-0.072	-0.254	-0.072	0.167	0.065	-0.163	0.421	0.182	-0.324	-0.014	-0.106	-0.271	-0.147	0.295	-0.282
Avg pH	-0.231	0.081	0.488	-0.080	-0.119	-0.001	-0.127	0.034	0.109	0.107	0.078	0.116	-0.126	-0.364	0.226	0.035	0.027	-0.304	0.415	0.081	-0.387
Sur pH	-0.229	0.187	0.377	-0.242	-0.169	-0.009	-0.223	-0.012	0.091	0.158	0.025	-0.007	0.015	-0.210	-0.208	-0.023	-0.122	0.022	-0.356	-0.371	0.481
Avg Tur	0.011	-0.345	0.212	-0.290	0.180	0.200	0.114	-0.019	-0.013	0.065	0.455	0.436	0.060	0.415	0.070	0.016	-0.080	-0.199	-0.192	0.001	-0.033
Sur Tur	0.054	-0.180	0.308	-0.386	0.132	0.126	0.280	-0.125	-0.269	-0.045	-0.330	-0.067	0.071	-0.124	-0.314	0.363	0.036	0.347	0.074	0.070	-0.154
Chl-a (0)	0.010	0.069	-0.072	-0.146	0.539	-0.292	-0.411	0.048	0.197	-0.004	-0.061	0.163	0.362	-0.133	0.231	0.171	0.219	0.181	-0.152	-0.032	-0.127
TSS (0)	0.109	-0.298	-0.115	-0.189	-0.203	0.321	-0.115	0.254	0.290	0.464	0.052	-0.328	0.278	-0.040	0.101	-0.165	-0.104	0.263	0.082	0.011	-0.125
TSS (0.5)	-0.105	0.006	0.099	0.297	0.060	0.211	0.427	-0.065	0.667	-0.199	-0.164	0.181	0.174	-0.157	-0.047	0.025	-0.126	0.033	-0.184	0.040	-0.040
Secchi D	-0.049	-0.209	-0.126	-0.196	0.109	0.451	-0.289	-0.395	-0.038	-0.459	-0.116	-0.232	-0.029	-0.176	0.179	-0.196	-0.046	-0.220	-0.080	-0.081	-0.009
Depth	0.059	0.382	-0.043	-0.326	-0.186	0.155	0.141	0.284	-0.065	-0.272	0.189	0.034	-0.084	-0.152	0.362	0.041	0.005	0.224	-0.230	0.437	0.104
Wind	-0.271	0.077	-0.162	-0.008	-0.302	0.226	-0.204	-0.218	-0.011	0.237	-0.453	0.453	-0.147	0.273	0.187	0.080	0.116	0.192	-0.053	0.043	-0.060
Humidity	0.066	-0.243	-0.262	0.089	-0.472	-0.138	0.031	-0.267	-0.124	-0.007	0.291	0.160	0.284	-0.393	0.007	0.371	0.096	-0.059	-0.158	-0.066	-0.072
PAR 0/0.5	0.219	-0.213	0.352	0.235	-0.055	-0.067	0.014	0.040	-0.284	0.107	-0.350	0.064	0.352	-0.077	0.241	-0.274	0.064	-0.195	-0.175	0.322	0.245

Table C-7: Correlation Matrix of Parameters Measured in Third Field Study

Variables	NIR-1	R	G	B	Avg Temp	Sur Temp	Sur Cond	Avg DO	Sur DO	Avg pH	Sur pH	Avg Tur	Sur Tur	Chl-a (0)	TSS (0)	TSS (0.5)	Secchi D	Depth	Wind	Humidity	PAR 0/0.5
NIR-1	1	0.587	0.987	0.981	0.387	-0.221	0.083	0.306	0.016	-0.113	0.041	-0.219	0.502	0.178	-0.079	-0.020	-0.167	-0.373	-0.156	0.233	-0.019
R	0.587	1	0.585	0.595	0.594	-0.399	0.092	0.549	0.033	-0.085	0.053	-0.154	0.417	0.163	0.029	0.032	-0.304	-0.554	-0.417	0.547	0.046
G	0.987	0.585	1	0.991	0.406	-0.219	0.099	0.315	0.061	-0.088	0.053	-0.176	0.496	0.174	-0.070	-0.035	-0.190	-0.369	-0.124	0.242	-0.009
B	0.981	0.595	0.991	1	0.425	-0.221	0.077	0.316	0.046	-0.062	0.031	-0.214	0.517	0.197	-0.097	-0.026	-0.167	-0.356	-0.178	0.240	-0.001
Avg Temp	0.387	0.594	0.406	0.425	1	-0.048	0.159	0.674	-0.115	-0.031	0.120	-0.175	0.219	0.031	0.000	-0.187	0.016	-0.774	-0.144	0.247	-0.057
Sur Temp	-0.221	-0.399	-0.219	-0.221	-0.048	1	0.087	-0.504	-0.438	-0.099	0.080	0.194	-0.588	-0.587	-0.435	-0.202	0.597	0.217	0.163	-0.496	-0.456
Sur Cond	0.083	0.092	0.099	0.077	0.159	0.087	1	0.060	0.270	0.053	0.995	0.080	-0.074	-0.327	-0.027	-0.165	0.064	-0.107	0.074	-0.064	-0.819
Avg DO	0.306	0.549	0.315	0.316	0.674	-0.504	0.060	1	0.440	-0.071	0.038	-0.171	0.542	0.451	0.262	0.039	-0.504	-0.555	-0.222	0.689	0.224
Sur DO	0.016	0.033	0.061	0.046	-0.115	-0.438	0.270	0.440	1	0.220	0.266	0.157	0.358	0.418	0.431	0.147	-0.527	0.167	0.106	0.503	-0.018
Avg pH	-0.113	-0.085	-0.088	-0.062	-0.031	-0.099	0.053	-0.071	0.220	1	0.060	-0.189	0.022	0.127	0.419	-0.259	0.078	0.137	0.004	-0.028	-0.031
Sur pH	0.041	0.053	0.053	0.031	0.120	0.080	0.995	0.038	0.266	0.060	1	0.051	-0.096	-0.321	-0.002	-0.152	0.078	-0.068	0.080	-0.080	-0.323
Avg Tur	-0.219	-0.154	-0.176	-0.214	-0.175	0.194	0.080	-0.171	0.157	-0.189	0.051	1	-0.216	-0.263	-0.312	-0.037	-0.060	0.057	0.130	-0.107	-0.075
Sur Tur	0.502	0.417	0.496	0.517	0.219	-0.588	-0.074	0.542	0.358	0.022	-0.096	-0.216	1	0.559	0.296	0.177	-0.549	-0.302	-0.347	0.543	0.384
Chl-a (0)	0.178	0.163	0.174	0.197	0.031	-0.587	-0.327	0.451	0.418	0.127	-0.321	-0.263	0.559	1	0.466	0.363	-0.585	-0.140	-0.203	0.506	0.511
TSS (0)	-0.079	0.029	-0.070	-0.097	0.000	-0.435	-0.027	0.262	0.431	0.419	-0.002	-0.312	0.296	0.466	1	0.154	-0.377	-0.092	0.274	0.355	0.145
TSS (0.5)	-0.020	0.032	-0.035	-0.026	-0.187	-0.202	-0.165	0.039	0.147	-0.259	-0.152	-0.037	0.177	0.363	0.154	1	-0.192	0.100	-0.197	0.235	0.174
Secchi D	-0.167	-0.304	-0.190	-0.167	0.016	0.597	0.064	-0.504	-0.527	0.078	0.078	-0.060	-0.549	-0.585	-0.377	-0.192	1	0.126	0.082	-0.594	-0.466
Depth	-0.373	-0.554	-0.369	-0.356	-0.774	0.217	-0.107	-0.555	0.167	0.137	-0.068	0.057	-0.302	-0.140	-0.092	0.100	0.126	1	-0.039	-0.362	-0.106
Wind	-0.156	-0.417	-0.124	-0.178	-0.144	0.163	0.074	-0.222	0.106	0.004	0.080	0.130	-0.347	-0.203	0.274	-0.197	0.082	-0.039	1	-0.255	-0.086
Humidity	0.233	0.547	0.242	0.240	0.247	-0.496	-0.064	0.689	0.503	-0.028	-0.080	-0.107	0.543	0.506	0.355	0.235	-0.594	-0.362	-0.255	1	0.451
PAR 0/0.5	-0.019	0.046	-0.009	-0.001	-0.057	-0.456	-0.819	0.224	-0.018	-0.031	-0.323	-0.075	0.384	0.511	0.145	0.174	-0.466	-0.106	-0.086	0.451	1

Table C-8: Eigen Values of Parameters Measured in Third Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
Eigenvalue	6.397	3.728	2.666	1.625	1.572	1.230	0.995	0.693	0.560	0.426	0.323	0.266	0.185	0.118	0.085	0.059	0.046	0.012	0.009	0.002	0.001
Variability (%)	30.461	17.755	12.695	7.736	7.487	5.859	4.740	3.301	2.665	2.030	1.538	1.266	0.882	0.561	0.403	0.282	0.218	0.059	0.044	0.012	0.006
Cumulative %	30.461	48.216	60.911	68.647	76.133	81.992	86.733	90.034	92.699	94.729	96.266	97.532	98.414	98.976	99.379	99.661	99.879	99.938	99.982	99.994	100.000

Table C-9: Eigen Vectors of Parameters Measured in Third Field Study

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21
NIR-1	0.245	-0.297	-0.160	-0.281	0.177	0.148	0.047	0.023	-0.004	0.002	0.019	0.044	0.012	-0.017	0.268	-0.101	0.147	-0.003	0.750	-0.148	-0.040
R	0.274	-0.216	-0.072	0.088	-0.119	-0.166	-0.092	0.018	-0.362	0.595	0.136	0.094	0.145	-0.200	-0.286	-0.117	-0.351	-0.136	0.052	0.037	-0.004
G	0.247	-0.297	-0.142	-0.277	0.174	0.191	0.020	0.069	-0.007	0.017	0.017	0.000	-0.087	-0.001	0.136	0.163	0.093	-0.105	-0.314	0.710	0.085
B	0.249	-0.294	-0.159	-0.275	0.187	0.133	-0.002	0.086	0.017	-0.017	0.028	-0.043	-0.062	0.066	-0.071	0.097	-0.045	0.108	-0.486	-0.640	-0.072
Avg Temp	0.182	-0.287	-0.066	0.467	-0.056	-0.049	0.037	0.236	0.120	-0.056	-0.071	-0.064	-0.252	0.478	-0.339	0.318	0.004	0.145	0.198	0.064	-0.007
Sur Temp	-0.313	-0.191	-0.142	0.036	0.018	0.016	0.074	0.343	0.179	0.018	-0.258	0.212	-0.158	-0.461	-0.313	-0.005	0.294	-0.376	0.009	-0.060	-0.096
Sur Cond	-0.036	-0.315	0.457	-0.065	-0.089	-0.086	0.040	-0.137	-0.036	-0.151	0.047	-0.053	-0.138	-0.149	-0.005	0.180	-0.079	-0.176	0.054	-0.127	0.701
Avg DO	0.302	-0.044	0.102	0.296	-0.159	-0.034	0.001	0.233	0.463	0.109	-0.072	-0.030	-0.103	0.004	0.479	-0.367	-0.169	-0.263	-0.138	-0.058	0.021
Sur DO	0.152	0.130	0.423	-0.177	-0.014	0.164	-0.094	0.405	0.318	0.195	-0.016	-0.160	0.257	-0.260	-0.170	0.141	0.031	0.436	0.080	0.067	0.001
Avg pH	-0.004	0.069	0.190	0.131	0.551	-0.143	-0.428	0.299	-0.338	-0.121	0.031	-0.326	-0.220	-0.032	0.013	-0.222	0.049	-0.086	0.010	0.013	-0.023
Sur pH	-0.048	-0.296	0.468	-0.065	-0.075	-0.112	0.063	-0.166	-0.022	-0.147	0.061	-0.041	-0.117	-0.103	0.095	0.209	-0.202	-0.118	0.022	0.046	-0.689
Avg Tur	-0.104	0.015	0.066	-0.105	-0.475	0.422	-0.317	0.411	-0.375	-0.167	-0.027	0.196	0.047	0.200	0.134	0.048	-0.088	-0.141	0.008	-0.047	-0.014
Sur Tur	0.334	0.049	0.022	-0.139	0.023	-0.045	-0.099	-0.148	0.035	-0.405	-0.614	0.023	0.246	0.026	-0.319	-0.199	-0.256	-0.124	0.040	0.073	-0.003
Chl-a (0)	0.266	0.254	0.019	-0.045	0.119	-0.081	0.133	0.187	0.130	-0.368	0.627	0.327	0.084	-0.008	-0.237	0.006	-0.081	-0.255	0.043	0.013	-0.007
TSS (0)	0.135	0.202	0.274	0.190	0.375	0.055	0.287	0.072	-0.242	0.182	-0.315	0.521	0.048	0.058	0.224	0.265	0.029	-0.009	-0.042	-0.079	0.022
TSS (0.5)	0.077	0.177	-0.026	-0.269	-0.209	-0.275	0.626	0.354	-0.292	-0.032	-0.111	-0.327	-0.196	0.031	0.018	-0.089	-0.028	0.011	0.007	0.013	0.003
Secchi D	-0.280	-0.204	-0.161	0.103	0.179	-0.180	0.099	0.244	0.031	-0.083	0.008	-0.229	0.686	0.085	0.194	0.246	-0.190	-0.187	-0.001	-0.017	0.038
Depth	-0.201	0.209	0.070	-0.460	0.122	-0.150	-0.175	0.027	0.285	0.368	-0.075	0.024	-0.173	0.439	-0.090	0.143	-0.184	-0.335	0.099	-0.019	0.015
Wind	-0.119	0.028	0.143	0.141	0.192	0.692	0.351	-0.122	-0.002	0.108	0.072	-0.295	0.026	0.103	-0.226	-0.202	-0.159	-0.237	0.033	-0.003	-0.015
Humidity	0.339	0.141	0.127	0.053	-0.185	-0.032	-0.080	-0.157	-0.081	0.124	0.000	-0.276	0.234	0.092	-0.042	0.191	0.630	-0.405	-0.044	-0.105	-0.079
PAR 0/0.5	0.165	0.341	-0.312	0.105	-0.012	0.156	-0.113	-0.059	0.019	-0.048	-0.062	-0.245	-0.233	-0.385	0.131	0.530	-0.324	-0.130	0.120	-0.065	0.011