APPLICATION OF HYBRID EVOLUTIONARYALGORITHM AND THEMATIC MAP FOR RULE SET GENERATION AND VISUALIZATION OF CHLOROPHYTA ABUNDANCE ATPUTRAJAYA LAKE

LAU CHIA FONG

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

APPLICATION OF HYBRID EVOLUTIONARY ALGORITHM AND THEMATIC MAP FOR RULE SET GENERATION AND VISUALIZATION OF CHLOROPHYTA ABUNDANCE AT PUTRAJAYA LAKE

LAU CHIA FONG
SGR080170

THESIS SUBMITTED IN FULL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **LAU CHIA FONG** I/C/Passport No: **850616-02-5013**

Regisration/Matric No.: **SGR080170**

Name of Degree: **MASTER OF SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

"APPLICATION OF HYBRID EVOLUTIONARY ALGORITHM AND THEMATIC MAP FOR RULE SET GENERATION AND VISUALIZATION OF CHLOROPHYTA ABUNDANCE AT PUTRAJAYA LAKE"

Field of Study: **ECOLOGICAL INFORMATICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work,
(2) This Work is original,
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work,
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work,
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained,
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

(Candidate Signature) Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: **PROFESSOR DR DATIN SERI DR AISHAH SALLEH**

Designation:

Witness's Signature Date:

Name: **DR SORAYYA MALEK**

Designation:

ACKNOWLEDGEMENTS

I would like to use this great opportunity to give thanks to everyone who hasgiven their support on my study and research work. Thousand thanks to my supervisor **Prof. Datin Dr. Aishah Binti Salleh** for consistent support, motivation, guidance, and full support on documentation and administration work. Great thanks to my co-supervisor **Dr. Sorayya Bibi Malek**for consistent support throughout my master research. I appreciate her guidance, patient and support in my system development and thesis writing.

Many thanks go to **Professor Dr. Rosli Bin Hashim** Dean of the Institute of Biological Sciences, Faculty of Science and **Prof. Dr. Mohd Sapiyan Baba** Department of Artificial Intelligence, Faculty of Computer Science & Information Technologyfor their kind advice, helping, and supporting.

Special thanks to **Prof.Qiuwen Chen (China)** and **Prof. Friedrich Recknagel (Australia)** for knowledge sharing on data arrangement, hybrid evolutionary algorithm and support on my system development. I have a great time in University of Adelaide.

I also like to express my appreciation to **Mr Yoo Sang Nge**, **Mr Mohamad Safwan bin Jusof** and **Mr Ahmad Ali-Emran bin Emran** for rendering assistance for the use of the of High Performance Computing (HPC) resources provided by Division of Planning & Research Computing, Information Technology Centre.

Last but not least, I would like to share the achievement of this work of mine to my friends and family especially my parents and my wife**Sharon Leong**, and also my brothers **Tang Chee Kuang, Cham Hui** and **Oh Jin Heng**. I wouldn’t have completed this research without their understanding, help and support.

ABSTRACT

This study describes application of a hybrid combination of hybrid evolutionary algorithm (HEA) and thematic map visualization technique in modeling, predicting and visualization of selected algae division growth, Chlorophyta for tropical Putrajaya Lakes and Wetlands (Malaysia). The system was trained and tested using five years of limnological time-series data sampled from tropical Putrajaya Lake and Wetlands (Malaysia). HEA is run on the training set in order to provide insights on the relationships between input variables and the algae abundance. Performances of the rule sets are assessed using Receiving Operating Characteristic (ROC) with true positive rate. The generated rules are tested with another set data to avoid biasness, which yielded accuracy rate of 73%. The rules generated by HEA are then integrated with thematic map technique for visualization of the Chlorophyta abundance. Input parameters are optimized using HEA to weed out insignificant input for predicting Chlorophyta abundance. The optimized variables are namely rainfall, wind speed, sunshine, temperature, pH, dissolved oxygen, Secchi, turbidity, conductivity, total phosphorus, ammonia (NH_3-N), nitrate (NO_3-N), biochemical oxygen demand, chemical oxygen demand and total suspended solids.

ABSTRAK

Kajian ini menerangkan aplikasi gabungan antara Hibrid Evolusi Algoritma (HEA) dan teknik visualisasi peta tematik dalam peragaan meramalkan dan visualisasi khas untuk divisi alga yang dipilih iaitu Chlorophyta dalam Tasik tropika Putrajaya dan Wetland (Malaysia). Sistem ini telah dilatih dan diuji dengan menggunakan limnological data siri masa lima tahun yang diperolehi dari Tasik tropika Putrajaya dan Wetland (Malaysia). HEA dijalankan pada set latihan untuk mendapatkan maklumat mengenai hubungan antara pembolehubah input dan kelimpahan alga. Prestasi set peraturan dinilai menggunakan Operasi Penerimaan Ciri-ciri (ROC) dengan kadar positif benar. Kaedah-kaedah yang dijana kemudian diuji dengan data yang lain dengan penetapan pada kadar ketepatan 73% untuk mengelakkan prasangka. Kaedah-kaedah yang dijana oleh HEA kemudiannya disepadukan dengan teknik peta tematik untuk visualisasi kelimpahan Chlorophyta. Parameter masukan dioptimumkan menggunakan HEA untuk menyaring input untuk meramalkan kelimpahan Chlorophyta yang tidak ketara. Pembolehubah dioptimumkan iaitu takungan hujan, kelajuan angin, cahaya matahari, suhu, pH, oksigen terlarut, kadar penembusan, kekeruhan, konduktiviti, jumlah fosforus, ammonia (NH_3-N), nitrat (NO_3-N), permintaan oksigen biokimia, permintaan oksigen kimia dan jumlah pepejal terampai.

ABBREVIATION

1) ANN = Artificial Neural Network
2) AUC = Area under the ROC curve
3) BMU = Best matching Unit
4) BOD = Biochemical oxygen demand
5) CA = Cellular automata
6) COD = Chemical oxygen demand
7) Cond = Conductivity
8) CW = Central wetlands
9) DO = Dissolved oxygen
10) EA = Evolutionary algorithm
11) EcoCA = Ecological Modelling with cellular automata
12) ESA = European Space Agency
13) FL = Fuzzy logic
14) GA = Genetic algorithm
15) GIS = Geographic information system
16) GMES = Global monitoring for environmental and safety
17) GP = Genetic programming
18) GSE = GMES Service Element
19) GUI = Graphic user interface
20) HEA = Hybrid evolutionary algorithm
21) LE = Lower east wetlands
22) NH3-N = Ammonia
23) NO3-N = Nitrate
24) PL = Putrajaya lakes
25) Pop = Population
26) RGB = color model in red, green, and blue
27) RMS = Root mean square
28) RMSE = Root mean square error
29) ROC = Receiver operating characteristic
30) Sec = Secchi
31) SOM = Self organizing map
32) SYKE = Finnish Environment Institute
33) Tchlorophyta = Total Chlorophyta
34) Temp = Temperature
35) TPO4 = Total phosphorus
36) TSS = Total suspended solids
37) Turb = Turbidity
38) TPR = True positive rate
39) UB = Upper bisa wetlands
40) UE = Upper east wetlands
41) UN = Upper north wetlands
42) UW = Upper west wetlands

Table of Contents

Contents

CHAPTER ONE

INTRODUCTION AND OBJECTIVES

1.1 Introduction

Algae are the most important indicators of water eutrophication in many lakes and reservoirs around the world. Algae respond to a wide range of pollutants hence a good indicator of eutrophication. The effects of eutrophication include worsening of water quality for human consumption, recreational usage, and extinction of marine life cause by dissolved oxygen below tolerable level and ecosystem degradation. Due to the character of different types of algae will predictably and rapidly response to certain pollutants, this provides potentially useful early warning signals of deteriorating conditions of the water bodies and the possible causes (Cairns *et al*., 1972). They provide benchmarks for establishing water quality conditions and for characterizing the minimally impacted biological condition of ecosystems. Preliminary comparisons suggest that algae indicators are a cost-effective monitoring tool for lake governance and maintenance.

Thestudy focus on a specific division of algae named as Chlorophyta, also called as green algae. This division of algae contains chlorophyll a and b and obtains energy through photosynthesis. Like most of the green plants, major storage product or food of Chlorophyta which is starch been stored in stroma. Most of the species of Chlorophyta have double boundary membrane similar to plants and also form cell plates during mitosis.Chlorophyta comprises about 26% of algae population in Putrajaya Lake and amongst its most dominant genera include desmids group such as *Staurastrum*, *Cosmarium*, *Closterium* and *Pediastrum* and micro-green algae such as *Scenedesmus*, *Chlamydomonas* and *Chlorella*. Desmids are generally more common and diverse in oligotrophic lakes and ponds (Gerrath, 1993). During excessive growth of Chlorophyta following by mortality of big amount of Chlorophyta will increase turbidity of water and cut off underwater activities and cause water pollution (Saravi *et al*., 2008).They are modeled in this study as they are highly sensitive to changes in the environmental parameters that could be considered as bioindicators for monitoring water quality (Coesel, 1983, 2001).

1.2 Problem Statement

Temporal dynamics of algal communities are influenced by a complex array of biotic and abiotic factors operating through both direct and indirect pathways (Carrillo *et al*., 1995). It has been demonstrated that artificial neural networks (ANN) and hybrid evolutionary algorithm (HEA) has been successfully applied to unravel and predict complex and non linear algal population dynamic (Recknagel *et al*., 2006). The advantages these computational methods such as HEA and ANN to those existing statistical methods are the universal non-linear modeling capability. They are also not limited by the form of the data distribution (Chen *et al*, 2004).

Even though both HEA and ANN are very competitive in classifying or predicting noisy data, ANN however lack in explicit representation of rules generated to explain the model. Thus HEA has been selected for this study due to its capability to generate rule sets from complex ecological data.

However both ANN and HEA do not represent knowledge visually. Data visualization is important as it enables communication of information clearly and effectively through graphical means. Thematic map is considered as an effective method of data visualization that is widely used for representation of ecological data (Few, 2010). Thematic map had proven success in many areas of ecological informatics research such as management of coastal greengold, detection of the toxic dinoflagellate and marine eutrophication (Congalton, 1998; Klemas, 2009).

1.3 Research Objectives

The research aims at

(1) Extracting generic relationship and pattern of Chlorophyta abundance with respect to water quality in Putrajaya Lakes and Wetlands using hybrid evolutionary algorithm;

(2)Developing datadata visualization system using Thematic Map Technology;

(3)Integrating rule sets discovered by HEA with thematic map technology for visualization of Chlorophyta abundance.

CHAPTER TWO

LITERATURE REVIEW

2.1 Lakes and Wetlands

Lakes contain a very small part of global amount of water that is around 0.01% and they are open system in exchange energy and mass with the environment (Jorgensen *et al.*, 1989). Lakes are influenced by controllable and non-controllable variables. Examples of controllable variables are inflow and outflow of water, nutrients, toxic substances and more. Examples of non-controllable variables are solar, wind, radiation and precipitation. The state of lakes determines by the use of internal variables of lakes such as phytoplankton, nutrients and fish concentration on those controllable and non-controllable variables. In consideration of the function of lacustrine ecosystem (also called as still water ecosystem), all these chemical, physical and biological factors must be taken into account.

Among Southeast Asian countries, lakes contain an approximate sum of 500cubic kilometers of high quality freshwater, so lakes are important in terms of ecology and economy (Lehmusluoto, 2003). Lakes and reservoirs also act as storehouses of waters, important ecological entities and sources of food and also help in preventing floods.

According to the Ramsar Convention on 1971 (Ramsar Convention Secretariat, 2013, wetlands are defined as land inundated with temporary or permanent water that is usually slow moving or stationary, shallow which the depth of low tide does not exceed six meters, either fresh, brackish or saline, where inundation determines types and productivity of soils and the plant and animals communities. Wetlands in Malaysia include lakes, rivers, mangroves, peat swamp and freshwater swamps and most of the areas come after the jurisdiction of State Government or other forms of protection such as forest reserves (Zakaria *et al.*,2009). Total areas of wetlands in Malaysia are 3.5 to 4.0 million hectares or 10% of the land areas (Aik, 2002).

Wetlands are natural medium in cleaning the river water from pollutants. Efficiency of wetlands are affected by retention times, pollutant loading rates, hydrology, sedimentation process, morphometry and biological processes. Management and maintenance of wetlands is to ensure that the system is not overloaded, in order to provide the diverse habitats for aquatic fish, to ensure a balance of phytoplankton and macrophyte communities, to prevent invasive weeds and excessive sedimentation and to control mosquito outbreaks. Many of the Man-made wetlands were constructed to minimize negative impacts of pollutants from urban and agricultural runoff. According to literature studies and reports, wetlands effectively improve water quality (Martin *et al*., 1994).

2.1.1Putrajaya Lakes and Wetlands

Putrajaya Lakes and Wetlands are located in the middle of west coast of Peninsular Malaysia. Peninsular Malaysia experienced Southwest Monsoon which is relative to dry weather from late May to September, and the Northeast Monsoon which brings heavy rainfall from November to March. West coast of Peninsular Malaysia is protected from the Northeast monsoon by the Titiwangsa mountain range, because annual rainfall in Putrajaya city is about 2200mm which is slightly lower than the average rainfall of 2500mm in peninsular Malaysia.

Figure 2.1 show the map of Putrajaya city. Putrajaya Lakes a man-made lake surrounded Putrajaya city and act as natural cooling system for Putrajaya city. Putrajaya lakes were created by inundating 400 hectares valleys of Sungai Chuau and Sungai Bisa. 170 hectares of Putrajaya Wetlands were constructed as natural treatment system to treat primary upstream inflow into the lakes (Perbadanan Putrajaya, 2006).

Expected results in increasing of pollutants from Sungai Chuau and Sungai Bisa, wetland system are constructed to straddle courses. Wetland system comprises 6 arms as shown in Figure 2.2 and divided to 24cells by a series of rock filled weirs along the six arms. All arms are connected to each other but all are differ in size, depths, plant communities and pollutants loads that it is designed to handle. They are important to maintain the functionality of wetlands as in providing a habitat for local fauna, primarily mammals, water birds, reptiles, amphibians, fish and invertebrates; providing flood detention area and reducing peak dischargers and flow velocities, and recreation. All the arms discharge into central wetland before flows down into the Putrajaya Lake (Perbadanan Putrajaya, 2006).

The Putrajaya Lakes are at the southern of the wetlands. It is categorized as shallow polymictic oligotrophic lake. Water flow to the lake came from wetlands 60% and direct discharge from bordering promenade 40%. The buffer feature along the lake shorelines contributed from 20 m width promenade. The total volume of the whole lake water is about 23.5 million cubic meters and the water depth is in range of 3 to 14 meters.

The design of Putrajaya lakes and wetlands features a multi-cell multi-stage system with flood retention capability. This will maximize space available for colonization by water plants. Those plants will also act as pollutants interceptor and will provide a root zone for bacteria and microorganisms to act as assistants in filtering and removing water pollutants.

Putrajaya wetlands were designed with multi-cell approach. There are 6 wetland arms with a total of 24 cells separated by two-three meters height of bunds. There are 8 cells each in Upper West Wetlands (UW1 to UW8) and Upper North Wetlands (UN1 to

UN8), 3 cells in Upper East Wetlands(UE1 to UE3), 2 cells each in Lower East Wetlands (LE1 and LE2) and Upper Bisa Wetlands (UB1 and UB2), and only one cell in Central Wetlands. This study has adopted the zonation of Putrajaya wetlands into the thematic map development.

Putrajaya lakes and wetlands ecological and environmental aspects have become very interesting areas for ecological scientists in further research. Putrajaya Lakes is located surrounding Putrajaya city. It provides recreation, landscape features, ecotourism, education, sports, tourisms and research. It built to fulfill the goal of Garden City as if turns nature for inspiration, provides a picturesque lakes in the landscape. As the objectives above, lake water quality is the biggest challenge and important factor for the achievement of its goal. Catchment of water indicates that it carries elevated level of pollutants. This is due to the upstream inflow and also boundary of Putrajaya development. As shows in studies, the concentration level of pollutants in Sungai Chuau and Sungai Bisa shows increment after times, the inflows to Putrajaya lakes also expected to increase the level of water pollution in the lakes.

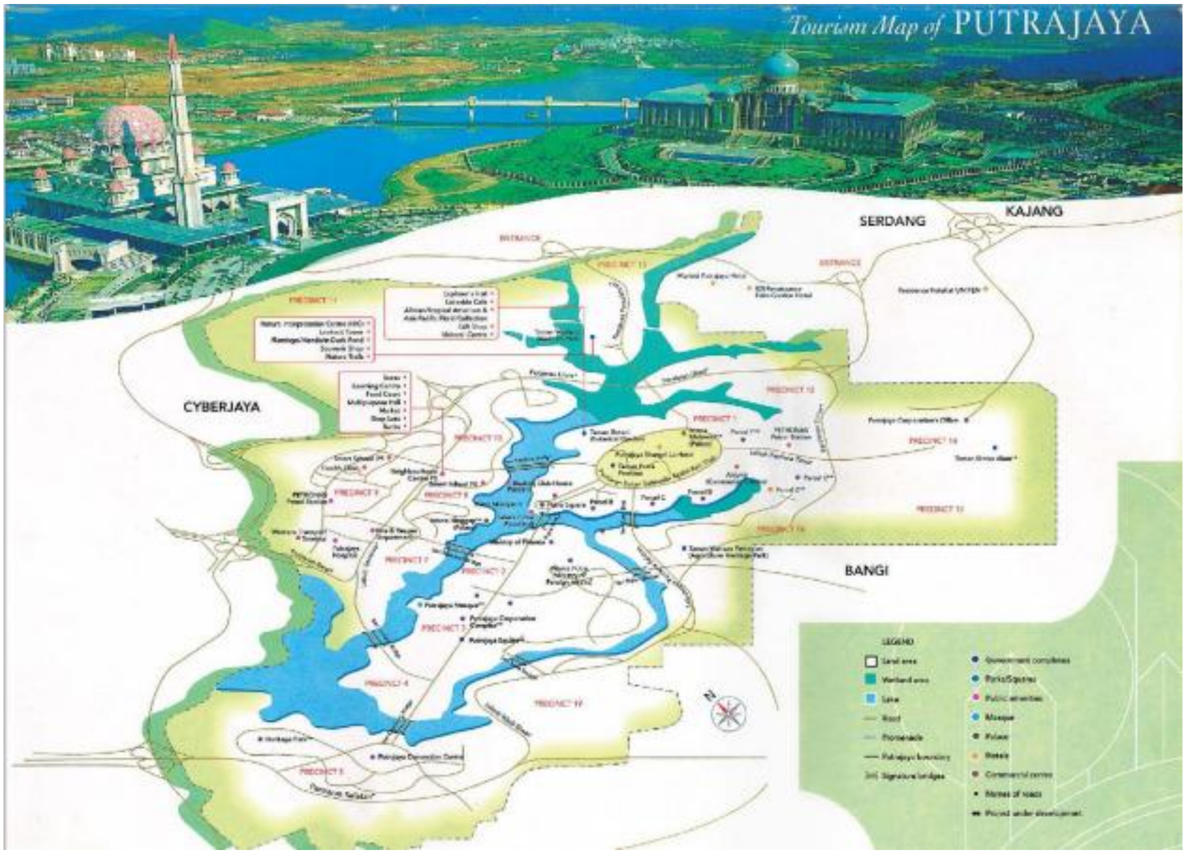

Figure 2.1 : Map of Putrajaya Lakes and Wetlands

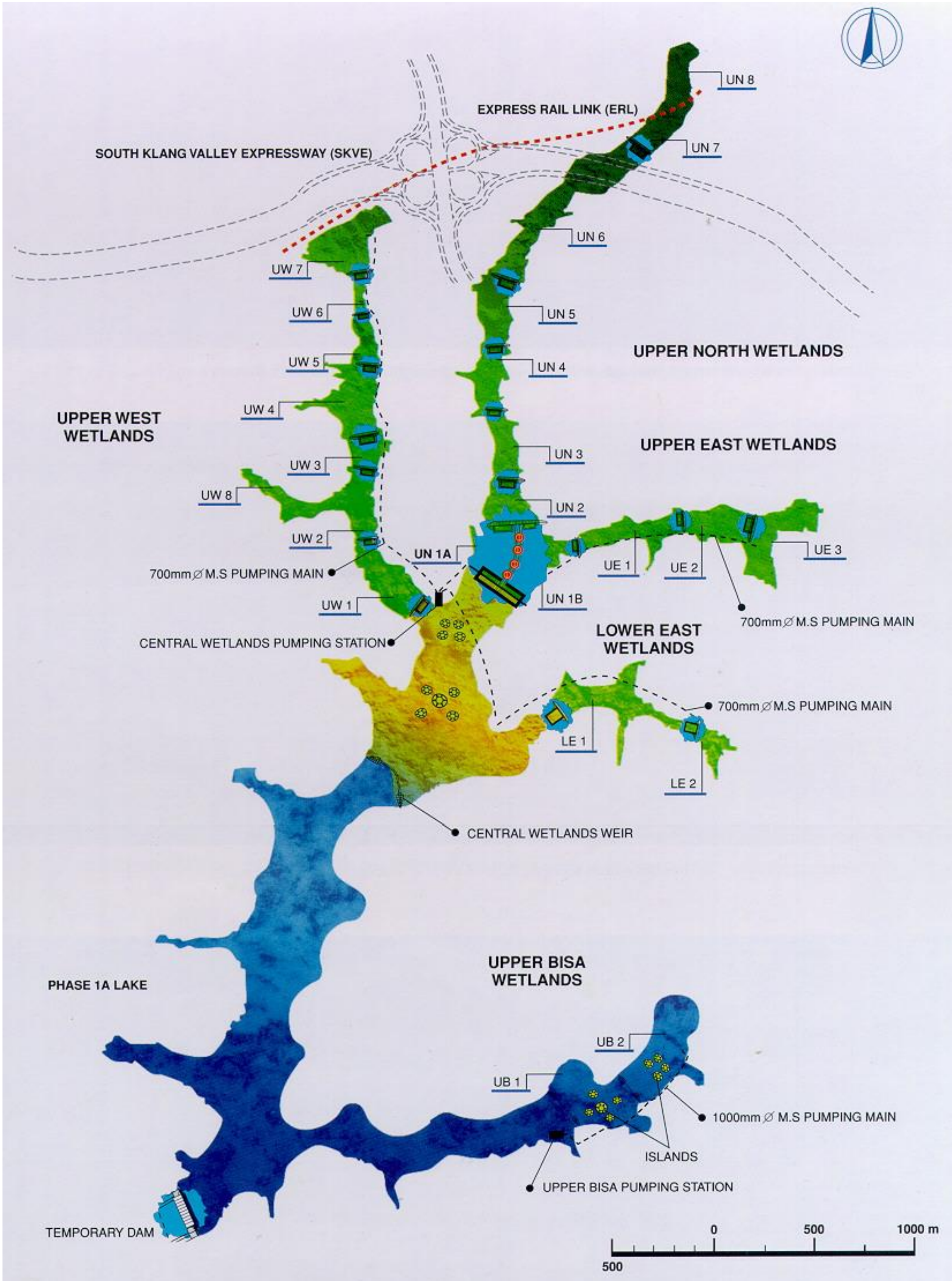

Figure 2.2 : The Putrajaya Wetland Cells and its location (Perbadanan Putrajaya, 2006)

2.2Algae

In common ways, algae are define as some micro plants which lack of true roots, stems, leaves and flowers. Algae are a large and diverse group of primarily aquatic plantlike organisms. Recently algae had been classified in major group called eukaryotes.

Algae had been customized with colors for each of the division, division in green called Chlorophyta, division in brown called Phaeaphyta, division in golden brown called Chrysphyta, division in blue-green called cyanobacteria*(*dangerous division with most of the species will cause water pollution*)* and division with red called Rhodophyta. Other characteristics of algae, such as type of photosynthetic food reserve, cell wall structure and compositions, have been important in further distinguishing the algae divisions.

As one of the larger diversity compare to other algae, Chlorophyta had caught the attention of researchers. Chlorophyta also been called as the green algae exist from the range of green to orange colour. Chlorophyta contain photosynthetic pigments which came from chlorophyll a and b which give green colour character to it. And those orange colour are form due to Carotenoids. Chlorophyta are predominately autotrophs.

Chlorophyta grow in almost every part of the world especially in wet places like lakes, ponds and streams as well as on the shaded sides of damp walls and trees. Growth of Chlorophyta is affected by water condition such dissolved oxygen, temperature, pH, salinity, and turbidity. In addition to these, Chlorophyta require nutrients such nitrate, phosphate and silicates for their survival and growth. All these factors determine the spatio-temporal abundance of Chlorophyta.

2.2.1 Chlorophyta

Main genera of Chlorophyta which are commonly found in Putrajaya Lakes and Wetlands are *Ankistrodesmus, Chlorella, Closteriopsis, Cosmarium, Crucigenia, Pediastrum, Scenedesmus, Staurastrum* and *Tetrahedron*.Chlorophyta also commonly known as the green algae contain more than 7000 species. It is the most diverse algae group growing in different environment. The green algae is excluded from Plantae, it consider as a paraphylethic group with containing chlorophyll. Chlorophyta contains two types of chlorophyll which enable them to photosynthesis. During photosynthesis, chlorophyll been used to capture light energy and with attendance of water to fuel sugar manufacturing, this does differ them from some of the other primary aquatic. As shown in the systematic treatment of McCourt (1995), there are three Classes of green algae in Chlorophyta :Ulvophyceae, Chlorophyceae and Trebouxiophyceae.

Ulvophyceae are mostly marine but some of them also occur in freshwater habitats. They are groupin range of uninucleate to multinucleate filaments to siphonaceous forms to giant unicells. Most of the diploid green seaweeds belong to this class. Some of the examples from this class *are Ulva*, *Cladophora* and *Codium*. *Cladophora* are mainly found in freshwater.

Most of the taxa of Chlorophyceae includes unicellular, colonial or filamentous. Chlorophyceae divided into two subgroups which known as directly opposed basal bodies clade (DO clade) and the clockwise arrangement of basal bodies clade (CW clade). Their flagella are non-scaly and its roots run in periphery of cell. Some of the examples from this class are *Pediastrum*, *Hydrodictyon* and *Oedogonium*.

As the third group of Chlorophyta, Trebouxiophyceae mainly found in soil. Trebouxiophyceae undergo distinctive mitosis called metacentric mitosis, which can also be explained by mitosis with polar centrioles. They are group in range of unicell to small sheets and filaments of cells.

Ankistrodesmus have long and needle shaped cells, and have high tolerance for copper treatments such as copper sulfate which use in algal growth control. Chlorella are one of the mainly found in latest science technology in producing superfoods which acts like vitamins and others type of supplements. *Chlorella* is a single-cell alga. It ability in photosynthetic efficiency can reach as high as 8%. Scientist believes that *Chlorella* might contribute in generating energy. *Closteriopsis* belongs to the class Trebouxiophyceae. *Cosmarium* are single-celled algae. According to Stamenković *et al*. (2008), *Cosmarium* and *Staurastrum* has high tolerance to water pollution as they found inhabit in alkaline and eutrophic freshwater ecosystems with containing toxic and heavy metals compounds, ß-radioactivity and considerable amount of mineral salts. *Crucigenia* and *Pediastrum* are from Chlorophyceae class. *Pediastrum* and *Scenedesmus* are nonmotile colonial green algae. *Pediastrum* are found in colonies of at least four cells with star alike pattern. *Scenedesmus* are commonly found in colonies of two to four cells and aligned in a flat plate(Lewis *et. al*., 2004).

2.2.2 Algae Bloom & Problems

Excessive algae growth in lakes, ponds, rivers and others water bodies cause serious problem to the water quality. It forms a thick layer of mats floating on the surface andprevents sunlight to pass through it. Such excessive algae growth phenomenon had been named as algae bloom.

As mentioned above, when the turbidity of water increases, sunlight cannot reach the deeper layers of water body and thus partially or completely inhibits decomposition of organic matter. After death and decay of algae, it also adds large amount of organic matter. Due to the turbidity problem of water body and rapid accumulation of organic matters, it causes serious water pollution. Some others algae also produce harmful and toxic substance to fishes and other aquatic animals. It also harms some of the land animals which drink this water. Most of these algae come from division of Cyanobacteria such as *Microcytis and Aphanizomenon.* Algae blooms generally will lead to stinky and oily water, fishy taste and not suitable to use as drinking water. During blooming session, many species from Cyanobacteria and Chlorophyta will cause unpleasant smell and tastes, large change in pH. Decomposition of large amount of organic matters also cause decrement of dissolve oxygen level, thus will causes endangers fish, high costs of water treatment plants operation, discouragement of tourism and might leads to poisoning of humans and other animals.

Type of water pollution and its polluted level in certain water body can be identified by analysis of the composition and growth pattern of algae. Such study and research of algal have been used to identify several types of water pollution problems. For example, increases of water acidity level will increase growth of filamentous algae. The pH level of the water body can be indicated accurately by changes in the species composition of diatoms as most of the algae and diatoms disappear in water below the pH level of 5.8 (Park, 1987), due to diatoms are highly sensitive to pH and in different pH values of water body different types of diatoms species will only be found.

Excessive addition of phosphates, nitrates, or organic matter will lead to blooming of algae such as *Microcytis, Scenedensmus*, *Hydrodictyon* and *Chlorella*. The blooming of algae which absorb and accumulate heavy metals such as *Cladophora* and *Stigeoclonium* will indicate heavy metals pollution in the water body. Oil pollution of water body can caused by excessive growth of algae like *Dunaliella tertiolecta, Skeletonema costatum, Cricosphaera carterae, Amphidium carterae, Cyclotella cryptica* and *Pavlova lutheri.*

A simple reference of trophic state index related to algal biomass base on Secchi disk transparency had been published (Carlson, 1977). Staurastrum is the main genus that dominates the water surface of the lake. Blooming of Staurastrum will lead decrement of water quality and water pollution. Besides that, present of Staurastrum will bring along uneasy smells, taste and intoxication to an aquatic ecosystem (Saravi *et al.*, 2008).

These Freshwater phytoplankton communities often undergo pronounced seasonal succession (Reynolds, 1984). The succession pattern in a lake is fairly repeatable among years, and patterns among lakes are somewhat predictable according to tropic status (Reynolds, 1984; Sommer *et al*., 1986). In order to prevent pollution problems in water body, various studies and research of the forces driving phytoplankton succession had been carry out, however, remains a difficult task since the temporal dynamics of algal communities are influenced by a complex array of biotic and a biotic factors operating through both direct and indirect pathways (Sommer, 1989; Vanni & Temte, 1990;; Carillo *et al*., 1995).

2.3 Ecological Modeling

Research in computational technologies to monitor algae growth for monitoring lake status has been developed for temperate lakes for the past 40 years. Aggregated-based ecological model is one of the oldest approach and it has been developed since the past 40 years. Those models will lumps species into biomass and formulate the dynamics into partial differential equations (PDE) form. But yet these approaches had proved to be unproductive by unable to reproduce realistic results where differences of individual properties and local interactions play a significant role in determining the relationship between populations, and between species and their surroundings.

Learning in neural networks is activated by changing the connection weights of the networking response to the example inputs and the desired outputs to those inputs. These adjustments of connection weights to learn the desired behavior is called the training period. This is followed by the operation period when the network works with fixed weight and produces outputs in response to new patterns. There are two types of ANNs training methods, namely supervised and unsupervised. Supervised type of ANNs models have been successfully implemented for eutrophication modeling and lake management in ecology (Melesse *et al*., 2008; Sorayya *et al*., 2009, 2010; Recknagel *et al*., 1997, 2006; Maier *et al*., 1998; Wilson and Recknagel, 2001). Meanwhile, self organizing feature map (SOM) which is an unsupervised type of ANNs allows knowledge discovery. SOM reduces the dimensions of data of a high level of complexity and plots the data similarities through clustering technique (Kohonen, 2001). SOM has been used effectively in ecological modeling of temperate water bodies (Recknagel *et al*., 2006). However, these models are mostly ‘black box’ in nature whereby the knowledge is hidden within the system parameters and little is made known in understanding the relationship of algae dynamics with regard to the environmental factors even though ANN models are able to make perfect predictions and are recognised as powerful, they are considered to be ‘black-box’ in nature. Therefore explanatory method such as HEA has been adopted in this study with the idea to clarify the‘black-box’ approach of ANNs. HEA approach can overcome the limitation of ANN approach. HEA allows discovery of predictive rule set in complex ecological data.

2.3.1 Hybrid Evolutionary Algorithm

This study adopted hybrid evolutionary algorithm based on Cao*et al*. (2006) to discovery generic rule set for Putrajaya Lakes and specific rule sets for each part of the wetlands. Hybridization of evolutionary algorithms is getting popular due to their capabilities in handling several real world problems involving complexity, noisy environment, imprecision, uncertainty and vagueness. The advantage of using these technique compared to artificial neural network is that algorithm that generates rule and provide a better performance in term of RMSE (Sorayya et. al., 2011).

Hybrid Evolutionary Algorithm (HEA) was the evolution of Evolutionary Algorithm (EA) with parameter optimization. It been developed to HEA to put as part of larger system. It also improves the ability to search for good solutions. 4 stages of EA that is initial population, mating pool, and 2 Offspring stage, the initial population stage is the part that known solutions, constructive heuristics, selective initialization and local search. Between the mating pool stage and the first offspring stage and also second offspring, there is a hybrid process called crossover and mutation. And it will involve use of problem-specific information in operators. HEA has been integrated and shown successin ecological data warehouse research in prediction and explanation of water quality and habitat conditions. Latest research of hybrid evolutionary algorithm also been integrated into some data visualization approach such as cellular automata. The main character of HEA was solving problems involving complexity, noisy environment, imprecision, uncertainty and vagueness. Due to the characteristics of HEA, it is very suitable to use as back-end tools for data visualization.

The hybrid evolutionary algorithms (HEAs) have been ad *hoc* designed as flexible tool for inducing predictive multivariate functions and rule sets from ecological data. Flow chart in Figure 2.3 shows conceptual framework of the application of HEA for rule discovery for one chromosome. It indicates that similar to supervised ANN, the training of HEA aims at the optimal approximation of the calculated output *Y*calculated to the original result *Y*original. HEA iteratively adjusts the rule structure and parameter values rather than input weights in order to minimize the error (*Y*original _ *Y*calculated). HEA framework (Figure 2.3) adopted in the present study was developed by Cao *et. al.*(2006).

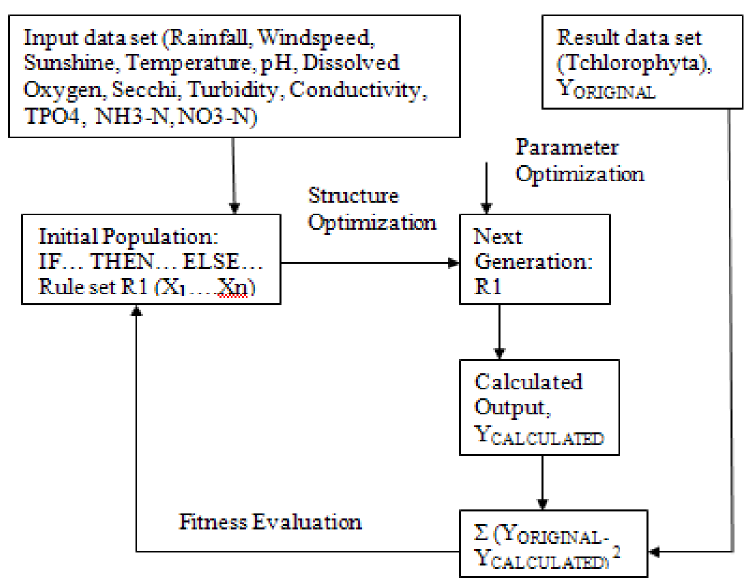

Figure 2.3: Conceptual Framework of Hybrid Evolutionary Algorithm for One Chromosome (One Run) (Cao *et al*., 2006)

Hybrid Evolutionary Algorithm (HEA) has been applied by Recknagel *et al.* (2006) on shallows and hypertrophic Lake Kasumigaura (Japan) was compared with the deep and mesotrophic Lake Soyang (Korea). Artificial neural networks (ANN) and evolutionary algorithms (EA) had been used for ordination, clustering, forecasting and rule discovery of complex limnological time-series data of two distinctively different lakes. One week ahead forecasting of outbreaks of harmful algae or water quality changes had been done using Recurrent ANN and EA. EA facilitate and discovering rule sets for timing and abundance of harmful outbreaks algal populations. Non-supervised ANN provides ecological relationships regarding seasons, water quality ranges and long-term environmental changes.Accuracy in forecasting and the ability in expalaning timing and magnitude of algal population make performance of EA to be superior compared to recurrent supervised ANN (Recknagel et. al., 2005).

Research had been done in shallow hypertrophic Lake - Lake Suwa in Japan, with both non-supervised artificial neural networks (ANN) and hybrid evolutionary algorithms (EA). Both approaches were applied to analyse and model 12 years of limnological time-series data in Lake Suwa. The results have improved understanding of relationships between changing microcystin concentrations, Microcystis species abundances and annual rainfall intensity. Non supervised ANN had revealed the relationship between *Microcystis* abundance and extra-cellular microcystin concentrationduring dry and wet years.The result successfully shows that dry year is higher compare to typical wet year. Non-supervised ANN also showed that high microcystin concentrations in dry years coincided with the dominance of the toxic Microcystis viridis whilst in typical wet years non-toxic Microcystis ichthyoblabe were dominant. Hybrid EA was used to discover rule sets to explain and forecast the occurrence of high microcystin concentrations in relation to water quality and climate conditions (Recknagel et. al., 2007).

ANN and HEA had been successfully applied to tropical water by Sorayya et. al., 2009-2011. In one of the study, four predictive ecological models; Fuzzy Logic (FL), Recurrent artificial neural network (RANN), hybrid evolutionary algorithm (HEA) and multiple linear regressions (MLR) had been applied to forecast chlorophyll- a concentration using limnological data from 2001 through 2004 of unstratified shallow, oligotrophic to mesotrophic tropical Putrajaya Lake (Malaysia). Performances of the models are assessed using Root Mean Square Error (RMSE), correlation coefficient (r), and Area under the Receiving Operating Characteristic (ROC) curve (AUC). Chlorophyll-a have been used to estimate algal biomass in aquatic ecosystem.

2.4 Data Visualization Approaches

Research in data visualization includes the studies in visual representation of data and information in some kinds of schematic form, maps, graphs and diagrams. Some of it does include the attributes and variables for the units of information. According to Friedman (2008), to communicate information in a clear and effective way through graphical means is the main goal of data visualization. And it is regardless of how beautiful and sophisticated of the design. Both aesthetic form and functionality need to join together and present into a rather sparse and complex data set by using a more direct way to communicate its key-aspects in order to deliver ideas effectively. Communicating information is the main purpose of data visualization, so the design of data visualization has to balance between design and function, and creatingstunning data visualizations.

In general, data visualization has split into four categories such as information graphics, scientific visualization, information visualization and statistical graphics. Nowadays, data visualization has been widely used in area of research, development, teaching and studies. According to Post *et al.*, (2002), data visualization has united the field of scientific and information visualization.

Information presentation is the main focus point for all kinds of approaches on the scope of data visualization. In his presumption, statistical graphics and thematic cartography are the two main parts of data visualization (Friedman, 2008). Statistical graphics is widely used in visualize quantitative. It includes histogram, probability plots, box plots and etc. Thematic cartography involves maps of specific geographic themes towards specific audiences such as population in certain country. Because of these, thematic cartography has been used in ecological informatics research such as cellular automata and remote sensing.

2.4.1 Thematic Map and Application

Thematic map is a type of mapdesigned with specific theme and topic. Most general thematic maphad been display or view and well known by public such as world map with population distribution or temperature. Thematic map does not contain any physical features such as rivers, roads and subdivisions. Thematic map features were to enhance understanding of its theme and purpose to everyone. Normally thematic map used city locations, countries map, rivers and other geographical locations as its base maps. Theme is added onto those base map using different mapping programs and technologies such as geographic information system (GIS).

There are a few important points to be considered in designing thematic cartography. The most important consideration is the map's end users. End users help to determine the thematic map content in addition to its theme. Secondly, the base map for thematic cartography has to be accurate, up to date and it has to come from reliable sources. In order to create an accurate thematic map, various ways to use that data and it has to take consideration on map's theme. There are univariate, bivariate and multivariate data mapping. As shown from the name of those ways, univariate data mapping is use for only one type of data, bivariate data mapping is use to show distribution of two data sets and correlations between them and as well as multivariate data mapping.

In thematic cartography, data can be presented in many creative ways. There are 5 most common ways of thematic maps techniques which are Choropleth map, Proportional or graduated symbols, Isarithmic or contour map, Dot map and Dasymetric mapping.

Chloropleth map represents quantitative data such as percentage, density and average value of an event in geographical area using colour. Different colour represents a certain range of data. Proportional symbols are the second type of thematic map that used symbols to represent data and associated with location points. Proportional sized of symbols been used to represents data in different occurrences. Symbols with proper geometrical shape such as circle, triangle and square are commonly used. The areas of the symbols are made proportionally to the values represented.

Third type of thematic map is contour map. It used contour line to represents continues values such as temperature and rainfall, as well as represents three-dimensional values such as attitude of a certain geographical area. The basic rule for a contour map is that it follows high and low side in relation to the isoline. Dot map is the fourth type of thematic map. It normally used to present an occurrence of a theme or a spatial pattern. On a dot map, a dot can represent one or several unit depending on what information on the map had been display.

The last one is dasymetric mapping. This map is a complex version of choropleth map with used the statistical analysis values and extra information to combine areas with similar values(Briney, 2009).In this study, the type of thematic map chosen was chloropeth map because it matches the aim of the study by showing Chlorophyta abundance in Putrajaya Lakes and Wetlands.

Thematic map had been integrated to assess and evaluate the coastal environment in Biscay, Spain. Different parts of coastal environment which to be used for human activities had been assessing their capability by integrating thematic map. A series of thematic maps with different aspect such as geological, biological and dynamic had been elaborate. With combination of all these thematic maps, the impact of homogeneous units with corresponding to a series of activities had been evaluated (Cendrero A.*et al*. 1979). The research shows the capability of thematic maps in presenting the data and also shows the effectiveness of thematic maps in elaboration of geological and biological data.

Thematic map had been used in enhancement for research in ecosystem modelling. Thematic mapenables research to move forward in application of visualization for prediction and forecasting. Thematic maps approach also been used to detect algal blooms in the Baltic Sea. By using Envis at MERIS satellite images, one to two satellite images are obtained per day. Algae being classified to four classes as in no, unlikely, potential and likely surface algae. Satellite image been combined with RGB colour to differentiate the algae blooms classes. Extra areas detected in the satellite image such as land areas been coloured by grey and clouds are all shown in white. The sizes of each pixel in each of the satellite image had been set to 300m x 300m and 1km x 1km. Data Also been provided and processed by Finnish Environment Institute (SYKE). Thematic map technology as a part of visualization system had been used by SYKE(Finnish Environment Institute, 2009)to present water quality (Secchi and turbidity) data pf a project that ended in 2011.

Qua *et. al.* (2008) have used cellular automata to study the effect of the selective and random harvest on the ecosystem sustainability and management. They demonstrated the advantages of cellular automata to stimulate more realistic predation-harvesting system.One of the latest researches in ecological informatics thathas been successfully developedin Australia and China using thematic map is cellular automata (Chen et. al., 2002). A Cellular Automaton is a mathematical system in which simple components act together to produce complicated pattern of behaviour. It starts from complete disorder and when irreversible evolutions go on, it will generate an ordered structure. This process also knows as self organization. It provides a powerful and flexible fuzzy logic modelling technique for uncovering patterns in ecological data.

A cellular automata system usually consists of a regular lattice of sites which called as cells or automaton, each site has some properties that are updated in discrete time steps according to local evolution rules. This has provided strong characters for cellular automata as in parallelism, homogeneity and locality. In further explanations, all cell states are updated simultaneously; all cells will follow the same evolution rules and will affect each others in the direct neighbourhood. Application of cellular automata is diverse and recently has caught the attention of biological scientist in the application of thematic cartography(Chen et. al. 2001).

In further research of ecological modeling system, a cellular automaton based prey predator (Ecological modeling with cellular automat also known as EcoCA) was developed. Effects of the cell size and configuration in cellular automata (CA) based prey-predator modeling was studied by Chen and Mynett (2003). They proposed to use principal spatial of studied ecosystem and apply Moore type cell configuration to achieve size-independent and consistent model behaviours.

An implementation of thematic map in monitoring water quality and algal bloom had been carried out in North Sea and South Sea of Germany, Europe. The program is called GMES which stands for Global Monitoring for Environmental and Safety had been carried out since 1998. The program provides services for users for data combination from in-situ measurements, space observation and model which will provide information and support policy for environmental safety. With combine force of ecology and space technology, the Europe commission contributes in research and development and European Space Agency (ESA) contribute in GMES Service Element (GSE) (Commission of the European Communities, 2003). ESA had developed sentinel series of satellites to deliver routine information about the ocean state such as surface height, waves, wind, temperature and ocean colour over the next 20 years (Drinkwater *et al*., 2005). A series of services named as downstream services took part in converting raw satellites data, in-situ observations and numerical models into ocean state and forecasting baseline information (Ryder, 2007). A prototype of marine downstream service is currently operating by the name Project MarCoast which deliveres satellite-based pre-operational services. The MarCoast Services deliver the water quality measurement such as suspended matter concentration, chlorophyll concentration, turbidity, algal bloom indicators and sea surface temperature to European users for monitoring the marine environmental monitoring.

Thematic maps had been generated by integrating satellite data combine with in-situ measurements and delivers through the MarCoast Services Portal (MarCoast, 2012).

Literature suggeststhat thematic map as a data visualization approach has never been used in ecological modeling of algae in Malaysia. Most of the research paper on application of data visualization and cellular automata in developing ecological modeling system had been widely develop and proven its stability and productivity in countries such Australia, China and Dutch. In Chen (2004), modeling on competitive growth and explanation of succession processes of two underwater species *Chara aspera* and *Potamogeton pectinatus* in eutrophicated lake has been done. Hence, data visualization approaches need to be developing using Tropical lakes data to improve the maintenance and monitoring of the selected topical lake.

CHAPTER THREE

METHODOLOGY

3.1 Study Area

Putrajaya city contain most of government departments, global commercial offices, and residential areas. Recreation park and water bodies surrounding it in Putrajaya city acts as natural cooling system. Putrajaya city is building up with covering of more than 30% of green areas from the total land space. Most distinctive features of this city arethe development of Putrajaya Lakes which cover 650 hectares. Putrajaya Lakes is created by the construction of a dam at the lower reaches of River Chuau and Sungai Bisa.

In balance of ecosystem for that area and to maintain the water quality standard, 23 cells of wetlands with total area of 197hectares had been constructed.Over 70 species of wetlands plants in total amounts of 12.3millions plants had been planted into the wetlands. The wetlands act as a natural filtering and treatment system for Putrajaya Lakes. As water flows from the rivers into wetlands, most of the pollutants had been filter before it enters the lakes. Functions of these wetlands are to purify inflows water by removing phosphorus, organic compounds, oxidizing ammonia and nitrates. Putrajaya wetlands also act as flood mitigation. The wetlands are design using a weir to separate each of the cells and also known as multi cell and multi stage approach. As the water flows across the wetlands, each of the cells gives different treatment to the water as they are all in different water levels and design for different purpose. The extra advantages of the design are good flow distribution, thus maximize shallow areas for the encouragement of macrophytes growth and facilitate a more cost-effective maintenance of weeds and insects. The construction of wetlands begins in March 1997 and completed in August 1998.

Figure 3.1 showing the water sampling points from Putrajaya Lakes and Wetlands. Water sampling for this study was carried out in the morning twice a month at 23 fixed substations of 13 major sampling stations during the years 2001 through 2006. The water samples were collected near shore at the depth of 0.5 m and samples were analyzed for each sampling stations. The sub-stations were divided arbitrarily into two sets (dataset A and B). Data from dataset A was used for training using HEA models and dataset B for testing. The dataset will be classified into Low, Medium and High. Water sampling for water quality and algae abundance analyses were carried out according to APHA (1995) and WHO. Water samples for algae identification were collected using plankton net with mesh size of about 30 µm. Smaller mesh size plankton net was not used because of problems of clogging and reduced water flow during the transfer of water to vials for subsequent analysis (Bellinger and Sigee, 2010). Each water sample for algae analysis was gathered from several scoops of the site water to reduce the chance of missing out smaller Chlorophyta from the sample. Algae were preserved by adding several drops of 4% formaldehyde into the water samples that were subsequently kept in 50 ml vials. Identification of algae was carried out using ordinary light microscope. Identification of algae genera was based on literatures such as Werh and Sheath (2003) and Bellinger and Sigee (2010). Algae abundance was calculated using sedimentation technique and procedure described by Evans (1972).The principal features of Putrajaya lakes and wetlands are listed in Table 3.1. Other information on the water characteristics of Putrajaya Lakes and Wetlands are shown in Table 3.2.

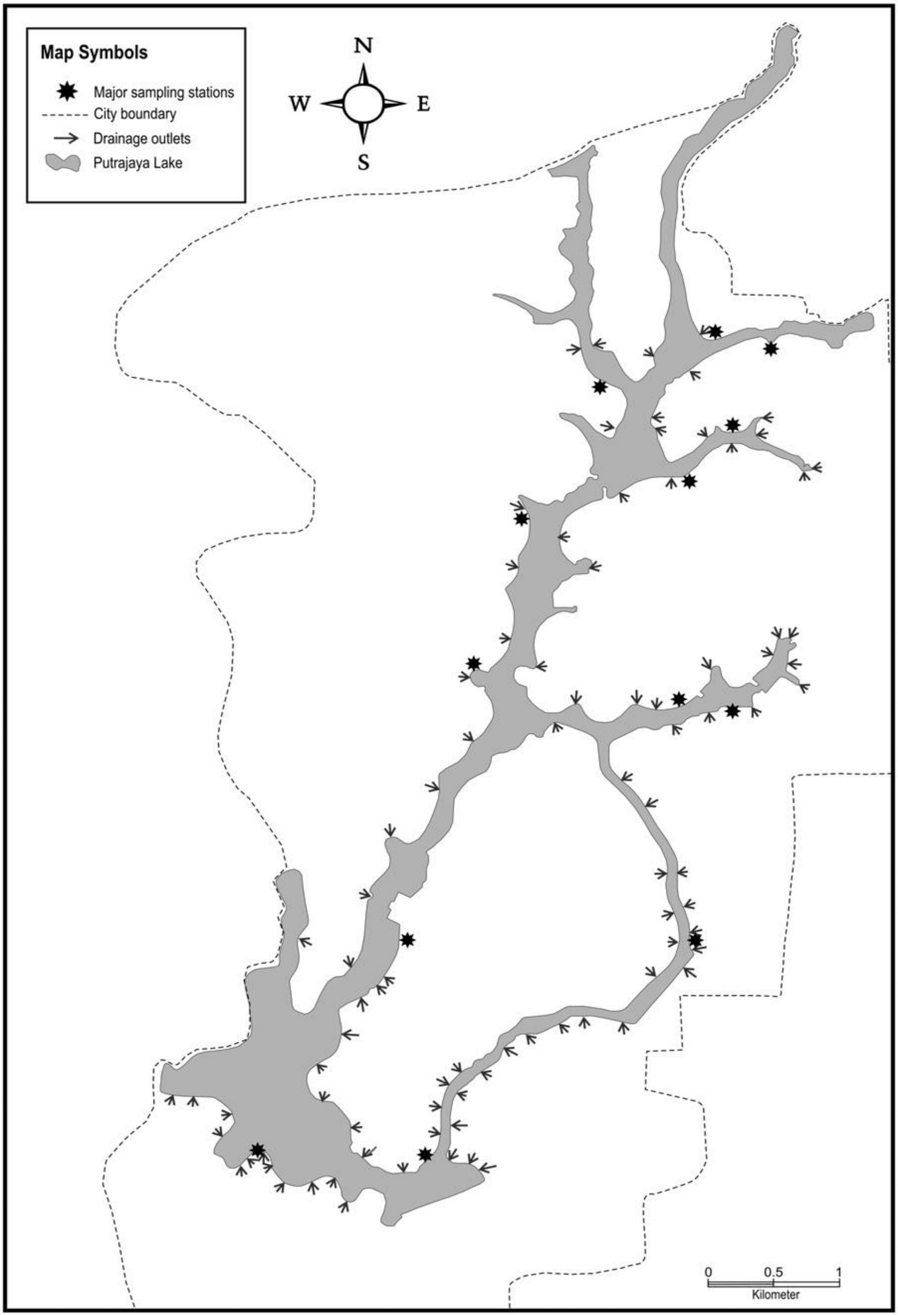

Figure 3.1 : Map of Putrajaya Lakes and Wetlands showing water sampling point

Table 3.1: General Characteristics of Putrajaya Lakes and Wetlands

Putrajaya Lakes and Wetlands	
Climate	Tropical
Trophic Status	Origotrophic
Putrajaya Wetlands	
Total Areas	197.2Hectares
Planted Area	77.70Hectares
Open Waters	76.80Hectares
Weirs and Islands	9.60Hectares
Zone of Intermittent Inundation	23.70Hectares
Maintenance Tracks	9.40Hectares
Putrajaya Wetlands	
Catchment Area	50.90 KM^2
Water Level	RL 21.00M
Surface Area	400Hectares
Storage Volume	23.50Mil M^3
Average Depth	6.60M
Average Catchments Inflow	200 millions L
Average Retention Time	132days

Table 3.2 : Mean limnological properties of Putrajaya Lake from 23 sampling stations collected from 2001 until 2006 (Perbadanan Putrajaya, 2006) .

Measure data of Putrajaya Lakes and Wetlands Year 2001to2006(Biweekly)	Minimum Values	Maximum Values	Mean	Standard Deviation
Rainfall (mm)	0.0	71.8	7.4	13.63488
Wind Speed (m/s)	0.23	1.73	0.7	0.23655
Sunshine (hr)	0.00	11.20	6.29	2.559912
Temperature (^{0}C)	25.77	35.68	30.37	1.086057
pH	4.23	9.17	7.12	1.086057
Dissolved Oxygen (mg/l)	2.17	11.81	7.04	1.034054
Secchi (m)	0.00	2.40	1.08	0.513018
Turbidity(NTU)	0.00	660.00	17.1	31.68181
Conductivity (uS/cm)	35	286	96.24	26.78943
TPO_4 (mg/l)	0.00	11.60	0.05	0.464765
NH_3-N (mg/l)	0.00	13.20	0.11	0.515626
NO_3-N (mg/l)	0.00	9.94	1.21	0.949132
Biochemical Oxygen Demand (mg/l)	1	20	2	1
Chemical Oxygen Demand (mg/l)	1	79	16	11
Total Chlorophyta (cell/ml)	1	537	57	56.06997

3.2 Limnological Database

Table 3.3 shows data from Year 2001 until 2006 collected biweekly from Putrajaya Lakes and Wetlands had been segmented into 7 parts accordingly to the design of multi-cell approach in Putrajaya lakes and wetlands zonation in the data visualization system. For all the 7 parts of Putrajaya Lakes and Wetlands, 60% of the data sets had been used for training using HEA in order to obtain the rule sets for prediction model, and 40% of the data sets had been categorized according to SOM into low, medium and high and used for ROC true positive testing for percentage of accuracy.

Table 3.3 : Segmentation and Grouping of Putrajaya Lakes and Wetlands Substation, Total Data Sets for each of the Segments and Seperation for Testing and Training Data Sets

Putrajaya Lakes and Wetlands Substations	Data Visualization Segmentation	Total Data Sets	Training Data Sets	Testing Data Sets
PLa, PLb, PLc, PLd, PLe,PLf, and PLg	Putrajaya Lakes	1040	624	416
CW	Central Wetland	79	47	32
LE1 and LE2	Lower East Wetlands	78	47	31
UB1 and UB2	Upper Bisa Wetlands	116	70	46
UE1, UE2, UE3, UE4,UE5, UE6, and UE7	Upper East Wetlands	33	20	13
UN1, UN2, UN3, UN4, UN5, UN6, UN7 and UN8	Upper North Wetlands	113	68	45
UW1, UW2 and UW3	Upper West Wetlands	82	49	33

3.3 Data Analysis

Figure 3.2 shows usage of SOM in this study for classifying and comparison purpose. Chlorophyta abundance had been group into Low, Medium and High according to SOM analysis. Initial result from testing data sets and predicted result obtained from HEA are grouped according to their receiver operating characteristic before they are compared.

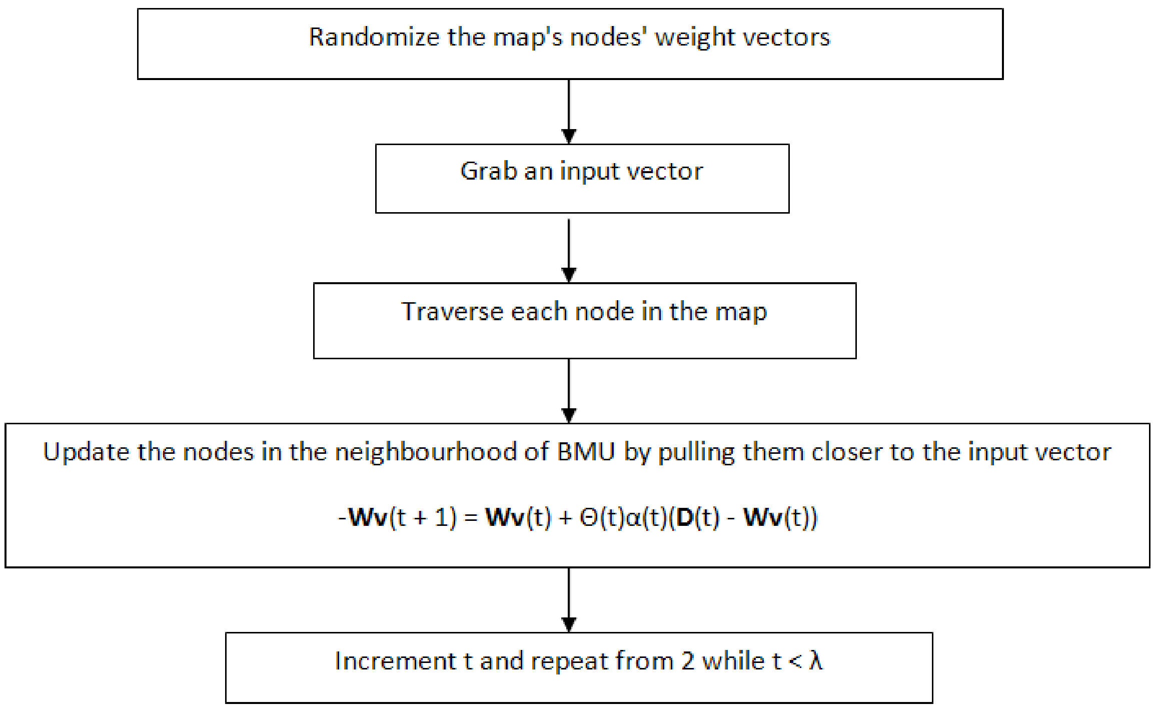

Figure 3.2: Processes involve in Self Organizing Map

Main advantage of using Self Organizing Map is they are very easy to understand.It’s very simple, if they are close together and there is grey connecting them, then they are similar. If there is a black ravine between them, then they are different. Besides than showing individual map, Self Organizing Map shows relationship between each of the parameters by µ-matrix and clusters map, classify data well and they are easily evaluate for their own quality so you can actually calculated how good a map is and how strong the similarities between objects are. Figure 3.3 shows the SOM of Putrajaya lakes and wetlands creating using MATLAB programming and Table 3.4 shows grouping and classifying of Chlorophyta abundance.

Data from dataset A is utilized to train the SOM that generates the component planes and cluster map. Putrajaya Lake trophic status is an oligotrophic lake, which is defined as having low productivity. The level of the water quality is well controlled as the diversity of the species is high with low number of individual Chlorophyta. This limits the possible categorization of Chlorophyta abundanceinto three ranges of categories only. The variables threshold range for each category is determined from the component planes of each variable are generated from SOM training. Threshold between less than 70 cells/ml, between 71 to 100 cells/ml, and higher than 100cells/ml were set for Chlorophyta biomass. True positive value of each extracted rule is calculated to determine the strength of the rule. Extracted rules are tested again with data Set A which is the training data. A different dataset which is not used for SOM training (namely dataset B) was used to test the effectiveness of the rule based system mainly to avoid producing biased testing results.

In this study, percentage of accuracy was calculated from receiver operating characteristic (ROC) curve graphs. ROC curve is a graphical plot of sensitivity or true positive rate versus false positive rate. In order to plot the ROC curve the Chlorophyta abundance was scaled according to the SOM analysis:low, medium, and high. The percentage of accuracy was then calculated to determine model performance using confusion matrix. Thresh-hold values of percentage of accuracy are ranges from 0% to

100%, where a score > 90% indicates outstanding discrimination, a score between 80-90% is excellent, and a score > 70% is acceptable.

Table 3.4: Total Chlorophyta Grouping

Range of Grouping:
0<= Low <70
70<= Medium < 100
100<=High

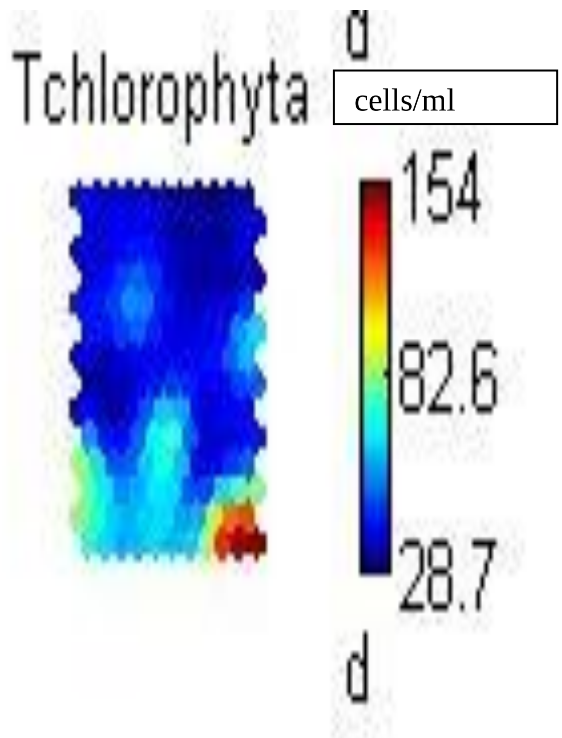

Figure 3.3: Self Organizing Map Obtained from Putrajaya Lakes and Wetlands Data from 2001 – 2006

3.4 System Design and development

3.4.1 System Overview

Figure 3.4 below illustrates the flowchart of data visualizationsystem. Ecological data will be uploaded into super computer after data analysis process. With execution of HEA program, high processing power of super computer will run 50cycles. On each cycles it will create an evolution of 100 generations of rules set. Out of all rules set obtained by super computer from hybrid evolutionary algorithm, 5 best rule sets will be evaluate and only one rule sets will be selected. The Selected rule sets then will be integrated into data visualizationsystem as the processor of the prediction model in the system. Regional data of Putrajaya Lakes and Wetlands will be inserted and appended into the data visualization to create map inside the system. Ecological database will then retrieve and injected into the system and will bond along with the Regional data that which they belong to. data visualizationsystem will populate a thematic map with Putrajaya Lakes and Wetlands, and Total Chlorophyta Abundance. The system development part are divided into two stages: 1) Application of HEA for rule sets generation and 2) thematic map development which is described in detail in section data visualization system Design and Development.

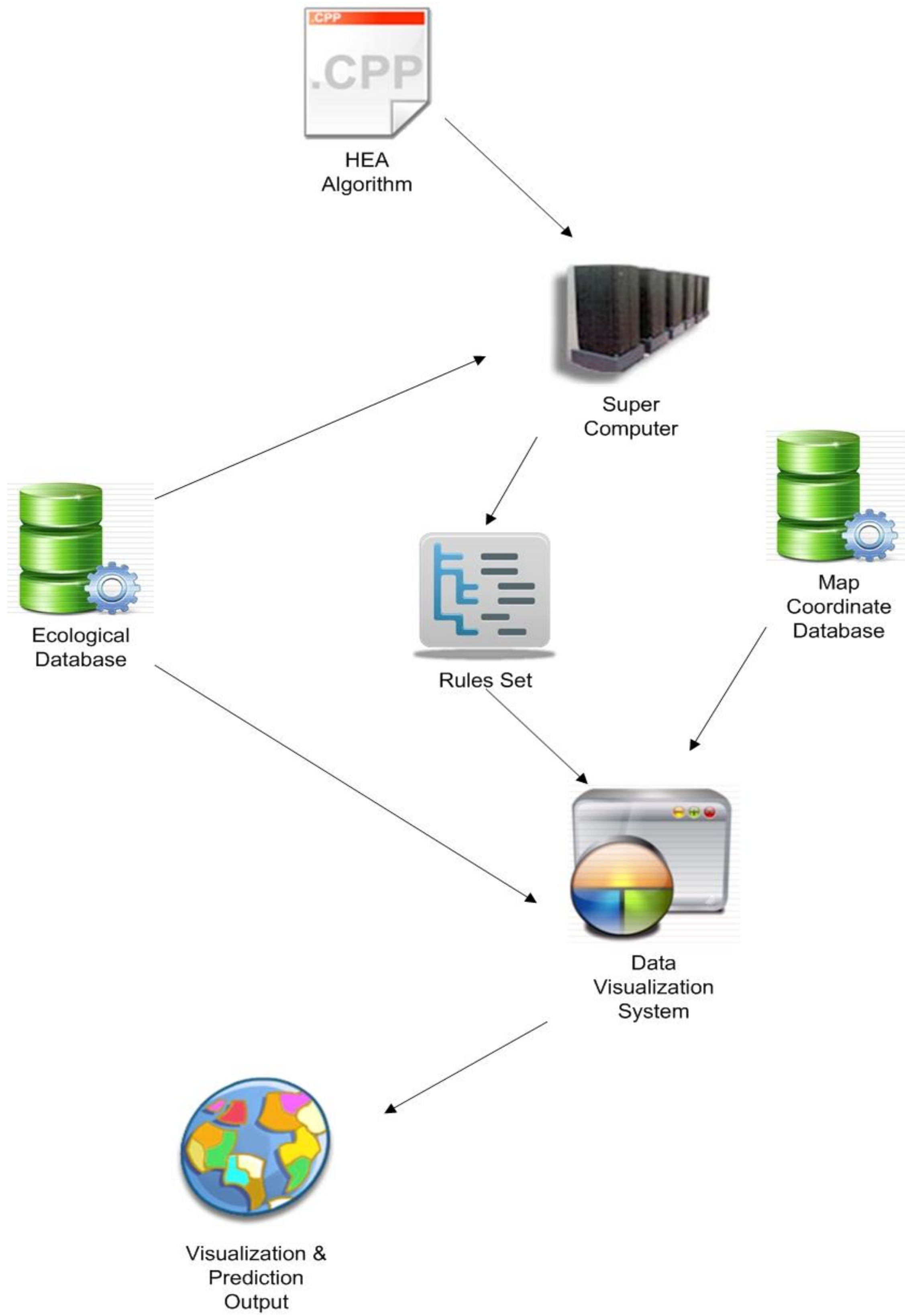

Figure 3.4 : System Development Flow Chart

3.4.2 System Requirement for Software

Table 3.5 shows minimum and recommended system requirement for software running in system development. hybrid evolutionary algorithms program requires high processing power of hardware and Linux based platform. It is running using Super Computer in Information Technology Center in University of Malaya.

Data visualization System requires operating system – Windows Xp or newer generation, running on Dotnet Framework 2.0. Hardware requirements are shown in Table 3.5.

Table 3.5 :HardwareRequirements for Software used in System Developement

Type	Minimum	Recommended	Remark
Processor	1.0 GHz	1.5 Ghz	
Memory	256 megabytes(Mb)	512 megabytes(Mb)	
Hard Disk Space	2 megabytes(Mb)	20 megabytes(Mb)	
Display	800 x 600 High colour	1024 x 768 True Colour	32 bits or 64 bits

3.5 Data Visualization SystemDesign and Development

3.5.1 Thematic Map Development

In this research, data visualization approach will combine thematic cartography with volume visualization. Volume visualization is a set of techniques which present and show an object without mathematical representing the other surface (Rosenblum, 1994). In this research, thematic map been selected to build in the system because of its features of representing particular theme, that is Chlorophyta abundance in Putrajaya Lakes and Wetlands.

Out of four properties, the main focus and concern in this research are area and distance. Equidistant projections had been selected. Figure 3.5 shows map of Putrajaya wetlands been used as skeleton of thematic map for data visualization. Figure 3.6 shows map of Putrajaya Lakes used as skeleton of thematic map for data visualization.

The actual Map of Putrajaya Wetlands in Figure 3.5 had been segmented into 7 parts. Grouping is based on different parts of wetlands which has been identify containing similar characters and Putrjaya lakes in Figure 3.6 had been grouped into one. The names of Lake and Wetlands group inside the system are UW represents Upper West Wetlands, UN represents Upper North Wetlands, LE represents Lower East Wetlands, CW represents Central Wetlands, UE represents Upper East Wetlands, UB represents Upper Bisa and PL represents Putrajaya lakes.Figure 3.7 is the Map sample from the system with different colour showing different parts of wetlands and Lake.Figure 3.7 later on will be used as the main part of data visualization system. Before running the visualization with time, different parts of lake and wetlands will be filled by different colour. When running visualization, different range of data will be representing by different colour and will be shown along with time line.

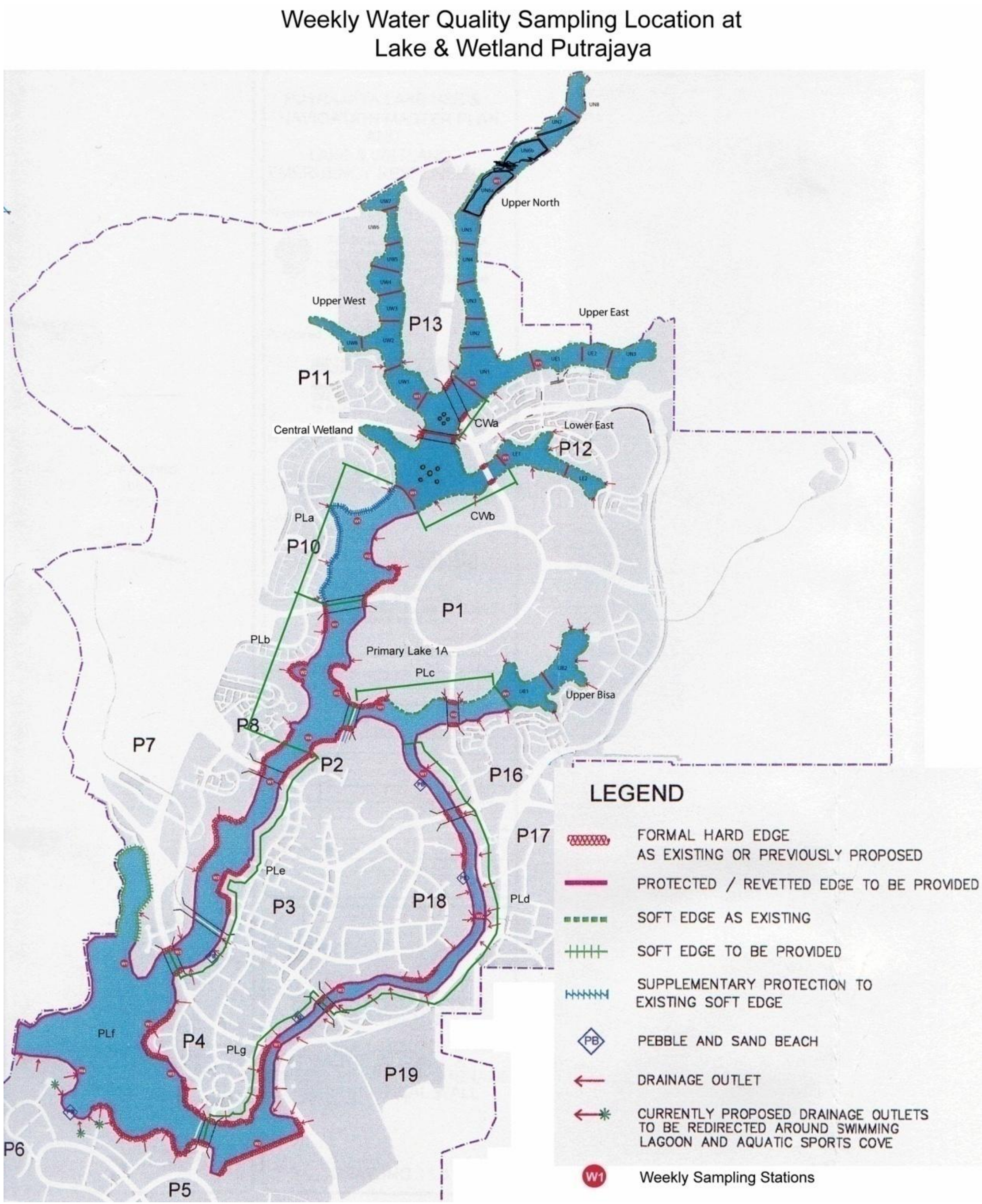

Figure 3.5 : Putrajaya Lakes and Wetlands map, and Water Quality Sampling Stations.

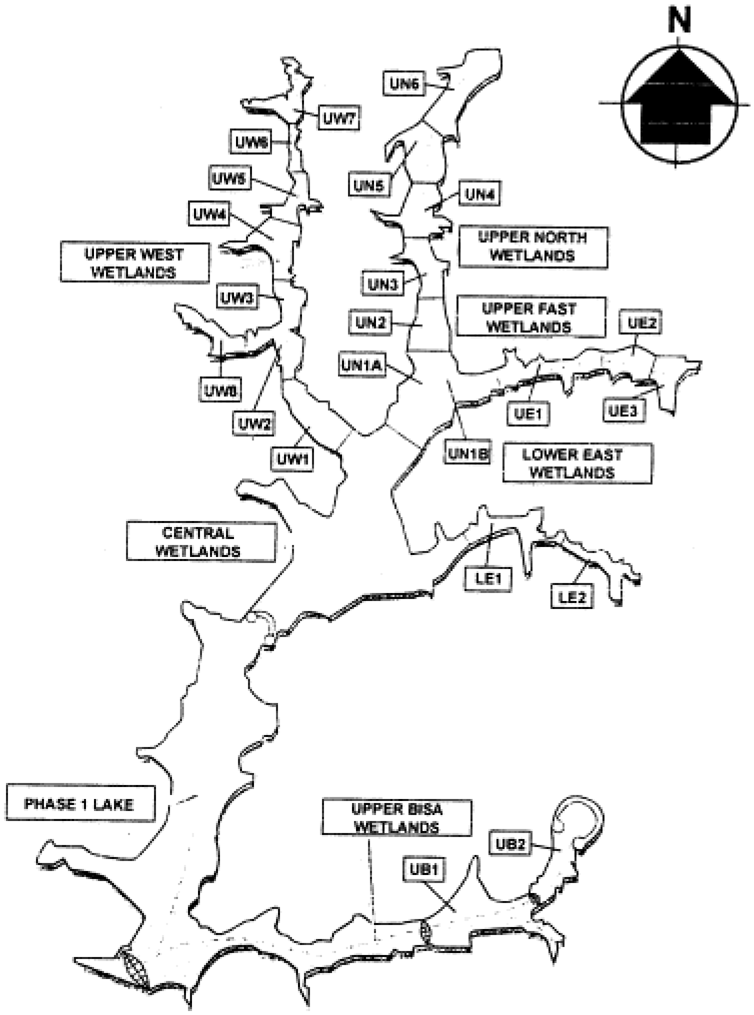

Figure3.6 : Additional Putrajaya Lakes Map (Ariffin, 1998)

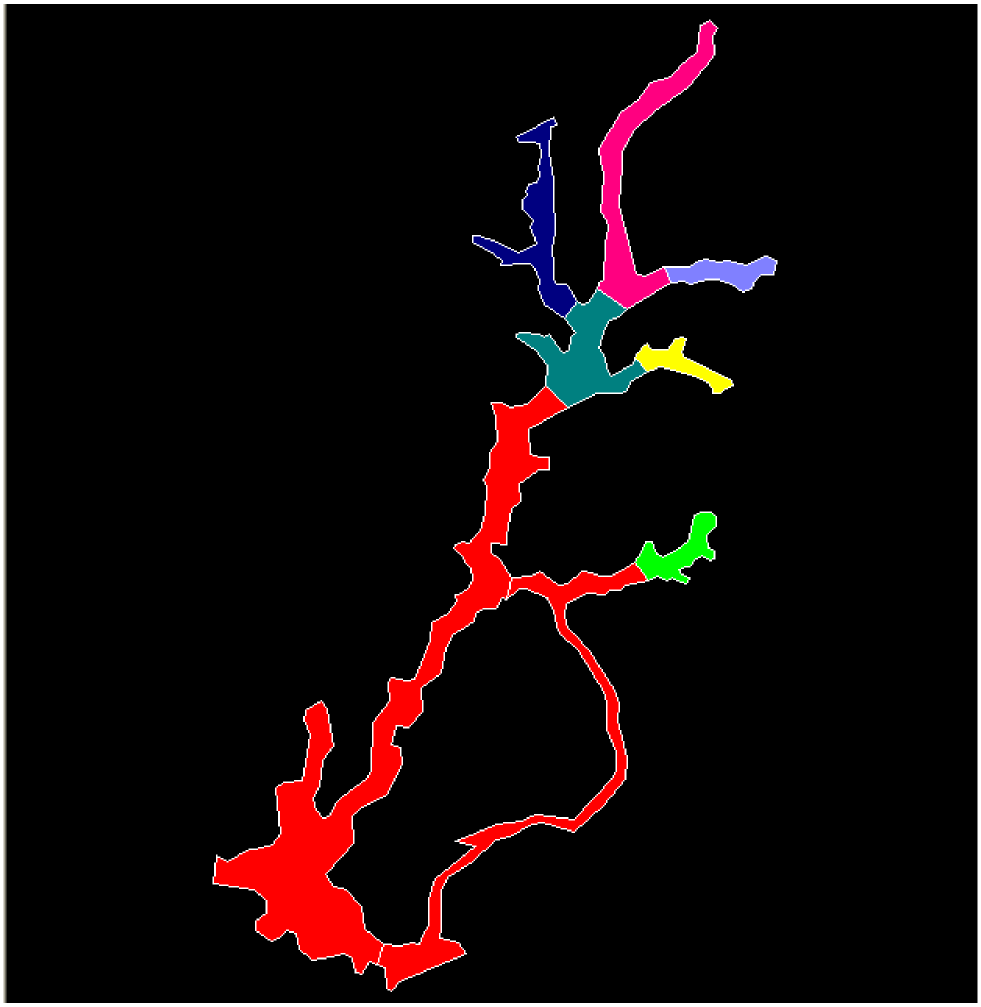

Figure 3.7: Map of Putrajaya Lakes and Wetlands in Data Visualization System combining Figure 3.5 and Figure 3.6

3.5.2 Hybrid Evolutionary Algorithm

Figure 3.8 shows the detailed algorithm for the rule discovery andparameteroptimization by HEA. The HEA program used in this study adopted from Cao et. al. (2006).HEA uses genetic programming (GP) to generate andoptimize the structure of rule sets and a genetic algorithm(GA) to optimize the parameters of a rule set. GPis an extension of GA in which the genetic populationconsists of computer programs of varying sizes andshapes. In standard GP, computer programs can be representedas parse trees, where a branch node represents anelement from a function set (arithmetic operators, logicoperators, elementary functions of at least one argument),and a leaf node represents an element from a terminal set(variables, constants, and functions of no arguments).These symbolic programs are subsequentlyevaluatedby means of ‘fitness cases’. Fitter programs are selectedfor recombination to create the next generation by usinggenetic operators, such as crossover and mutation. Thisstep is iterated for consecutive generations until thetermination criterion of the run has been satisfied. Ageneral GA is used to optimize the random parameters in the rule set.

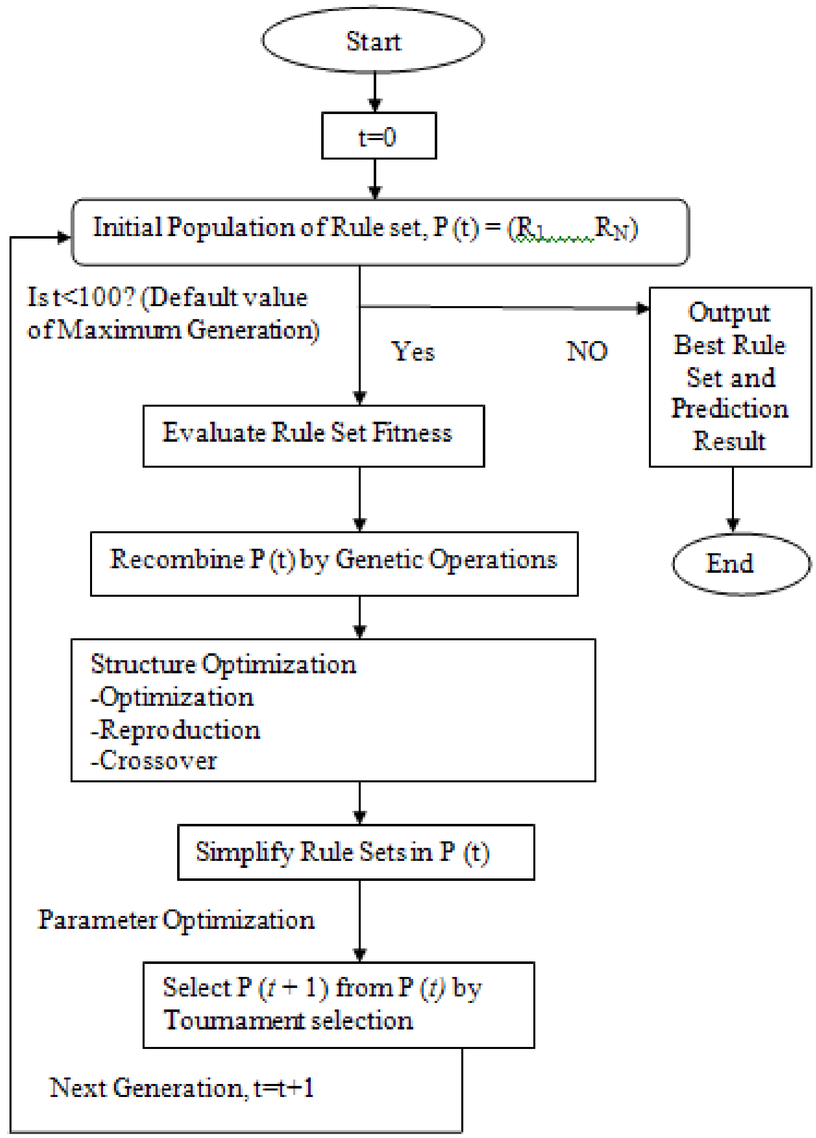

Figure3.8 : Flow chart of the hybrid evolutionary algorithm HEA for predictive modeling of time-series (from Cao *et al*. 2006)

50 out of 10000 best rules (100 Generations out of 100 Chromosome) will be select and study for based on their accuracy. Data predicted will be put under test and study. Only best rule set will be selected for each part of the wetlands and lakes, and will be programmed into data visualization phase as back-end engine. Cao *et. al*. (2006).

3.5.3 Data Arrangement for Hybrid Evolutionary Algorithm

Data collected and arranged into Microsoft Excel Format (.xlsx) will be converted into text (.txt) form. Figure 3.9 show data for Putrajaya lakes which has been prepare for hybrid evolutionary algorithm. The data parameters used in this study are from table 3.2.

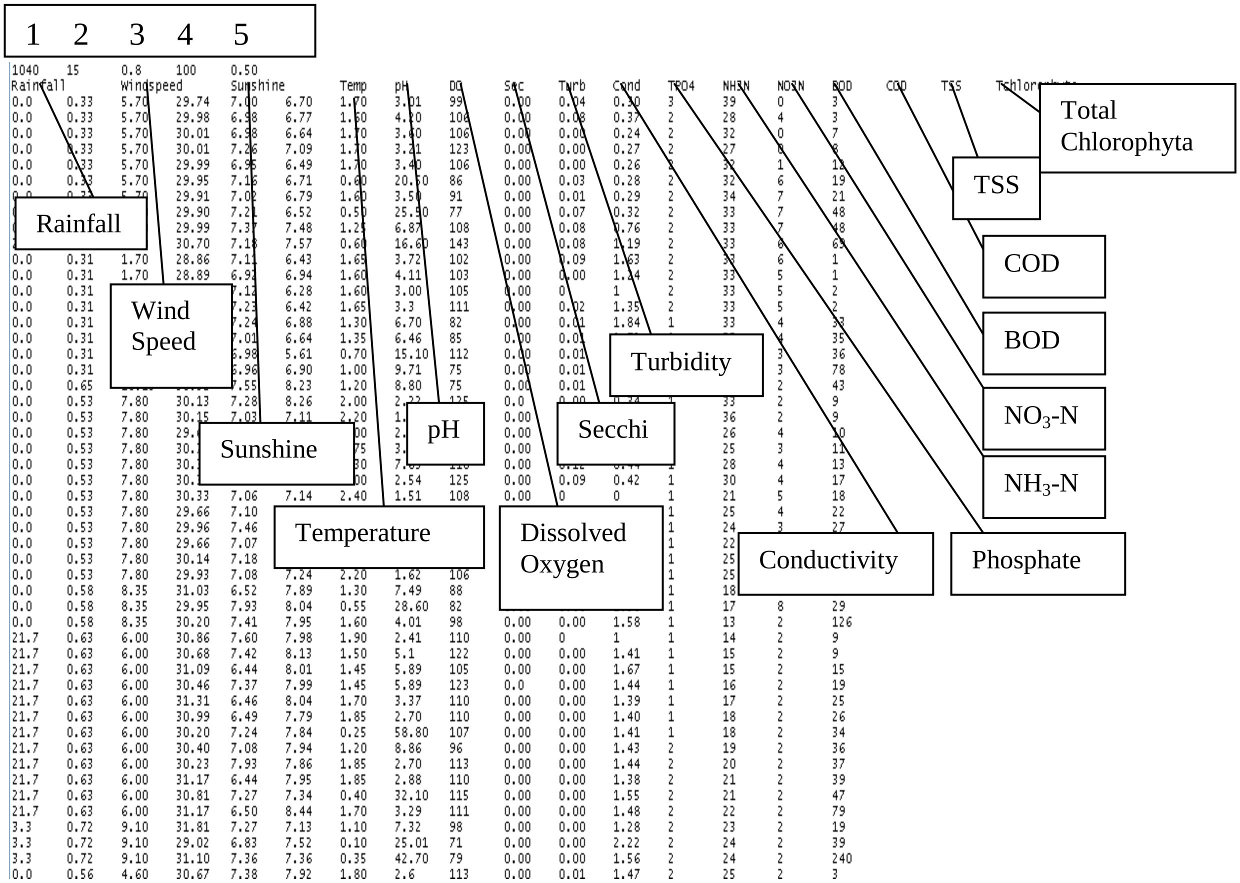

Figure 3.9: Sample Data Sets for Hybrid Evolutionary Algorithms

As refer to figure 3.9, the first row shows setting of hybrid evolutionary algorithm command parameters. In order from left to right, those values represent 1) Total amount of data, 2) Total input variables without counting the output which is Total Chlorophyta, 3) Percentage of Training data sets, 4) Numbers of runs of evolution and 5) Numbers of rules selected. From the figure above, hybrid evolutionary algorithm is running with total amount of 1040 data sets, 15 variables are involve. List of 15 variables are rainfall, windspeed, sunshine, temperature, pH, dissolved oxygen, Secchi, turbidity, conductivity, total phosphorus, NH_3-N, NO_3-N, biochemical oxygen demand, chemical oxygen demand and total suspended solids. The output is Total Chlorophyta. The third value shows the percentage of data use for training which 0.8 means 80%. The fourth value shows that these data will undergo 100 runs of evolution which each run contain the evolutions of 100 generations. The last values show that 50 best rules set will be selected for further evaluation based on RMS values.

3.5.4 User interface design and development

The system user interface consists of five main parts as illustrated in Figure 3.10. The explanations for the numbered parts of the system are as follows: Number 1 -Menu Bar, Number 2 - Thematic Map, Number 3 – Region List and Control Buttons, Number 4 – Coordinate List and Number 5 - Time line.

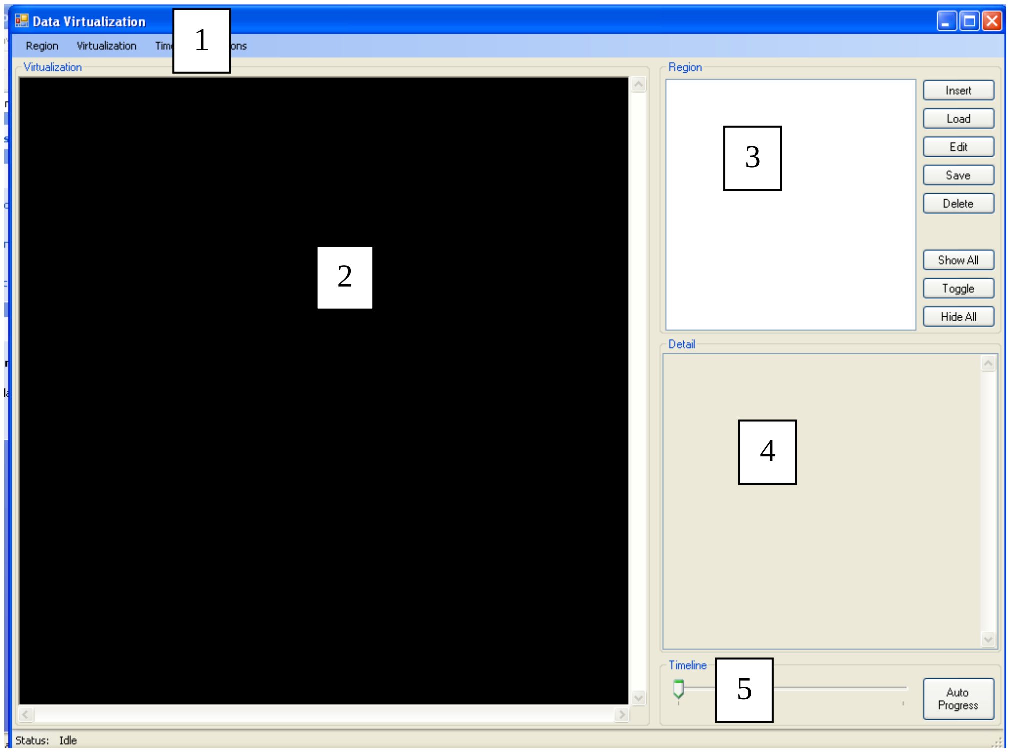

Figure3.10 : Design Layout of Data Visualization System

Referring to Figure 3.10, Number 1 shows the menu of the system. Inside the Menu, The Region menu bar contain inset, load, edit, save, delete and exit. Insert buttonis used to create new profile of the lake part include name, coordinate, representing colour, rule set and data sets. Loadbutton is used to loadthe existing lake system file. Edit button is used to edit profile of the selected lake part. Save button is used to save current work and convert it into a specific format which can be load and use later on. Deletebutton is used to delete a selected lake profile from your system. Exit button is to exit the system.

The Visualization menu bar contains controls of thematic map representation. The timeline menu bar contains controls of the running timeline for data visualization. The Options menu bar contains the basic main function in data visualization system.

Number 2 shows the Area for the system to plot and visualize the map or certain regions which had been selected to visualize. Number 3 show a blank area and 8 buttons on the side. After loading the system file, the system will load the map and shows the regions inside the map in this blank area. The insert button allows user to add new region into the current map. The Load button allows user to load a certain region data into the current map. Other controllers such as edit, save and delete were use to modify, store and delete a particular region. Another 3 buttons below are design to ease user to control the selection of visualization for the map such as show all, toggle and hide all.

Number 4 contain a grey area. The grey area is the place to shows the coordinate of regions being selected.The bottom of number 4 which is number 5 is the area which controls the time line of visualization. User can use the auto progress button and the system will automatically show the output. The timeline also can be adjusted manually by dragging the arrow on the time line.

3.6 Data Visualization System Development

Data visualization System is divided into twomodules: Prediction Module and Visualization Module.The rules generated using HEA are integrated with thematic map for Chlorophyta abundance visualization over time. Thematic map used is thematic cartography combine with volume visualization. By using volume visualization, different colours will be used to represent different range of data. There will not be any mathematical representation.Quick decision and respond can rapidily be made based on the displayed result.

3.6.1 Overall System Review

Refering to overall system flowchart in Figure 3.11, there are three types of information to be uploaded. The first is information on a selected region (regional data) to be uploaded that is the coordinates and default color setting as example red colour been used to represent Putrajaya Lakes and light green to represent Upper Bisa Wetlands for thematic map construction.The coordinate of certain part of Putrajaya Lakes and Wetlands were obtained from the map of Putrajaya Lakes and Wetlands. The second type of information involves uploading of required input parameters data and the ecological database of the selected region for Chlorophyta abundance prediction. Finally the HEA rule sets generated for Chlorophyta prediction is uploaded. Validation check in terms of data format and consistency is performed at each step to ensure integrity of information uploaded.

The first process in the system is to get data of a specific region.The process separates into two parts according to the two modules of the system. The first module which is the prediction module takes two types of data which are the ecological database and rule set generated from HEA. The visualization module takes only regional data. After uploading process, system will perform integrity check on both types of data. On the ecological database part, the system will go through extra two processes: first to extract all the element from the rule set and second to predict the result. Further details of the processes involved will be explained in prediction module part. For regional data, the system will undergo an integrity check such as the matching of coordinates and formation of a map. In this process, the system will passed the selected output list from the prediction module to attach with the regional data for the ease of grouping and matching the colour for the data range.

The updated process will come next for the global minimum and maximum value of both regional data and ecological database. Optimization process will take place to enhance the performance of the system. The system then will seek for the next region. Every region will repeat the same processes before visualization take place. If there is no next region exists, the system will visualize data from the regions that desired to be displayed.

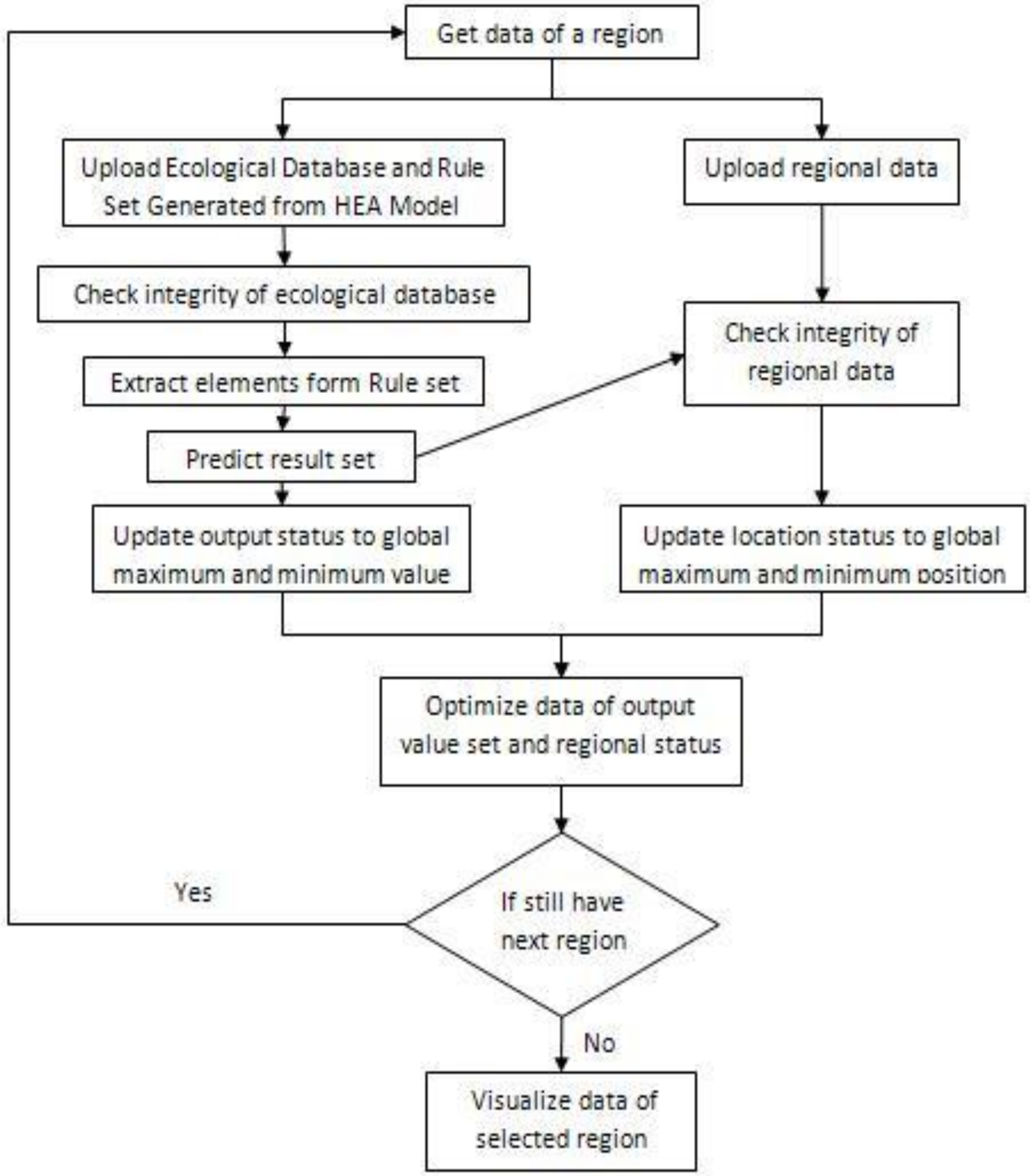

Figure 3.11 :System Flowchart

3.6.2 Prediction Module

Prior to Chlorophyta abundance prediction by means of rule sets generated by HEA, the rules are required to be analyzed by the system. The whole process of this module is illustrated in Figure 3.12. The uploaded rules are analyzed by extracting each element from the rule set such as the arithmetic operators, logical operators, variables and constants. Variables extracted from the rule sets are added into variable lists to indentify types of input parameters required from the data sets. Once the system finished extracting the required elements, the input data is uploaded to predict abundance of Chlorophyta. The predicted output in terms of threshold category of “Low”, “Medium” and “High” would be stored before displaying it over the constructed thematic map.

The first step of this module is to upload rule set generated by HEA and the get ecological database of a region. Half of this step involves of injecting rule sets generated by HEA to attach with the particular region and insert it into the data visualization. Another half of this step involves inserting the data set of the region to attach along with the rule sets. The system will then evaluate the expression of the rule sets. The rule set generated from HEA will be verified by matching with the elements from the rule set with the predefined operators and reserved keywordsin the system. The system will then determine the prerequisite of the operator or keyword.

After understanding and determine prerequisite of the operator and keyword, the system will split to verification step. In the verification step, the verification for the system will only concern of numerical expression, Boolean expression, assignment expression, and keyword. If the system found a numeric expression, the system will match it according to the variable that been previously stored or it is a new variable. If it is a new variable, the system will add it to the new variable list. The continue of the numeric expression is the same as when the system found Boolean, assignment expression or keyword, it will directly go to the next step which is searching for next expression from the rule sets. If yes, the system will repeat from the get next expression step again.

After the system gets all the expression from the rule sets, it will populate the variable list. The system will then come out with the option of choosing the output to be visualized. If predicted Total Chlorophyta been chosen, the system will emulate the function of prediction model to calculate the predicting output and then only it will populate output value list. If any variables or the original data sets of Total Chlorophyta had been chosen, the system will directly go to the final step of this module which is populating output list.

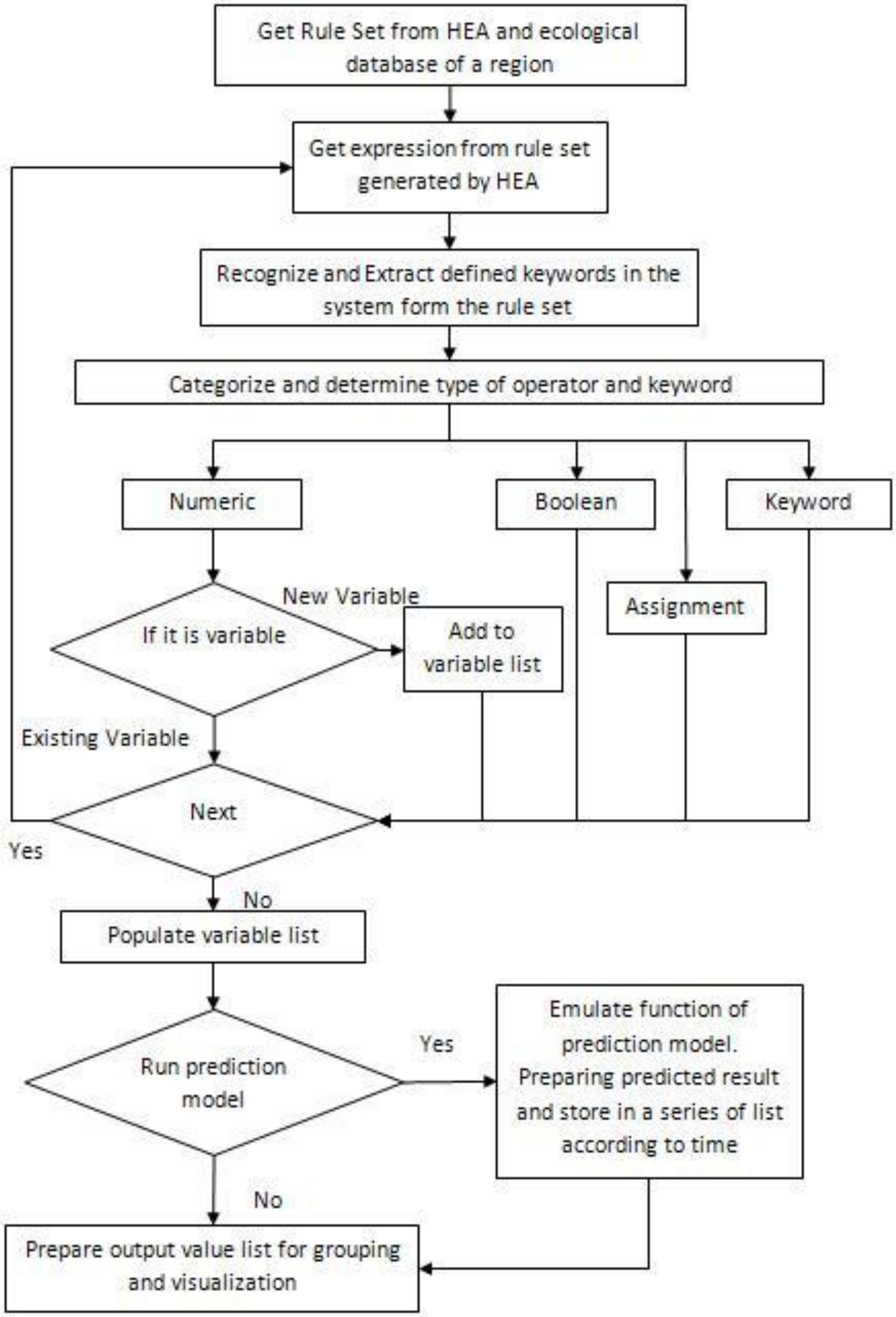

Figure 3.12 :PredictionModule

3.6.3 Visualization Module

Figure 3.13 show the visualization module. The final module involves visualization of predicted Chlorophyta Abundance over time. Thematic map are constructed using coordinates of the uploaded selected region. Two main processes are involved in visualization of Chlorophyta over time on constructed thematic map. Firstly the constructed map needs to be resized according to the preferences setting by the system user such as the window size and map ratio. Next the output of predicted Chlorophyta abundance is displayed iterating over time of the input dataset. The predicted abundance of Chlorophyta is showed by varying colors on the thematic map over time, where red is associated with high abundance, green is associated with medium abundance and yellow with low abundance.

The first step of this module is for the user to select regions of the map to be viewed. This step allows the system to retrieve the detail of regions selected for viewing. The subsequent step starts with two concurrent processes. The first process involves the retrieval from the prediction module of output values of regions according to time. Thre regions are coloured according to their output values. Data that are used in this process are linked with timeline and the system determines the date format and averages the data from the earliest to the latest data sets. The system then determined the colour to represent the output value. The system also provides colour display options for the user. The colour can be multi-contrasted like natural colour shifting gradually from green for low output value to red for high output value. The user can also choose just green, yellow, and red to represent respectively low, medium, and high to output values. The output data are then matched with the range of values to which it belongs. The second process involves the extraction of the region's coordinate to map it on the global map area of the system. The system then used outputs from both processes to draw and fill the selected regions with colours that represent the output values. The entire steps and processes are repeated for the remaining selected regions. Upon completion, the result of the entire selected area of the map is displayed for visualization. All these are summarized in Figure 3.13.

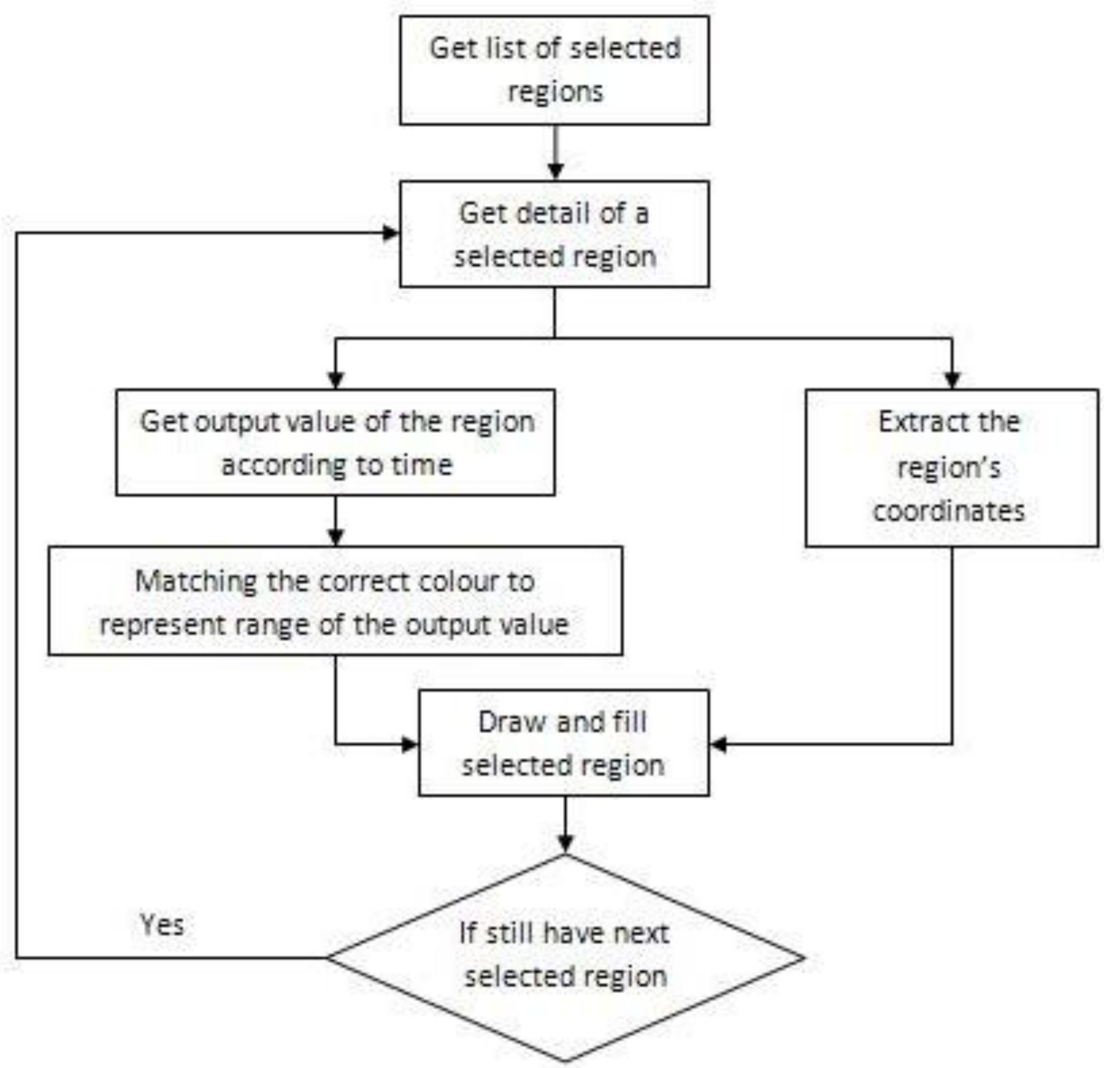

Figure 3.13: Visualization Module

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Results

This section contains detail information on application of the system to model Chlorophyta abundance at the selected study site Putarajaya Lake and Wetlands using data collected from year 2001 to 2006. The first part would be application of the system itself until visualization of the Cholorophyta abundance. The second part presents detail explanation of rules sets generated by HEA.

4.1.1 Data Visualization System (Thematic map and Application)

This section describes application of Chlorophyta abundance using thematic map visualization. User's guide to the system is described in Figure 4.1 to Figure 4.12.

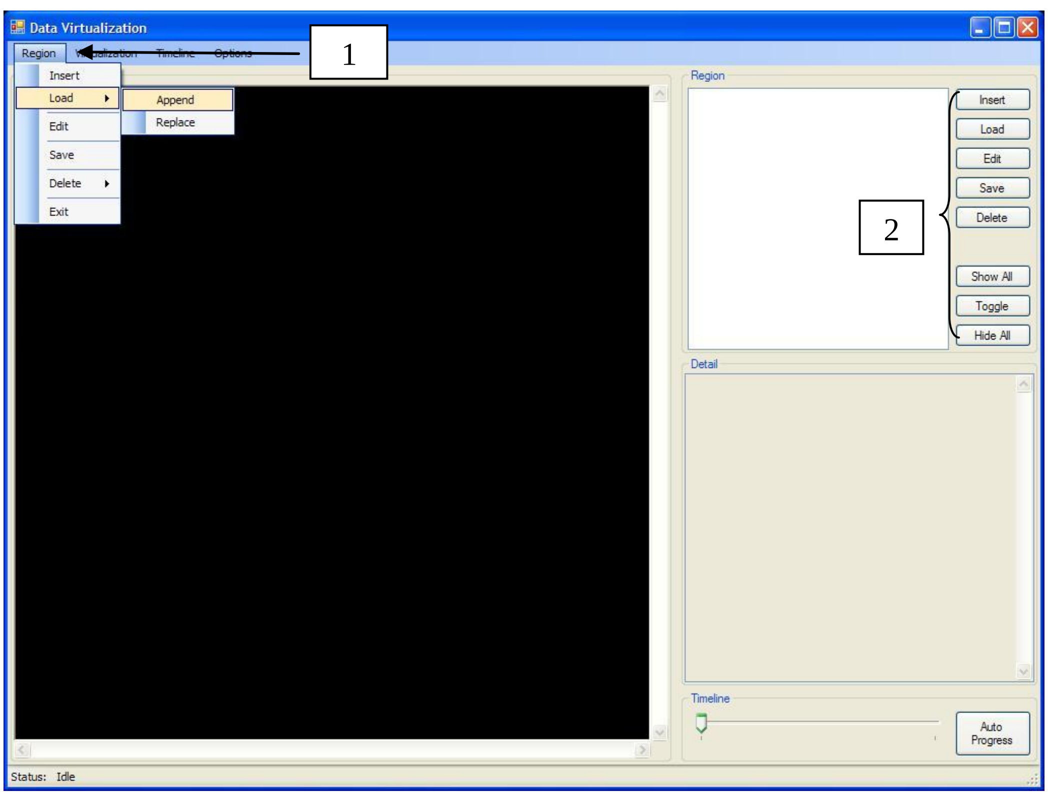

Figure 4.1 : Import Region Definition

Figure 4.1 represents the main graphic user interface (GUI) of the system. There are 4 items in the menu bar which are Region, Visualization, Timeline and Options. The Region menu which had been labeled as number 1 in Figure 4.1 contains Insert, Load, Edit, Save, Delete and Exit. The Number 2 in Figure 4.1On the right hand side, there are similar bottoms which act as the same function which contain in the Region menu.

The Insert function is used to create new lake profile. Lake profile is a file created by the data visualization system. It stored the coordinates, rules sets and data sets of different regions (parts) of the lake. There are two options for the Load function. Append option allows to add in new region into the current lake without replacing existing current region. The replace function will delete selected region from the existing current lake and replace with new region. Edit function allows modifying region coordinate, rule sets and data sets on a specific region. Save function allow to keep all works done. Delete function contains two options. It allows users to delete only selected region or the whole lake profile. Exit option uses to quit the program.

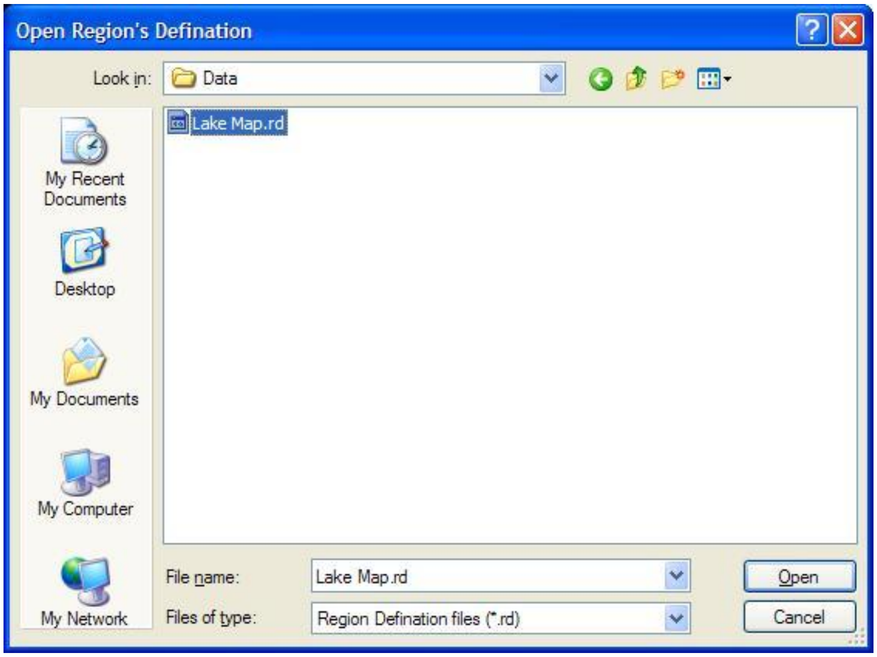

Figure 4.2 :Loading Lake Profile

Figure 4.2 show the loading of the Lake profile using the Load function. The Lake profile will be stored as an extension of region definition files (* .rd).

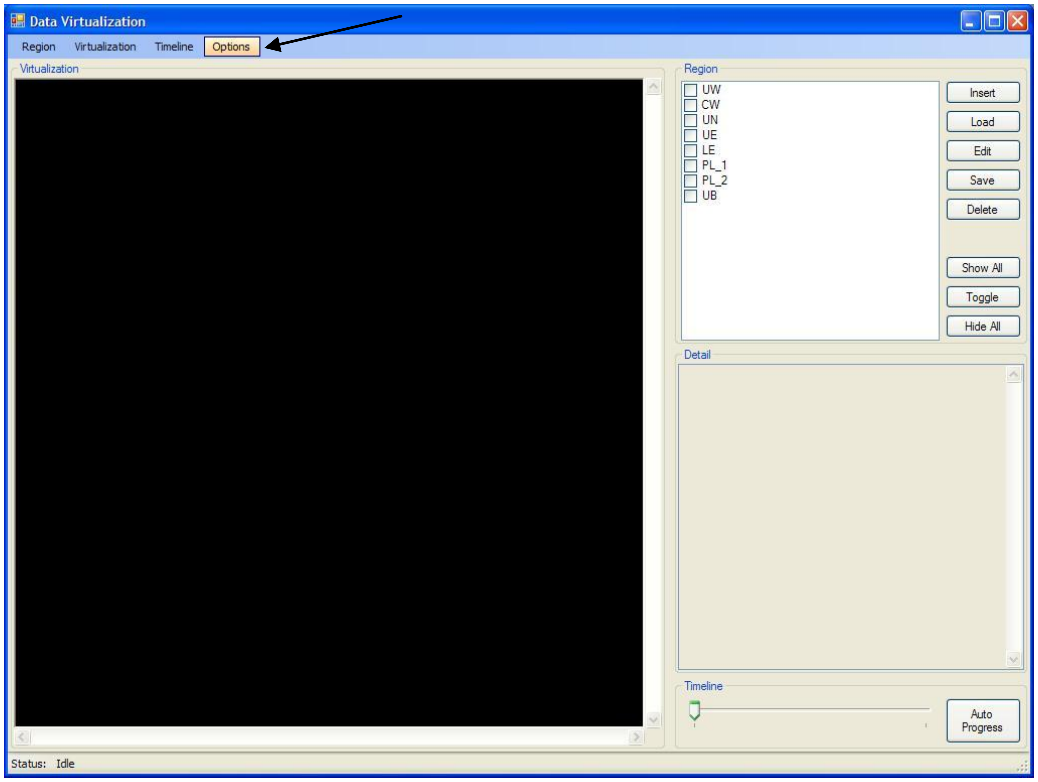

Figure 4.3 : Select Option

Figure 4.3 shows how the system responds after loading the lake profile. Lake Map display can be manipulated by selecting option menu. This will open up pop up window that allows changing of the map display. Due to the shape of Putrajaya Lakes, it has been split into two parts : PL_1 and PL_2. Both PL_1 and PL_2 contain similar data.

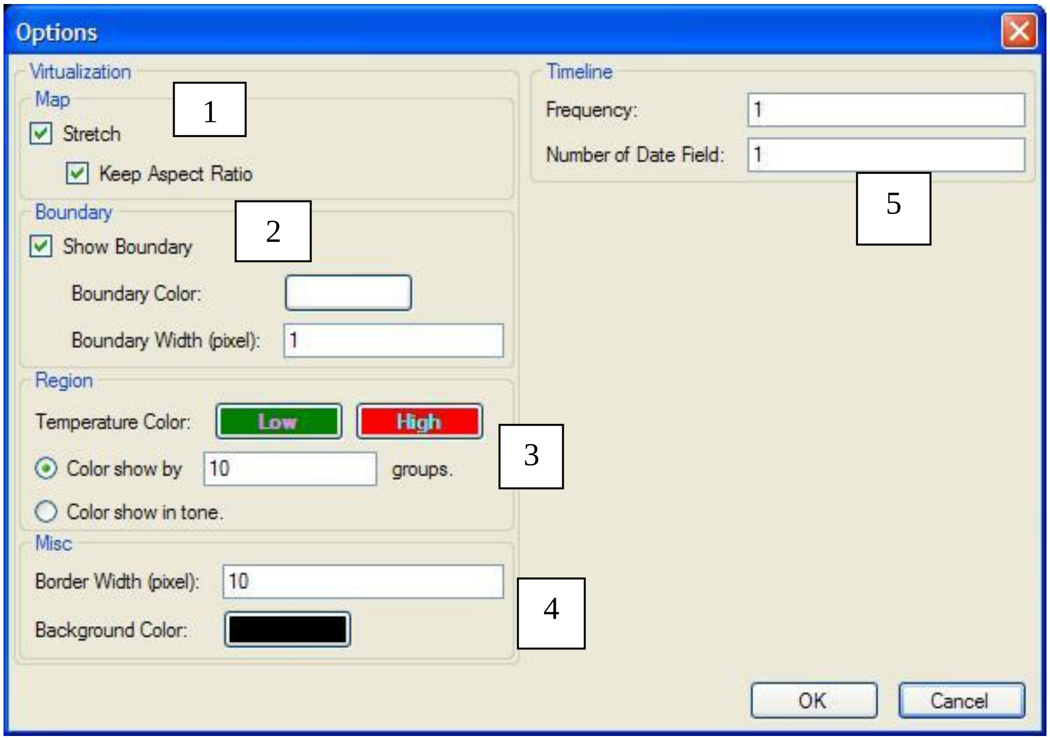

Figure 4.4 : Set Options for map display

Figure 4.4 shows the pop up window of the option menu. Number 1 is showing the map option. It allows the user to stretch the Lake map. Number 2 is the boundary option, it is usedto show boundary between each regions of the lake and change colour and width of the boundary. Number 3 is the region option. This option is uses to adjust colour to represent the Chlorophyta abundance intensity from low to high. The hue and saturation of the color can be manipulated as well. User can choose to set range of Chlorophyta abundance by setting colour that attach with it and also allow letting the system automatically separate the Chlorophyta abundance according to the colour tone from the computer. Number 4 is the Misc option and it allows user to control and setting the border width and the background colour.

Timeline option is shown on the number 5 in Figure 4.4. Frequency and number of date field text box in the timeline is used to set the frequency of displaying the date of data visualization map. The number of date field need to be set according to date field in Lake data set. If the date is set as 2.01.2005, then the number of date field has to set to 1; else if the date is set separately such as 2nd, January and 2005(date are separate to between Day, Month and Year), so the date field has to set to 3.

Figure 4.5 to Figure 4.10 show the user interface of Insert New Region and the steps to setup before the visualization take place. Figure 4.5 show the initialize screen of the insert new region. Every part is empty without any fill in.

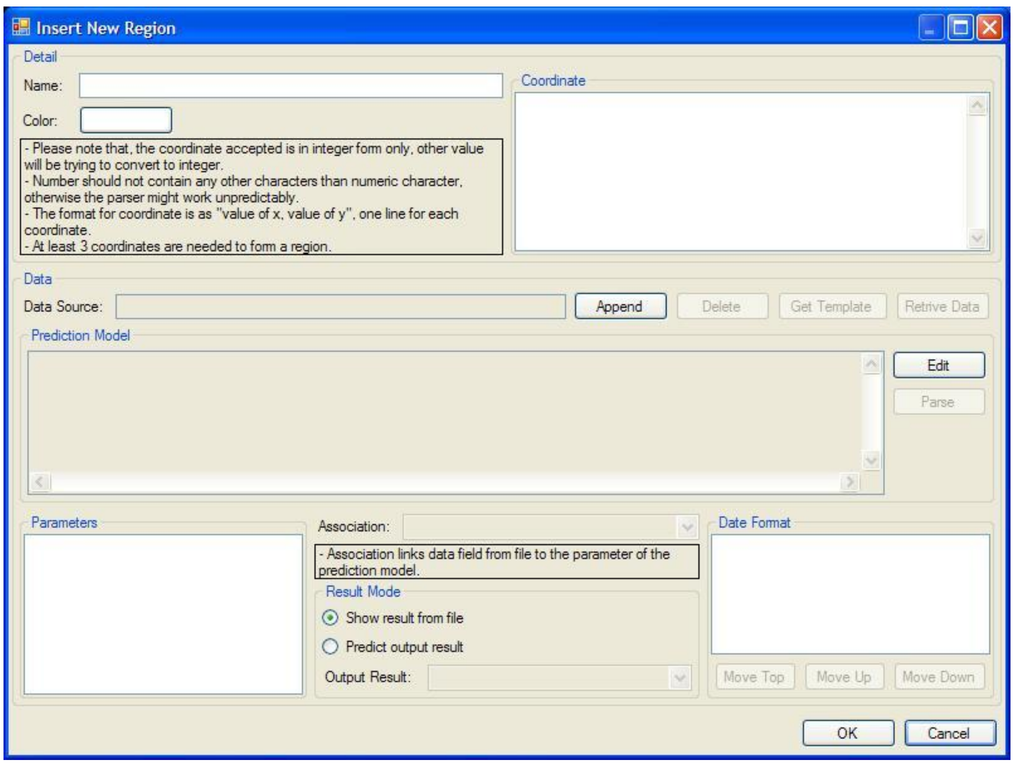

Figure 4.5 : Screen display of Insert New Region

Figure 4.6 shows naming and setting the colour to represent a certain region of the Lake. The specific region of the Lake shown in the figureis Upper West Wetlands and using blue colour to represent it. Figure 4.6 also shows insert the list of coordinates of the region on a particular lake.

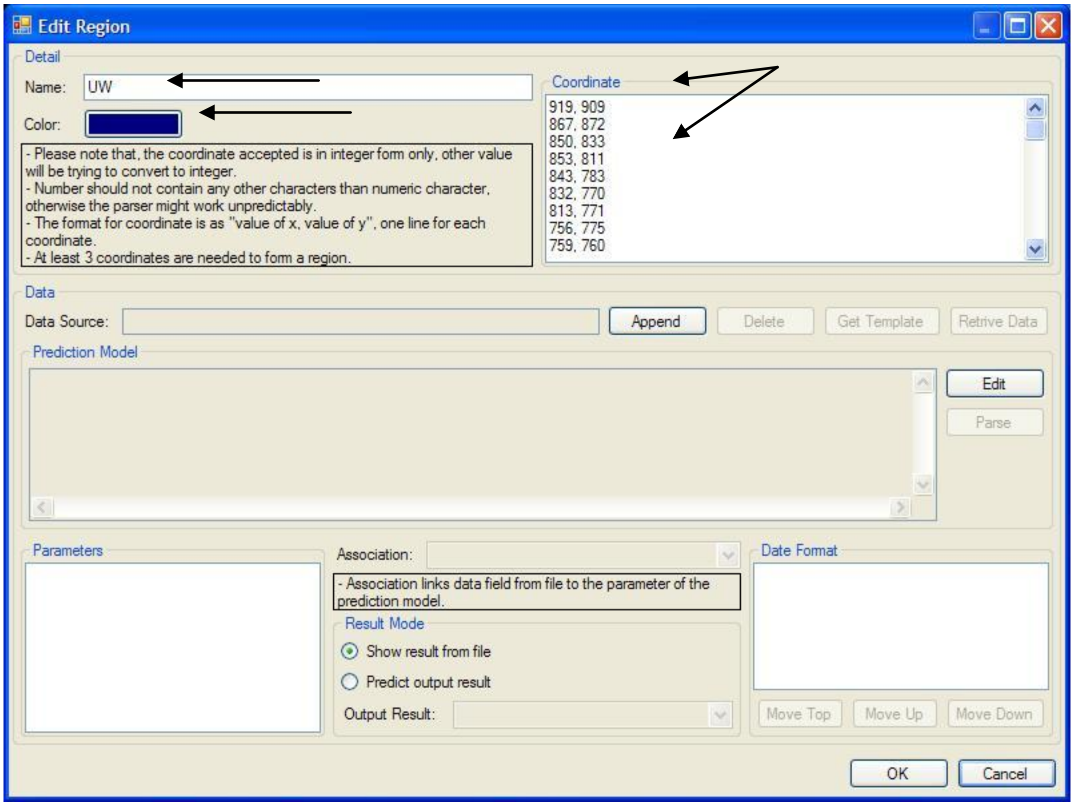

Figure 4.6 : Naming, Choosing Representation Colour and Setting Coordinates of a certain Region

Figure 4.7 shows insert the data source of this particular region. For the 1st time, only Append option will be shown. All the other options will be shown if the region already attach with data sets.

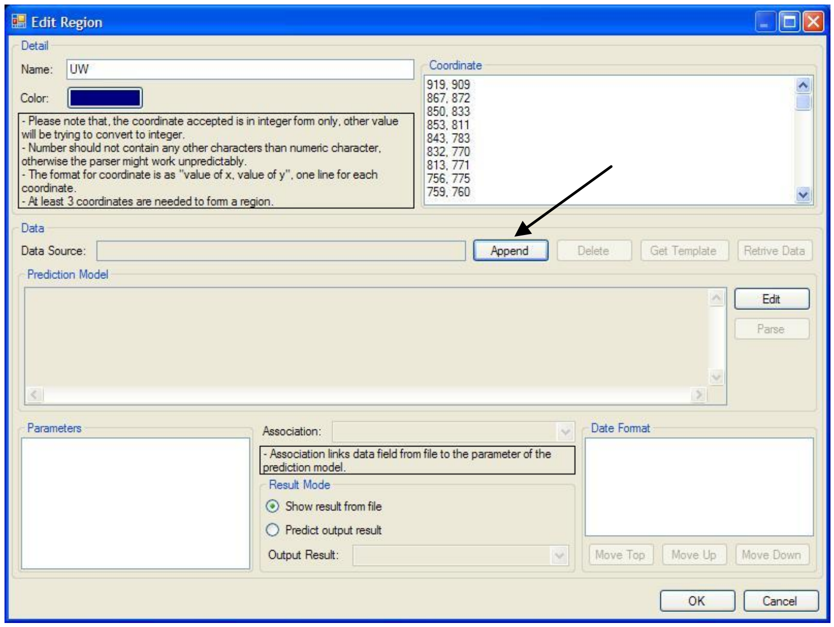

Figure 4.7 : Append data sets for a specific Region

Figure 4.8 shows the Prediction model. It allows users to insert the rule sets that generate for the particular region. The rule sets have to be written in C or C++ programming language. User can edit the rule sets or parse it to attach with the region. After insert the rule set, user has to use the parse button in order to send the parameters (variables) from the rule sets to the system. The Parameters column will display all the parameters which contain inside the rule sets. User has to associate the parameters of rule sets with parameters from the data sets by selecting from the Associate dropdown button.

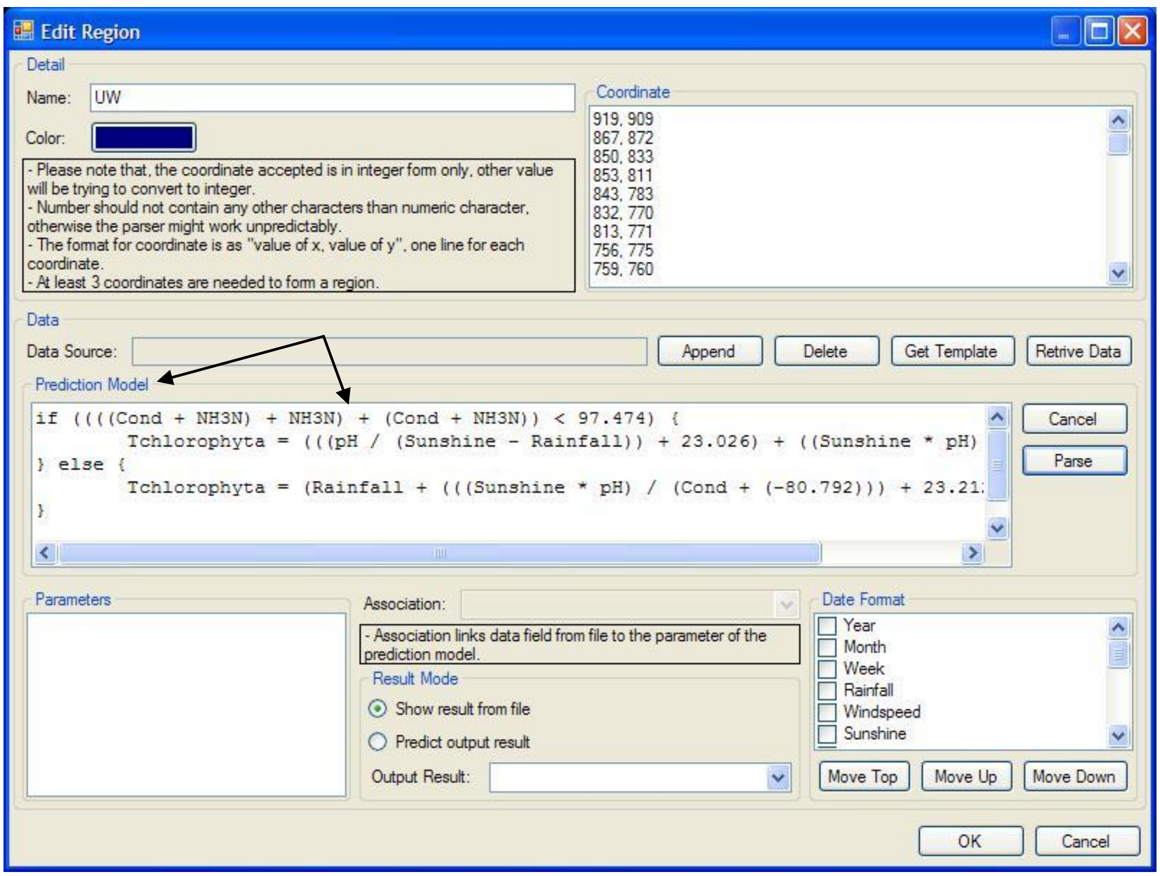

Figure 4.8 : Edit Prediction Model and Insert Rule Set

Figure 4.9 shows the result mode allows user to choose to visualize any parameter from the data sets, for example temperature, pH or total Chlorophyta. User can also use the system to predict and visualize the output by choosing the predict output result.

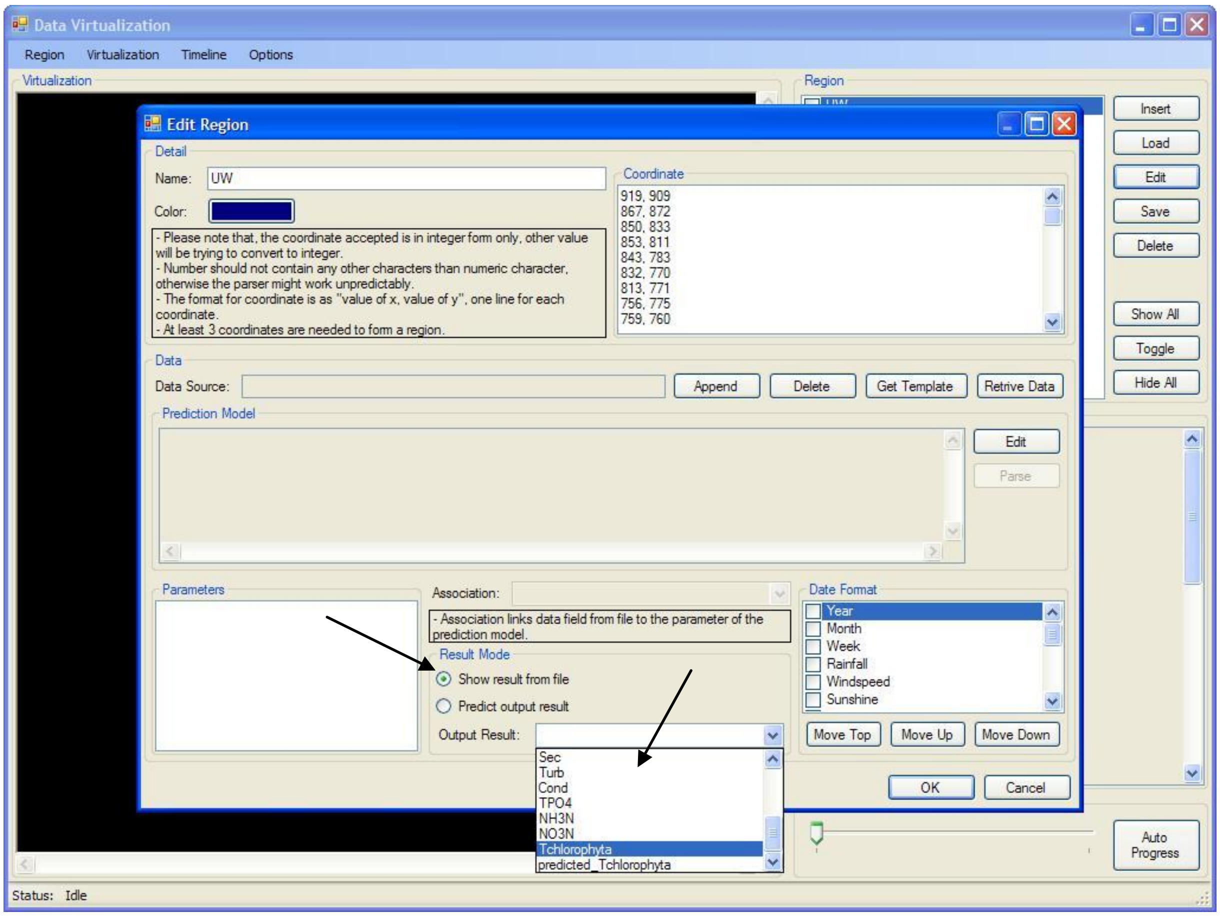

Figure 4.9 : Select Desired Output Data Field

Figure 4.10 shows the final step of inserting the region which is setting the date format. User has to choose the parameters that represent date. If it is more than one parameter then user has to choose and arrange it accordingly to days/weeks, months and years by using the moving up and down button.

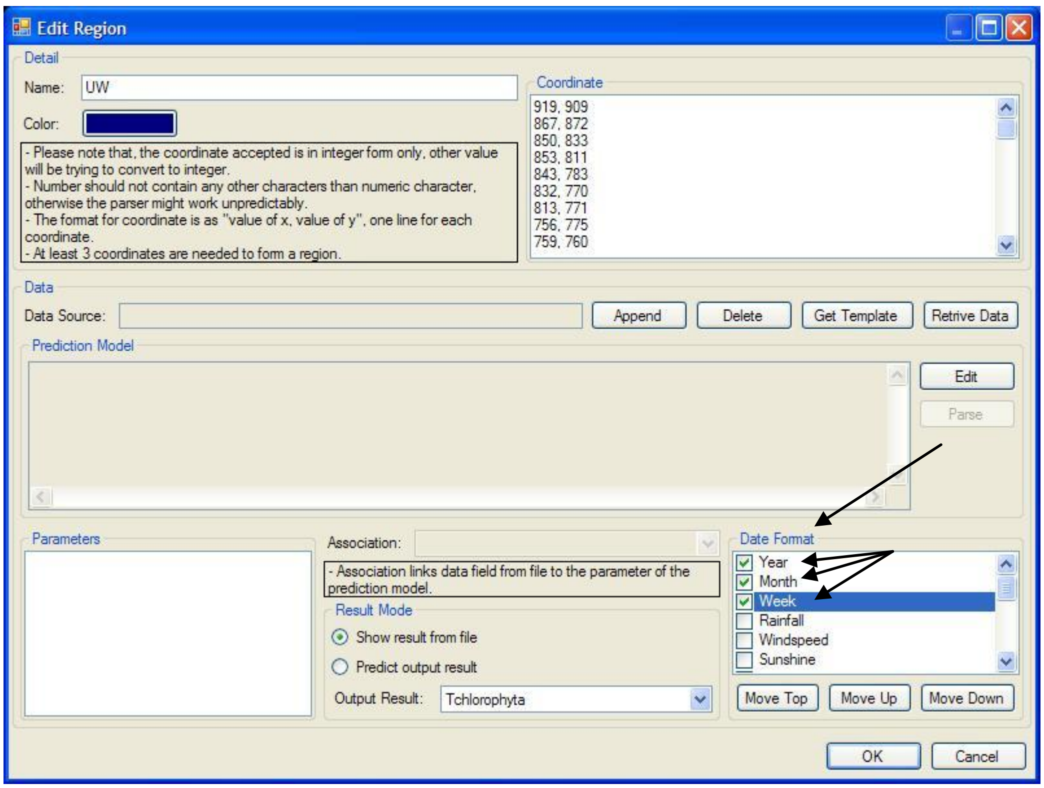

Figure 4.10 : Select Fields which indicates Date

Figure 4.11 shows the map of the lake by combining the entire regions. User can choose to hide one or more regions by double clicks uncheck the regions from the list as shown in number 1 in the figure. Some handy buttons had been added for the ease of user to show all the regions, hide all the regions and also toggle the current selection. When user click on a certain region, the coordinate of that region will appears on the grey area below the list in number 2.

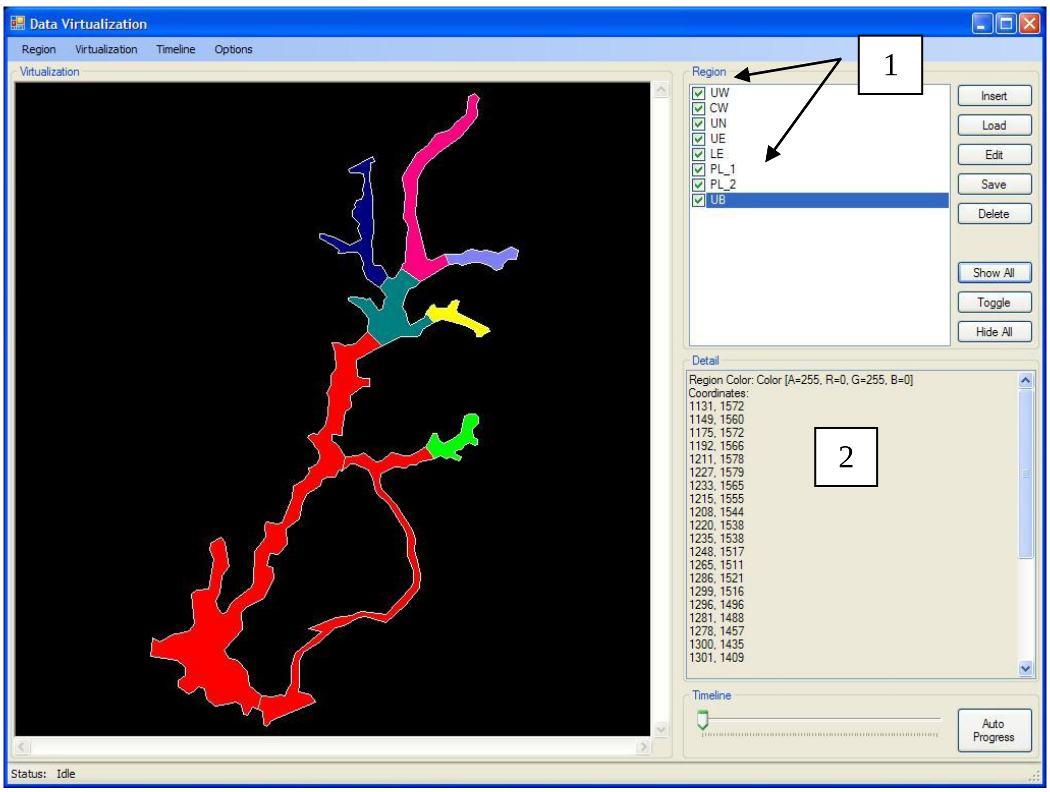

Figure 4.11 : Select Desired Output Region

Figure 4.12 shows the system running on prediction model and visualizing the data of the Putrajaya lakes and wetlands. Time line had been set by the system according to the amount of timeline contain in the data sets. For example if some of the selected region contains data from year 2001 to 2006 and others contain data from year 1998 to 2003, the system will set the timeline from 1998 to 2006. To display it daily, weekly, biweekly, monthly or yearly is all depends on how the data sets. The timeline can be drag or let it run automatically and the date will be shown on top of the program.

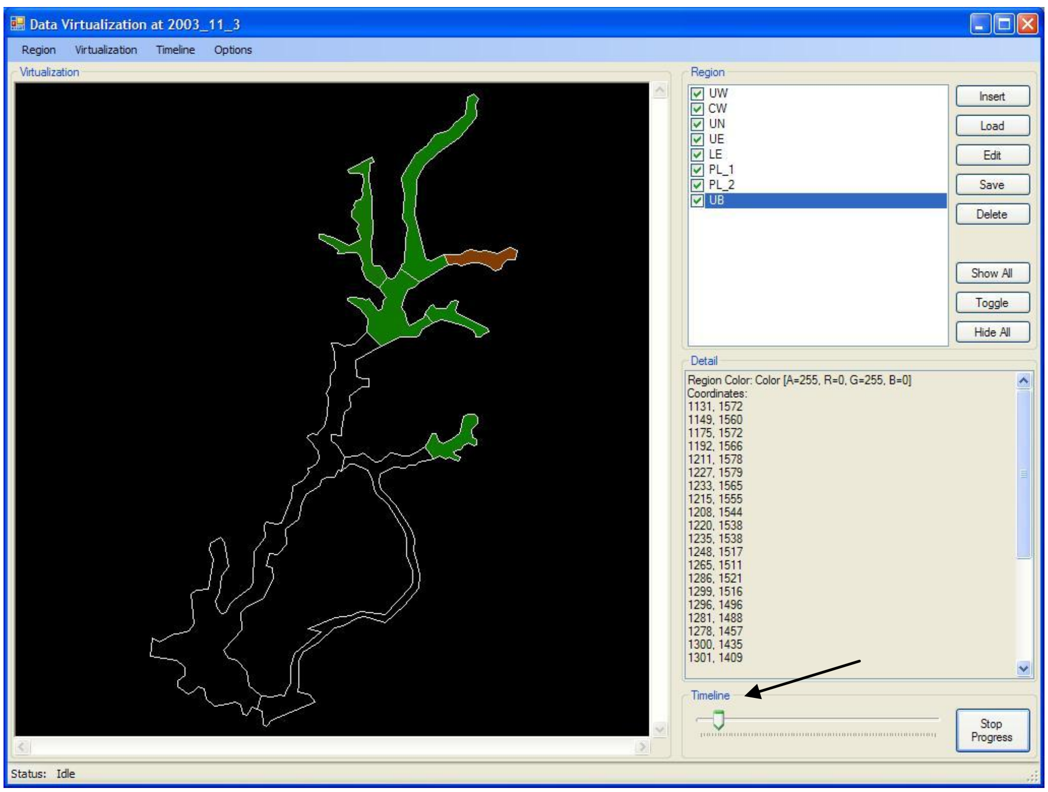

Figure 4.12 : Click the Auto Progress for Automatic Progression of Each Date Time

4.1.2 Rule sets and Results from Hybrid Evolutionary Algorithm

Table 4.1 shows that HEA rules have been generated as prediction model for data visualization system from all the seven parts of the lakes and wetlands from year 2001 to 2006. Each of the section contains one best rule set which the rule set can be divided into two brunches which are IF brunch and ELSE brunch. The percentage of accuracy of ROC true positive had been calculated for each of the part of the lakes and wetlands.

Table 4.1 : Rule Sets Generated from HEA and Percentage of Accuracy for Putrajaya Lakes and Wetlands

<table>
<tr><th rowspan="2">Putrajaya Lakes and Wetlands Sections</th><th colspan="3">Rule Sets</th><th rowspan="2">Percentage of Accuracy (%)</th></tr>
<tr><th>Condition Statement (IF)</th><th>IF Brunch (THEN)</th><th>Else Brunch (ELSE)</th></tr>
<tr><td>Putrajaya Lakes (PL)</td><td>(ln(|TPO4|)>=99.933)</td><td>Tchlorophyta=((NO3N*Windspeed)+(NO3N+39.627))</td><td>Tchlorophyta=((((NH3N*34.126)+22.975)/((NO3N*Turb)+pH))*((Temp-(Windspeed-COD))-((21.369-Sunshine)-(Sunshine/BOD))))</td><td>72.79</td></tr>
<tr><td>Central Wetlands (CW)</td><td>((((Sunshine*Sunshine)>=79.010)AND(Cond>=52.825))AND((DO*DO)>=49.065))</td><td>Tchlorophyta=(exp(NO3N)*50.846)</td><td>Tchlorophyta=(((NO3N/exp(NO3N))*45.347)-(((NO3N*(-10.179))/(149.747-Cond))*51.734))</td><td>91.14</td></tr>
<tr><td>Lower East Wetlands (LE)</td><td>(((158.687/Rainfall)>90.737)OR(((Cond+Sunshine)>=74.965)OR((Cond>16.596)AND(Rainfall>10.321))))</td><td>Tchlorophyta=(NH3N+(Temp+(-1.215)))</td><td>Tchlorophyta=((((Turb+234.147)+(-1.610))+(-1.610))+(-1.304))</td><td>93.59</td></tr>
<tr><td>Upper Bisa Wetlands (UB)</td><td>((Turb*428.965)>428.965)</td><td>Tchlorophyta=((85.916/((41.978/NO3N)-Cond))+28.769)</td><td>Tchlorophyta=226.054</td><td>96.55</td></tr>
<tr><td>Upper East Wetlands (UE)</td><td>((Sunshine<=9.583)AND((Rainfall>=8.487)OR(Rainfall<=6.549)))</td><td>Tchlorophyta=(exp(exp(Windspeed))+(exp(exp(Windspeed))+23.026))</td><td>Tchlorophyta=(Turb+236.425)</td><td>96.97</td></tr>
<tr><td>Upper North Wetlands (UN)</td><td>(((exp(NO3N)/TPO4)>19.590)OR((Sunshine/TPO4)<=86.085))</td><td>Tchlorophyta=((ln(|ln(|Turb|)|)+5.129)+(Cond/(exp(pH)/(Cond-(-144.302)))))</td><td>Tchlorophyta=(((32.515/NO3N)+121.739)+Turb)</td><td>84.96</td></tr>
<tr><td>Upper West Wetlands (UW)</td><td>((((Cond+NH3N)+NH3N)+(Cond+NH3N))<97.474)</td><td>Tchlorophyta=(((pH/(Sunshine-Rainfall))+23.026)+((Sunshine*pH)/exp(Cond)))</td><td>Tchlorophyta=(Rainfall+(((Sunshine*pH)/(Cond+(-80.792)))+23.212))</td><td>95.12</td></tr>
</table>

4.1.3 Results of Input Sensitivity and Discussion

Following best rule set generated from HEA, graphs had been plotted to study the sensitivity of the inputs elements against the abundance of total Chlorophyta.

- Putrajaya Lakes (PL)

For Putrajaya Lakes, the condition statement of rule application determines by threshold value of ln(TPO4). The rule indicates that IF ln(TPO4) is more than or equal to 99.933 THEN the Total Chlorophyta abundance is determined by the equation Tchlorophyta=((NO3N*Windspeed)+(NO3N+39.627)) ELSE by the equation Tchlorophyta=((((NH3N*34.126)+22.975)/((NO3N*Turb)+pH))*((Temp-(Windspeed-COD))-((21.369-Sunshine)-(Sunshine/BOD)))).

Figure 4.13 shows the input sensitivity of the IF brunch of Putrajaya Lakes (PL) section. IF rule distinguishes between conditions of low abundance with lower concentration of NO3-N (mg/l) and weaker wind speed, and conditions of high Chlorophyta growth reflected by higher concentration of NO3-N (mg/l) and stronger wind speed. Putrajaya Lake as the lake is categorized as an oligotrophic lake. The highest Concentration of TPO4 (mg/l) is 11.60. This extreme condition is not happening in Putrajaya Lakes.

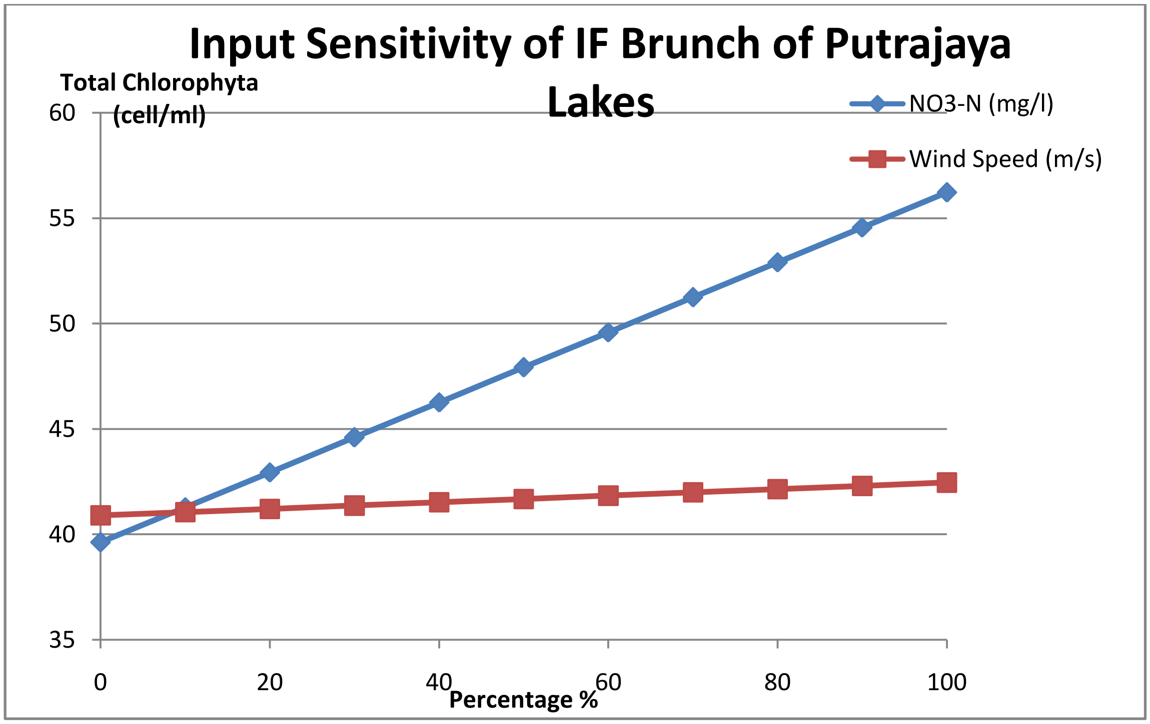

Figure 4.13 : Input Sensitivity of IF Brunch of Putrajaya Lakes, Total Chlorophyta against Input Range (%)

Figure 4.14 shows the input sensitivity of the ELSE brunch of Putrajaya Lakes (PL) section. ELSE rule shows that abundance of Chlorophyta is dependent on NH3-N, NO3-N, turbidity,pH, temperature, wind speed, chemical oxygen demand (COD), sunshine and biochemical oxygen demand (BOD). Among all these variables, total Chlorophyta are highly sensitive towards NH3-N, NO3-N, turbidity and COD. Others variables like pH, windspeed, sunshine and BOD show a very consistent value of Chlorophyta abundance.

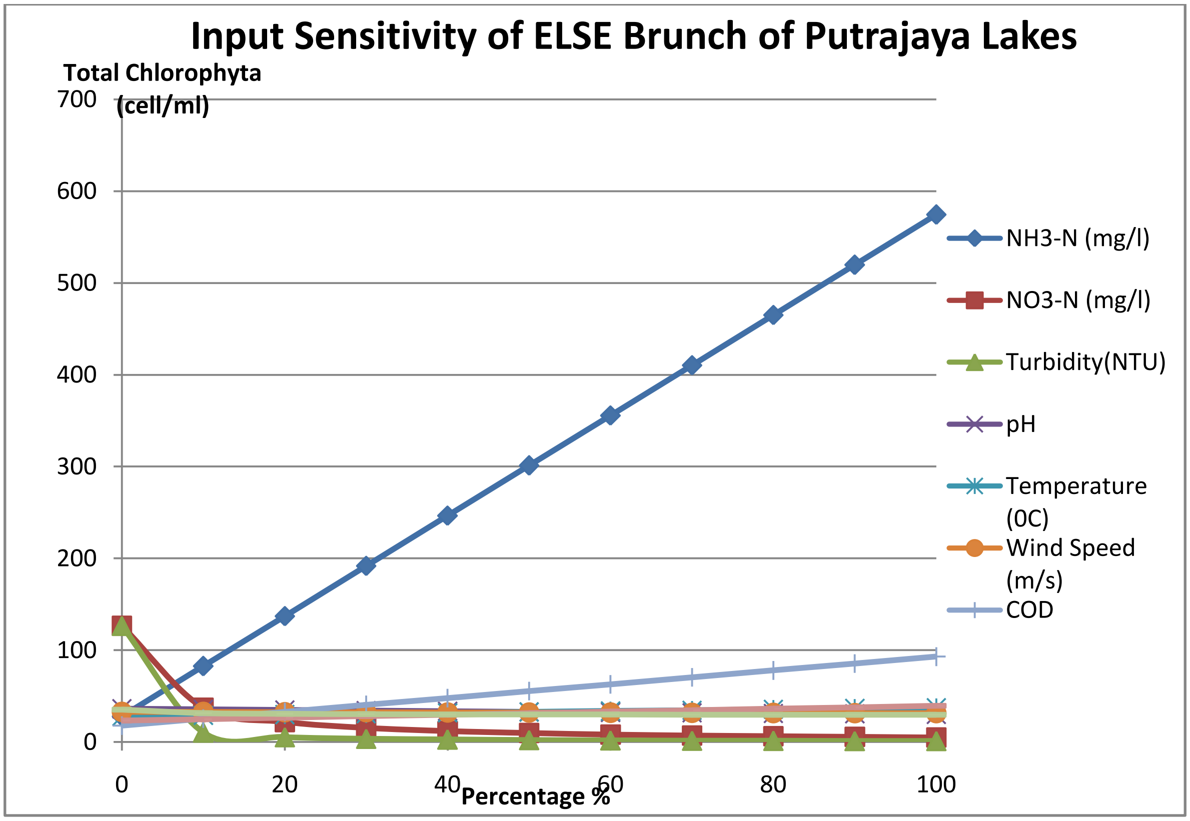

Figure 4.14: Input Sensitivity of ELSE Brunch of Putrajaya Lakes, Total Chlorophyta against Input Range (%)

- Central Wetlands (CW)

For central wetlands (CW), the condition statement of rule application determines by threshold value of sunshine, conductivity and dissolved oxygen. The rule indicates that IF sunshine squared is more than or equal to 79.010 AND conductivity is more than or equal to 52.825 AND dissolved oxygen squared is more than or equal to 49.065 THEN the Total Chlorophyta abundance is calculated by the equation Tchlorophyta=(exp(NO3N)*50.846) ELSE by the equation Tchlorophyta=(((NO3N/exp(NO3N))*45.347)-(((NO3N*(-10.179))/(149.747-Cond))*51.734)).

Figure 4.15 shows the input sensitivity of the IF brunch of Central Wetlands (CW) section. IF rule distinguishes between conditions of low Chlorophyta abundance with lower concentration of NO_3-N (mg/l), and conditions of high Chlorophyta abundance reflected by higher concentration of NO_3-N (mg/l).

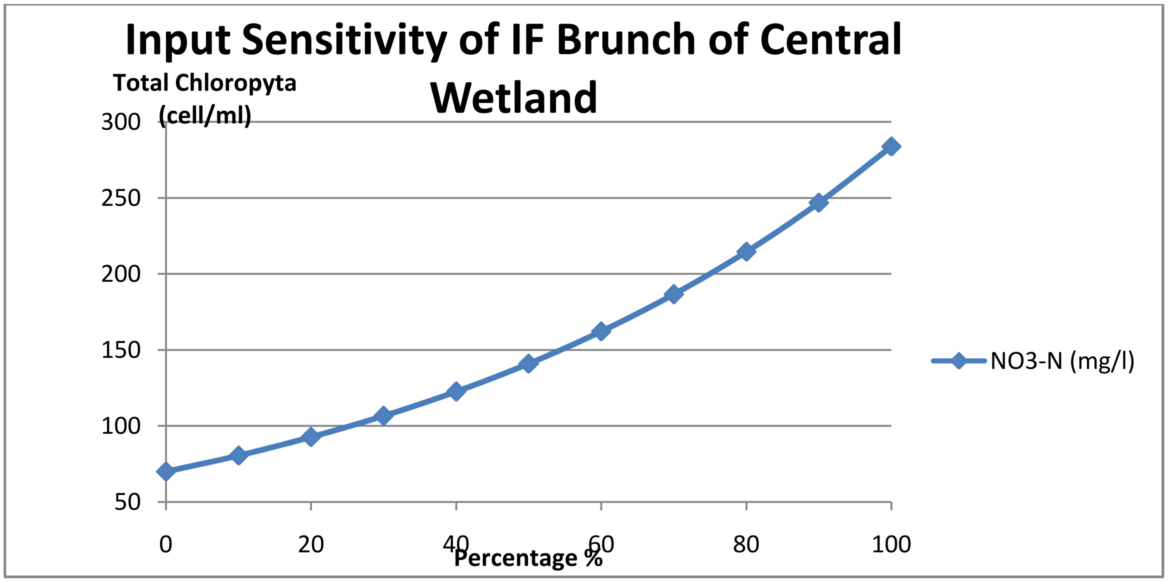

Figure 4.15 : Input Sensitivity of IF Brunch of Central Wetland, Total Chlorophyta against Input Range (%)

Figure 4.16 shows the input sensitivity of the ELSE brunch of Central Wetlands (CW) section. ELSE rule shows that abundance of Chlorophyta is dependent on NO3-N and conductivity. Result shows that Chlorophyta abundance increases when both conductivity and NO3-N increase.

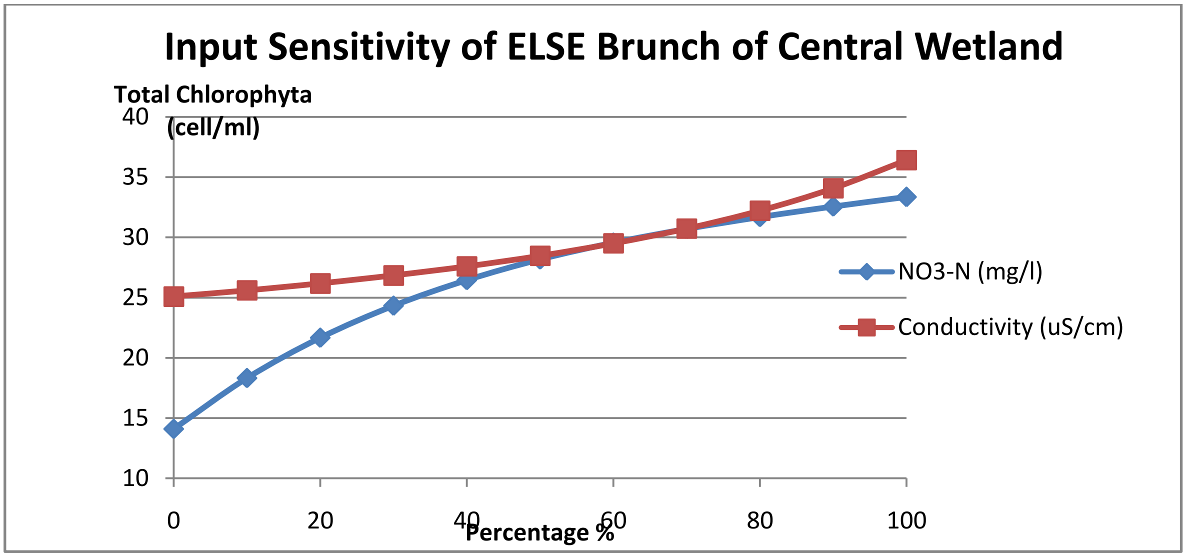

Figure 4.16 : Input Sensitivity of ELSE Brunch of Central Wetland, Total Chlorophyta against Input Range (%)

- Lower East Wetlands (LE)

For Lower East Wetlands, the condition statement of rule application determines by threshold value of rainfall, conductivity and sunshine. The rule indicates that IF rainfall is less than 1.75 OR conductivity plus sunshine more than or equal to 74.965 OR conductivity is more than 16.596 AND rainfall is more than 10.321 THEN the Total Chlorophyta abundance is calculated by the equation Tchlorophyta=(NH3N+(Temp+(-1.215))) ELSE by the equation Tchlorophyta=((((Turb+234.147)+(-1.610))+(-1.610))+(-1.304)).

Figure 4.17 shows the input sensitivity of the IF brunch of Lower East Wetlands (LE) section. IF rule show that Chlorophyta abundance is positively correlated to level of concentration of NH_3-N and temperature. This means that either one of temperature increase or level of concentration of NH_3-N increase, Chlorophyta abundance will increase.

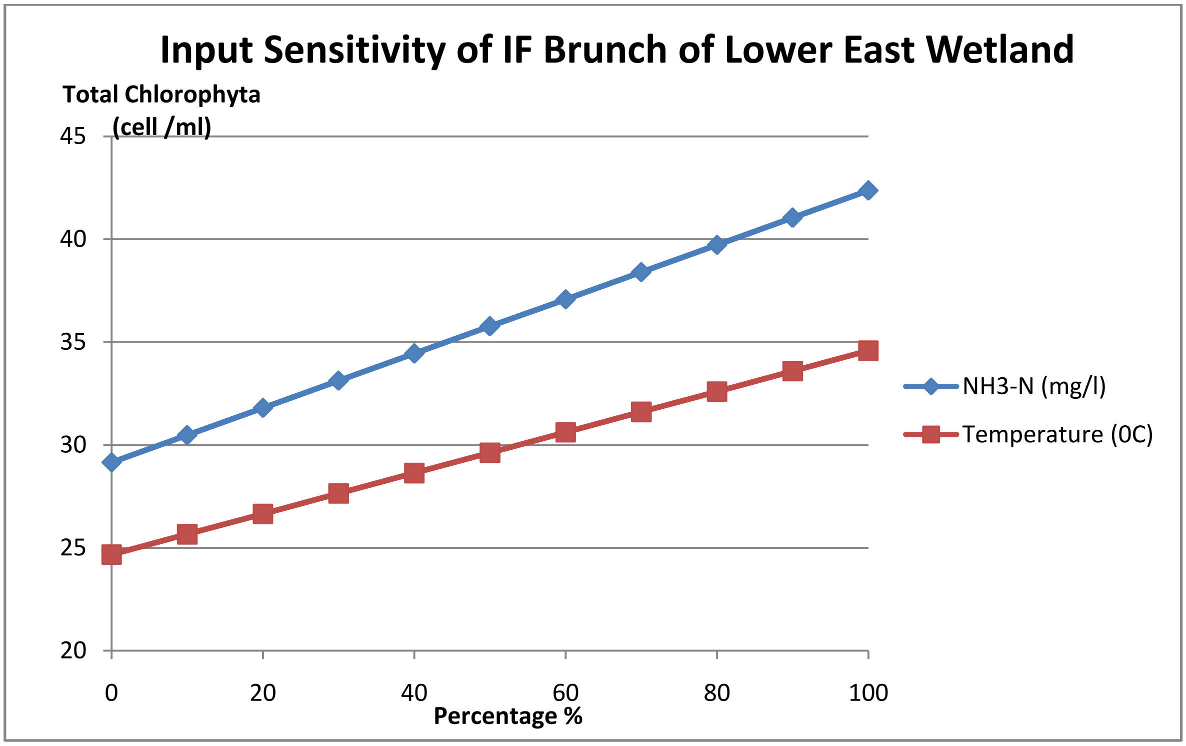

Figure 4.17 : Input Sensitivity of IF Brunch of Lower East Wetland, Total Chlorophyta against Input Range (%)

Figure 4.18 shows the input sensitivity of the ELSE brunch of Lower East Wetlands (LE) section ELSE rule shows that abundance of Chlorophyta is dependent on turbidity. Result shows that Chlorophyta abundance increases when turbidity increase.

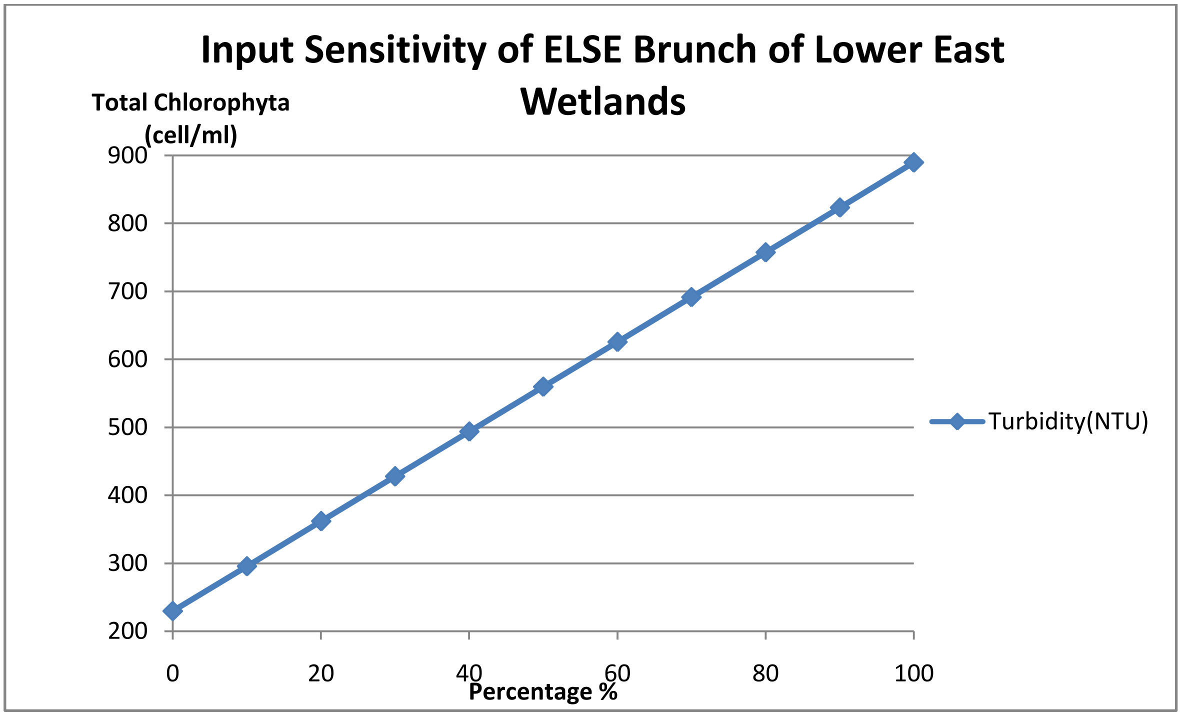

Figure 4.18 Input Sensitivity of ELSE Brunch of Lower East Wetland, Total Chlorophyta against Input Range (%)

- Upper Bisa Wetlands (UB)

For Upper Bisa Wetlands, the condition statement of rule application determines by threshold value of turbidity. The rule indicates that IF turbidity is more than 1 THEN the Total Chlorophyta abundance is calculated by the equation Tchlorophyta=((85.916/((41.978/NO3N)-Cond))+28.769) ELSE by the equation Tchlorophyta=226.054.

Figure 4.19 shows the input sensitivity of the IF brunch of Upper Bisa Wetlands (UB) section. IF rule shows that Chlorophyta abundance is positively correlated to level of concentration of NO_3-N and conductivity.Both NO3-N and conductivity give a consistent result of Chlorophyta abundance which is around 27cells/ml.

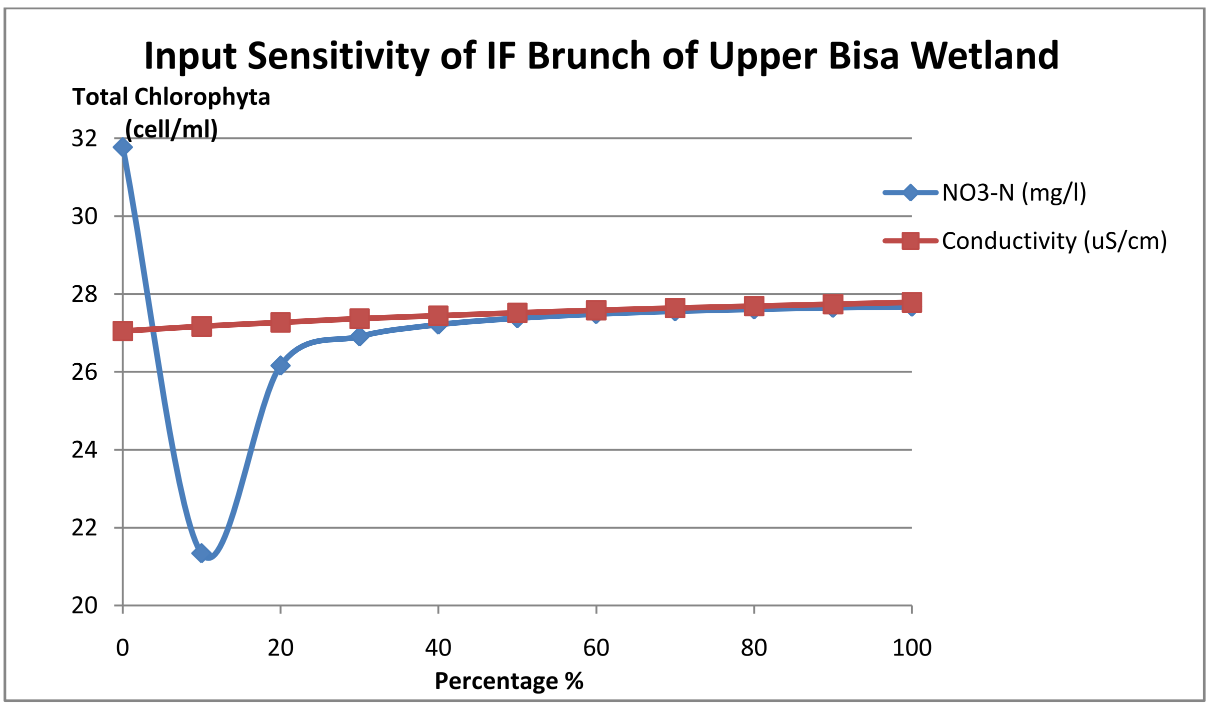

Figure 4.19 : Input Sensitivity of IF Brunch of Upper Bisa Wetland, Total Chlorophyta against Input Range (%)

Figure 4.20 shows the input sensitivity of the ELSE brunch of Upper Bisa Wetlands (UB) section ELSE rule shows that abundance of Chlorophyta is always consistence at 226cells/mlwhen turbidity between range of 0 to 1 (NTU). This shows that turbidity in this range does not have much effect on the Chlorophyta abundance.

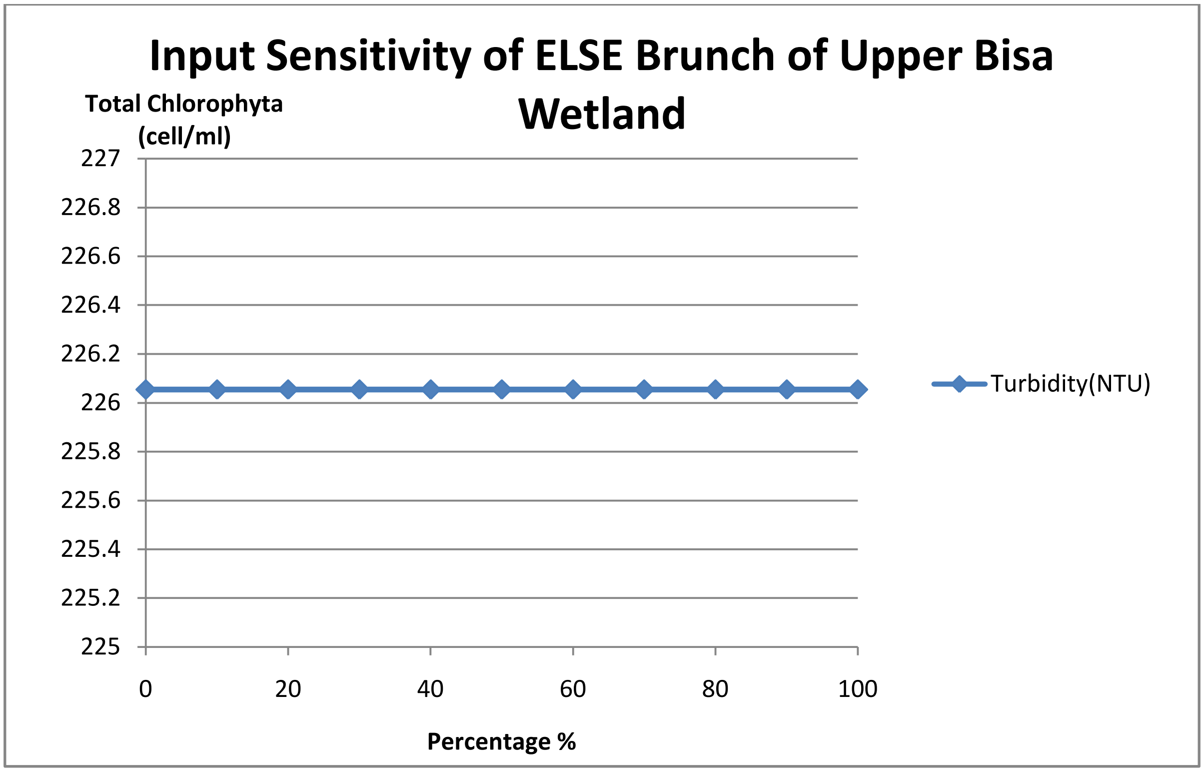

Figure 4.20 : Input Sensitivity of ELSE Brunch of Upper Bisa Wetland, Total Chlorophyta against Input Range (%)

- Upper East Wetlands (UE)

For Upper East Wetlands, the condition statement of rule application determines by threshold value of sunshine and rainfall. The rule indicates that IF sunshine is less than or equals to9.583 AND rainfall more than or equals to8.487 OR less than or equals to6.549THEN Total Chlorophyta abundance is calculated by the equation Tchlorophyta=(exp(exp(Windspeed))+(exp(exp(Windspeed))+23.026))ELSE by the equation Tchlorophyta = (Turb+236.425).

Figure 4.21 shows the input sensitivity of the IF brunch of Upper East Wetlands (UE) section. IF rule show that Chlorophyta abundance is positively correlated to level of wind speed. This means that when wind speedincreases, Chlorophyta abundance will increase.

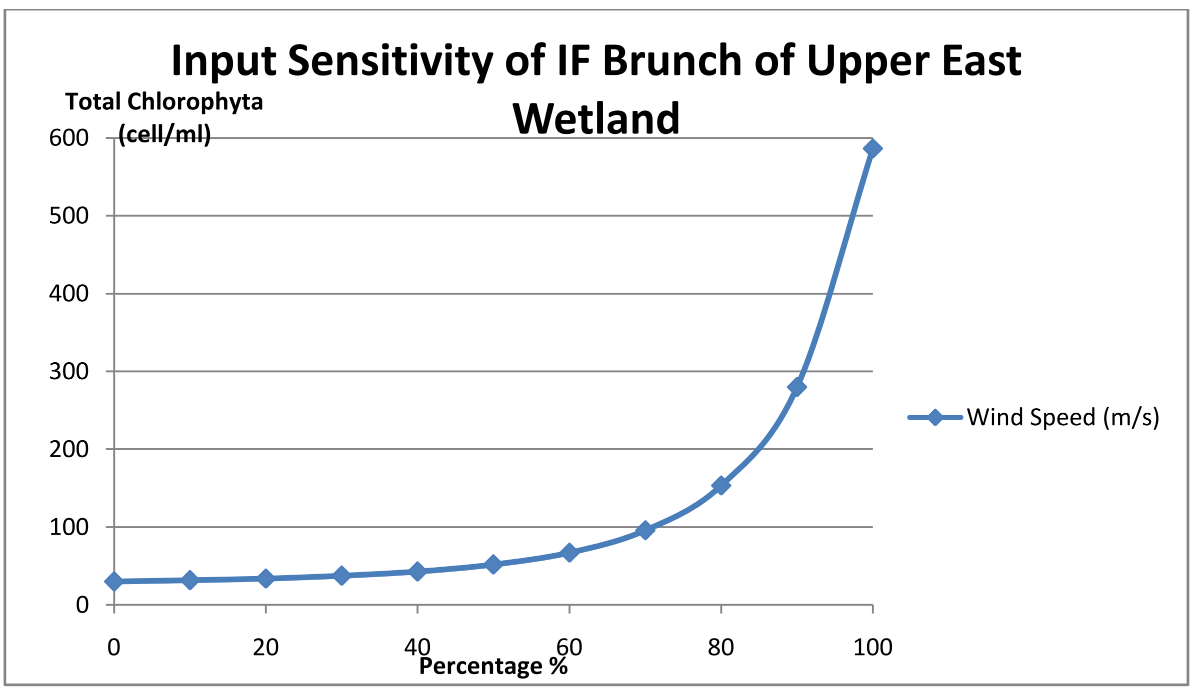

Figure 4.21 : Input Sensitivity of IF Brunch of Upper East Wetland, Total Chlorophyta against Input Range (%)

Figure 4.22 shows the input sensitivity of the ELSE brunch of Upper East Wetlands (UE) section. ELSE rule shows that abundance of Chlorophyta is positively correlated to turbidity.

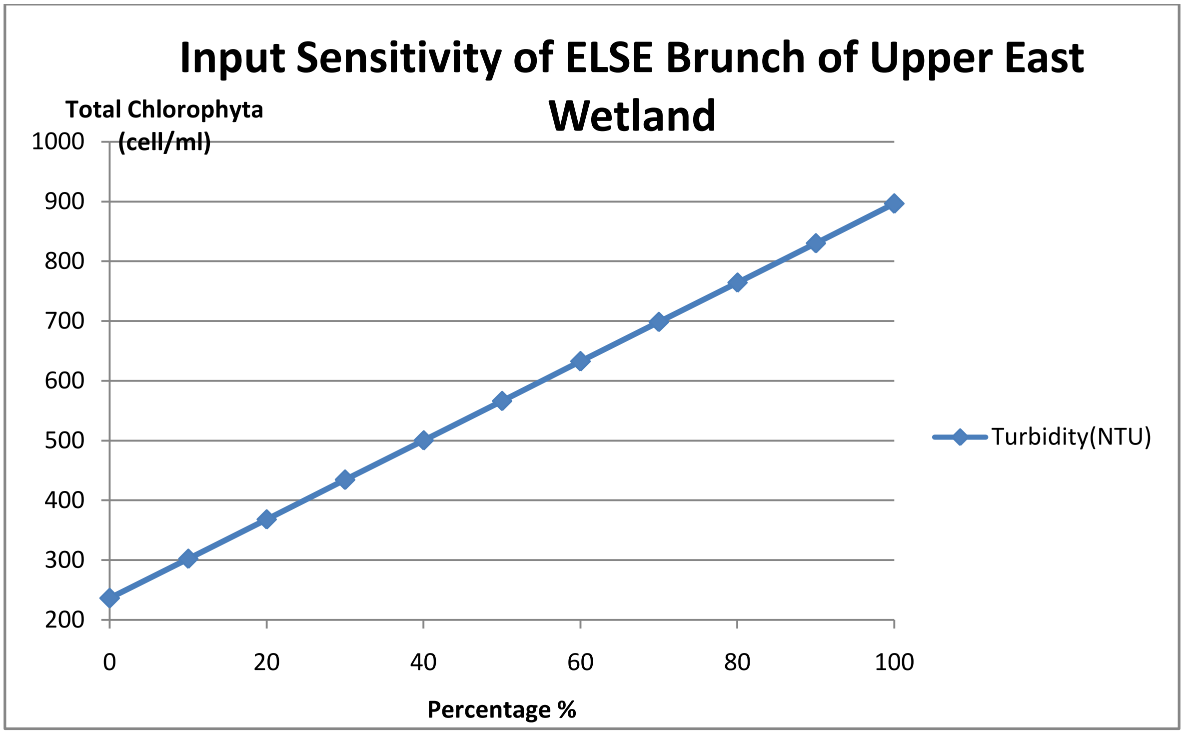

Figure 4.22 : Input Sensitivity of ELSE Brunch of Upper East Wetland, Total Chlorophyta against Input Range (%)

- Upper North Wetlands (UN)

For Upper North Wetlands (UN), the condition statement of rule application determines by threshold value of NO_3-N, TPO_4 and sunshine. The rule indicates that IF (exp(NO3N))/TPO4is more than 19.509 OR (Sunshine/TPO4) is less than or equal to 86.085 THEN the Total Chlorophyta abundance is calculated by the equation Tchlorophyta=((ln(|ln(|Turb|)|)+5.129)+(Cond/(exp(pH)/(Cond-(-144.302))))) ELSE by the equation Tchlorophyta=(((32.515/NO3N)+121.739)+Turb).

Figure 4.23 shows the input sensitivity of the IF brunch of Upper North Wetlands (UN) section. IF rule show that Chlorophyta abundance is positively related to level of turbidity,conductivity and negatively related to pH. This means that when turbidity, conductivity increases or pH decreases, Chlorophyta abundance will increase.

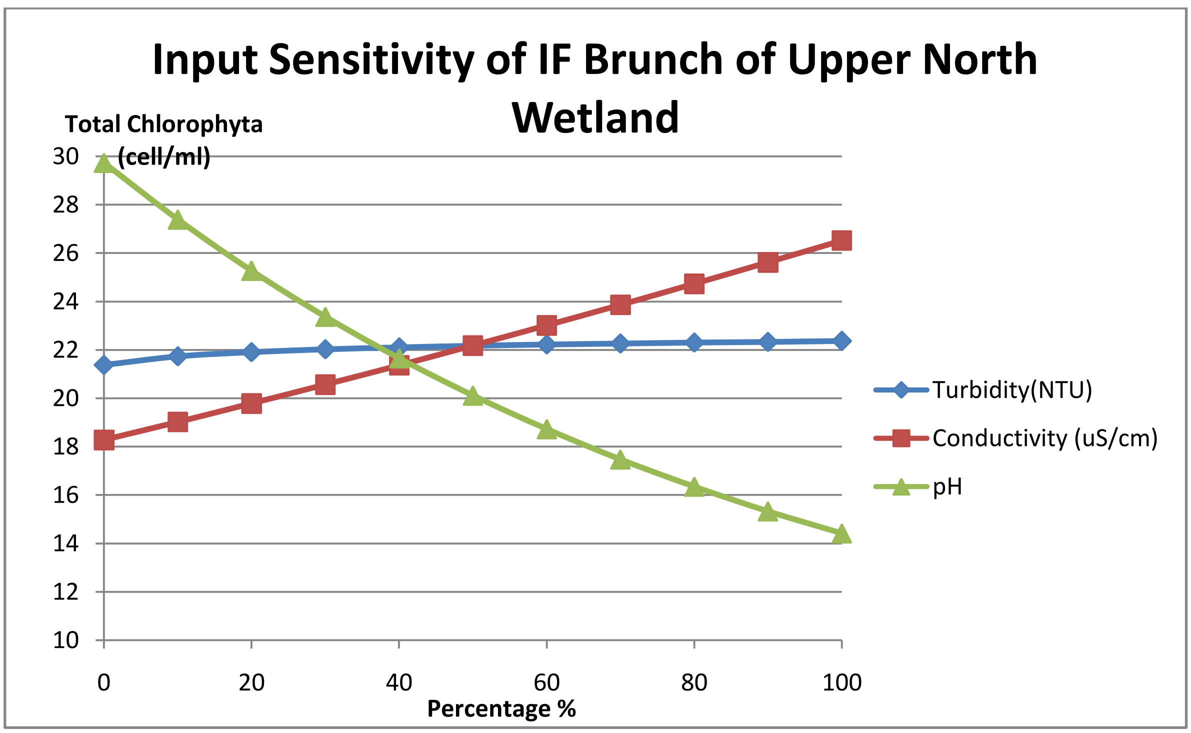

Figure 4.23 :Input Sensitivity of IF Brunch of Upper North Wetland, Total Chlorophyta against Input Range (%)

Figure 4.24 shows the input sensitivity of the ELSE brunch of Upper North Wetlands (UN) section.ELSE rule shows that abundance of Chlorophyta is dependent on turbidity and NO_3-N. Chlorophyta abundance increases while the concentration of NO3-N decreases and turbidity increases.

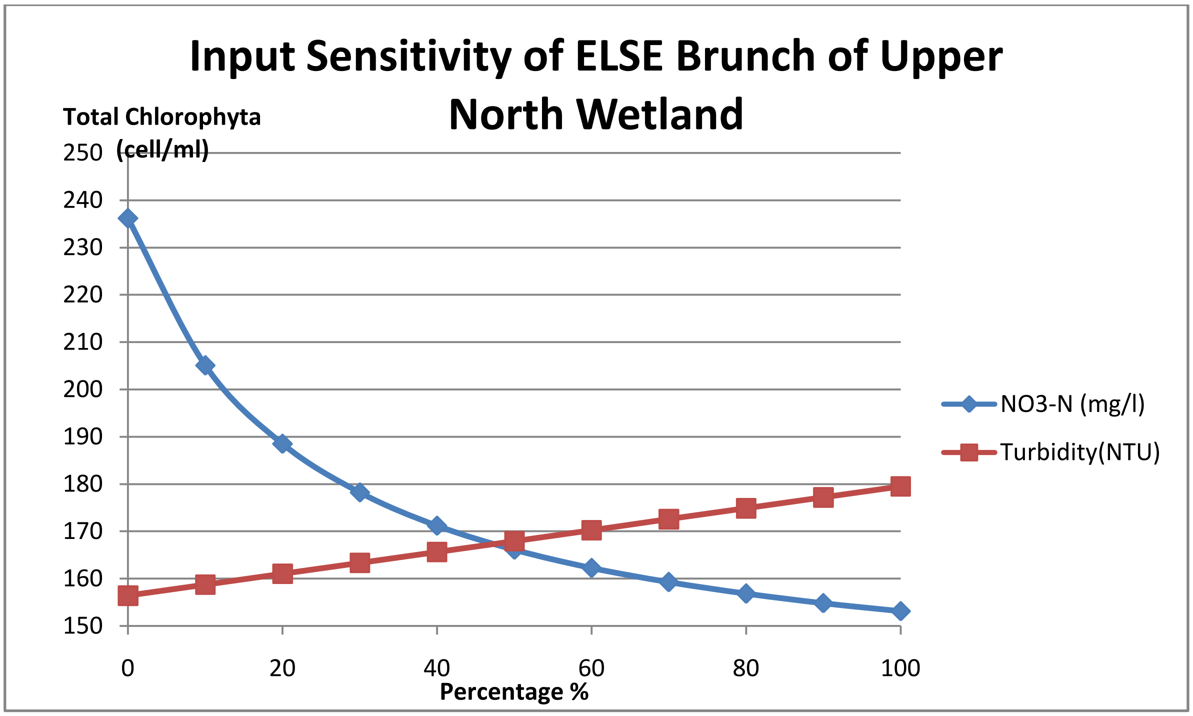

Figure 4.24 : Input Sensitivity of ELSE Brunch of Upper North Wetland, Total Chlorophyta against Input Range (%)

- Upper West Wetlands (UW)

For Upper West Wetlands, the condition statement of rule application determines by threshold value of conductivity and NH_3-N. The rule indicates that IF ((Cond+NH3N)+NH3N)+(Cond+NH3N) is less than to 97.474 THEN the Total Chlorophyta abundance is calculated by the equation Tchlorophyta=(((pH/(Sunshine-

Rainfall))+23.026)+((Sunshine*pH)/exp(Cond))) ELSE by the equation Tchlorophyta=(Rainfall+(((Sunshine*pH)/(Cond+(-80.792)))+23.212)).

Figure 4.25 shows the input sensitivity of the IF brunch of Upper West Wetlands (UW) section. IF rule shows that total Chlorophyta abundance is much rely on the constant value of 23.026. Total Chlorophyta abundance is not sensitive to pH, sunshine and conductivity.According to reports from Upper West wetland, the range of pH level is 6.89 to 7.94, the range of sunshine is 1.85 to 9.21 hours, and the range of conductivity is 84.7 to 122 uS/cm. From the sensitivity analysis, with changes of pH, sunshine and conductivity, total Chlorophyta abundance will still remains 21 to 22cells/ml.

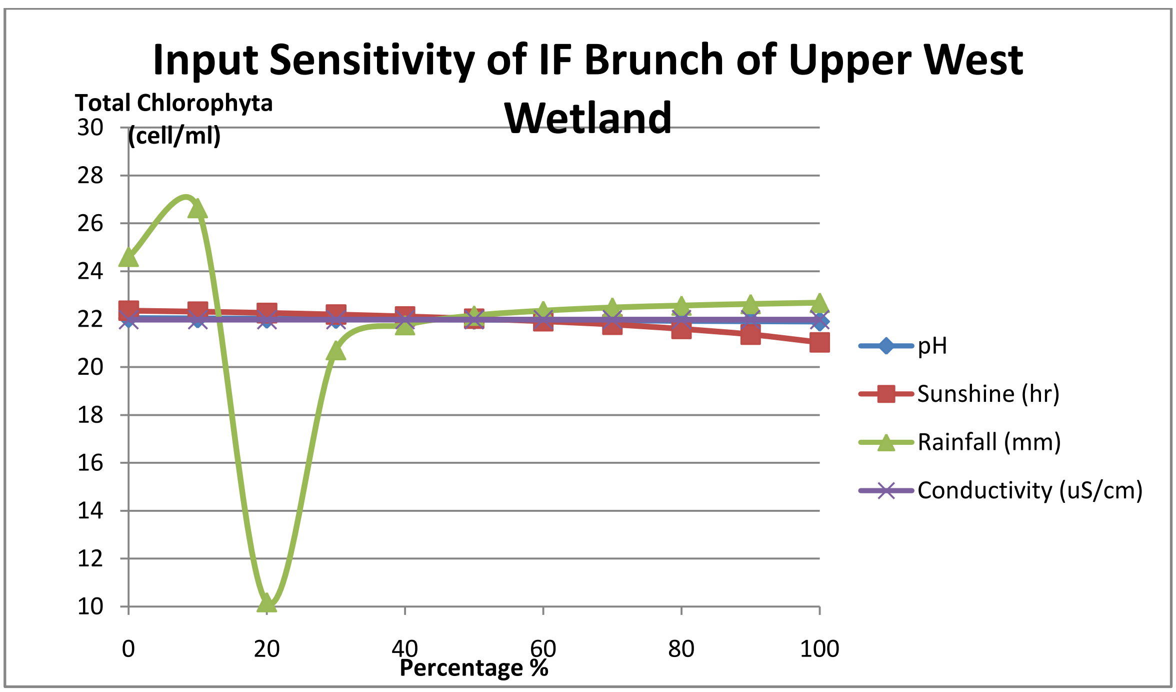

Figure 4.25 : Input Sensitivity of IF Brunch of Upper West Wetland, Total Chlorophyta against Input Range (%)

Figure 4.26 shows the input sensitivity of the ELSE brunch of Upper West Wetlands (UW) section. ELSE rule shows that abundance of Chlorophyta is lesser dependent on sunshine and pH. Total Chlorophyta is positively related to rainfall and negatively related to conductivity. Conditions of low Chlorophytaabundancehappened with lower rainfall and higher conductivity level.Conditions of high Chlorophyta abundance will reflected by lower conductivitylevelandheavier rainfall.

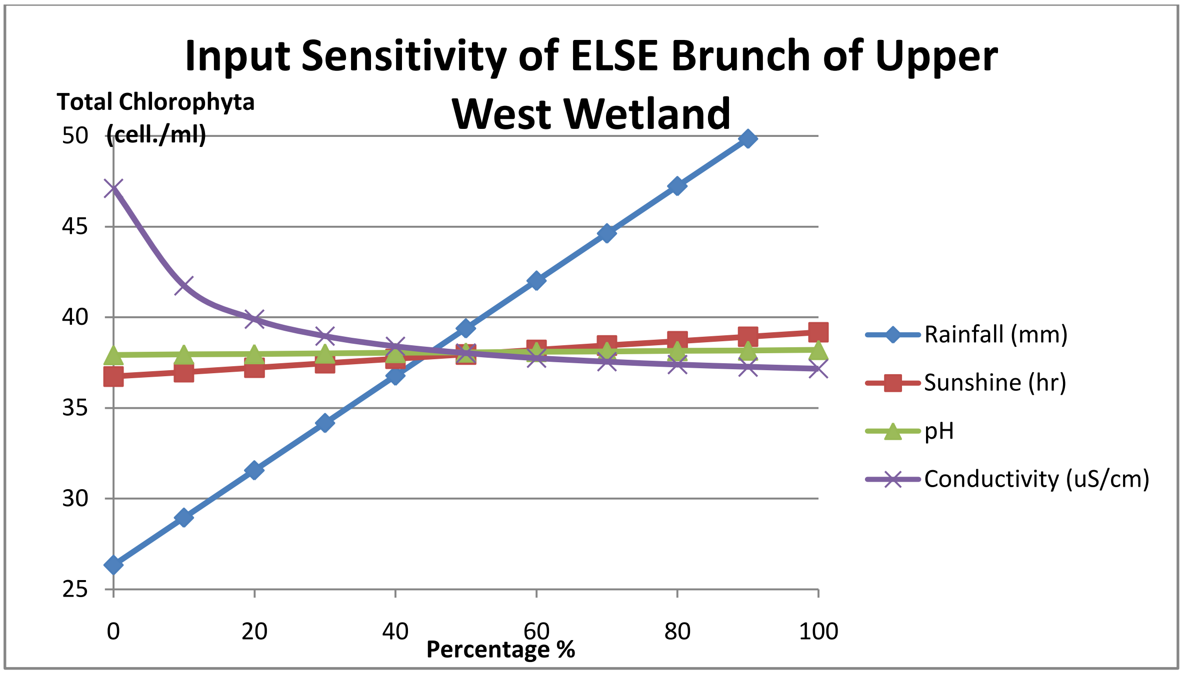

Figure 4.26 : Input Sensitivity of ELSE Brunch of Upper West Wetland, Total Chlorophyta against Input Range (%)

4.2 Discussion on Water Quality and External Variables affecting Chlorophyta Abundance

In this study, generic rule sets have been successfully used to distinguish seven zones in Putrajaya Lake and Wetlands where eache zone is characterized by a generic rule set that is determined from their water quality parameters. Most of the wetlands parts are getting percentage of accuracy over 90%. UN wetlands only manage to get 84.96% due to the amount of substation inside and each substation has a slight different in geographical characteristic. As water flowing from UN8 towards UN1, each and every parts of the substation in UN wetlands do filter up the nutrients and pollutant.

Putrajaya Lakes manage to get 72.79%, which is acceptable with just slightly above the threshold value. As Putrajaya Lakes are built surrounded Putrajaya city, each substation of Putrajaya Lakes only interacted with 2 neighbouring substations. Therefore, geographical characteristic and water quality are different for substation in different directions. Break down of Putrajaya Lakes into a few parts willgive a better result.

Chlorophyta abundance at Putrajaya Lakes and Wetlands decreases with increasing nitrate concentration and Chlorophyta are reported to be high in abundance at very low concentration of nitrate. This finding is supported from the sensitivity curve graph where highest Chlorophyta abundance occurs at lowest nitrate nitrogen concentration of 0.2 mg/l. A few parts of Putrajaya Lakes and Wetlands such as Putrajaya Lakes, Upper Bisa Wetlands and Upper North Wetlands have given the same results which support this finding.

This indicates that Chlorophyta has actively assimilated nitrate nitrogen to support growth. However the Chlorophyta abundance reduces when nitrate nitrogen concentration increases. The reduction is not significant which can be seen from the sensitivity curve.Mean nitrate nitrogen concentration of Putrajaya Lakes and Wetlands during the years 2001 to 2006 was 1.2mg/l which was comparable to that of Lake Chinias reported by Kutty *et. al.* (2001). They reported that nitrate nitrogen of the lake as 1.1mg/l with most algae comprises of species belonging to the division Chlorophyta. Desmids are reported by the authors as the majority component of the Chlorophyta in Lake Chini Pahang.

This stude shows that Chlorophyta can tolerate low nitrate nitrogen. They rarely profuse in water bodies unless nutrient levels are high, which is frequently the situation in shallow, well-mixed lakes (Jensen *et. al.* 1994; Jeppesen, 1998). Such condition is similar to that of Putrajaya Lakes and Wetlands. Furthermore Chlorophyta population at Putrajaya Lake and Wetlands also comprises of genera such *Scenedesmus* which is susceptible to loss by grazing due to their small size as the nitrate concentration increases, which explains the sensitivity graph plot findings at higher concentration of nitrate.

Light is an important factor for phytoplankton growth through photosynthesis, in which light energy is used to transform inorganic molecules into organic matter (Graham and Wilcox, 2000). With the present of more than 8 hours of sunshine per day, process of photosynthesis of Chlorophyta will be encouraged. As the growth of Chlorophyta abundance depends on process of photosynthesis. Sunlight is an important ingredient to transform inorganic material to organic matter (Graham and Wilcox, 2000). This indicates that Chlorophyta has actively assimilated nitrate nitrogen to support growth(Kutty *et al*., 2001). Besides that, increment of nitrate and sunlight lead to high Chlorophyll values contain inside Chlorophyta. Higher chlorophyll values will improve the effectiveness in process of photosynthesis (Hutchings *et al*., 1995, J. T. Pennington*et al*., 1999). In this study, as refers to the rule sets generated by HEA for Putrajaya Lakes and Upper West Wetlands, sunshine was the important factor that increases the Chlorophyta abundance.

Increment of nutrient concentration such as nitrate will increase the turbidity level in the water body. Reduction of light availability will prevent extensive photosynthesis, (Wetzel, 2001). While turbidity level increases, the water turns cloudy and sunlight started to get harder to reach the water. This will eliminate the process of photosynthesis and will decrease Chlorophyta growth. But Putrajaya Lakes and Wetlands were recorded with average 96.24uS/cm of high conductivity level, the outflows of the water prevent the increment of the turbidity level. With high conductivity level and long hours of sunlight, high Chlorophyta abundance will occurred (Wetzel, 2001).

Average water temperature in Putrajaya Lakes and Wetlands are around 30 degrees, this is the optimal water temperature for desmids species of Chlorophyta to correspond with (Coesel & Wardenaar, 1990). Furthermore many cellular processes are temperature dependent and their rates increases with higher temperature. Reynolds (1984), states maximal productivity of algae occurs between 25 ^{0}C and 40 ^{0}C. Shen (2002) reported that most favorable temperature was 30^{0}C which conforms to findings from this study which average temperature of Putrajaya Lakes and Wetlands is 30.37^{0}C. According to Craig (1991), maximum growth of larger green algae is most commonly found during summer time with warm water condition. This further supported the behavior pattern of Chlorophyta in Putrajaya Lakes and Wetlands. Some of the species of Chlorophyta such as *Chlamydomonas, Chlorella* and *Scenedesmus* belong to category of micro-green algae which are also found at Putrajaya Lake and Wetlands. These algae speciesmostly reside and contain in nutrient poor oligotrophic lakes(Priddle and Happy-Wood, 1983). Maximum growth is favored by warm water conditions and higher light irradiance which is a typical of tropical climate at Putrajaya Lakes and Wetlands which justifies high abundance of green algae at Putrajaya Lakes and Wetlands.

When the rainfall increase, water contain in the water body will increases. It will wash away the floating mats and allows sunlight to reach bottom part of the water body. Rainfall also will cause slight erosion at littoral zone of the Wetland. The littoral zone is situated near the shore where rooted plants grow. With heavy rainfall, the nutrients from sediments such as nitrate, phosphorus and ammonia will dilute into the water. These nutrients will act as fertilize towards Chlorophyta and increase the productivity (Spencer and Lembi, 1981). These will increase and encourage the process of photosynthesis, the productivity of Chlorophyta abundance will increases as well.

Putrajaya Wetlands were built to filter water inflows from Sungai Chuau and Sungai Bisa before the water flows down to Putrajaya Lakes. As the wetlands filtered most of the heavy metal and toxic, the highest change of increasing of COD levels are narrowed down to landslide which caused by construction of land development and agro-based industrial effluents. This will cause the increment of concentration of phosphorus and ammonia contain inside Putrajaya Lakes. ammonia and phosphorus are both suppliers of nitrogen-containing nutrient for algae growth (Karl, 1998). In another words, both ammonia and phosphorus can also be called as fertilizer for algae. Ammonia can be converted to nitrite (NO_2) and nitrate (NO_3) by bacteria, and then used by plants. Increase inCOD tends to be also accompanied by an increase in ammonia and phosphorus which will lead to increase in Chlorophyta abundance.

CHAPTER FIVE

CONCLUSION

5.1 Conclusion

Data visualization system had been successfully developed in this study based on thematic map and hybrid evolutionary algorithm for Chlorophyta abundance in Putrajaya Lakes and Wetlands using data collected from year 2001 to 2006. Prediction model of the system which integrated rule sets from HEA produced promising results with percentage of accuracy range between 72.79% to 96.97% for different parts of Putrajaya Lakes and Wetlands. In this study, HEA has proved to be practical and successful technique when dealing with datasets of complex relationship and without clear distinction of membership. HEA also proved the ability of explanatory by generate rule sets with conditional operators of IF and ELSE by combination of logical operators and mathematical equations. Thematic map has given the accessibility to successful visualized the abundance of Chlorophyta over a period of time according to the data sets in the system's file based.

Although the data visualization system has been successfully developed, there are setbacks that need to be resolved. One of the biggest aspects is dealing with the arrangement of data. Along the research, the most expensive task in terms of time constraint and human power is data arrangement. A stable database approach needed to be implemented for all the ecological data collection to ease and save time for data arrangement. Second, the prediction model can only take coding from visual basic, C and C++ programming language. In future to come, natural language processing should be applied into the prediction model to make it more readable and understandable for the majority.

In term of opportunity, the system being made to be industrial driven. It can be put to practice directly on Putrajaya Lakes and Wetlands for a fine tuning period and expand to other water bodies and ecological departments. Beside freshwater reservoir, the system can be expanded and linked to other ecological studies such forest and population researches.

Data visualization system developed in this study can also be improved to make it compatible with web-based applications. This will enable fast and accurate exchange of information between scientists around the world. Industry is instrumental to development of data visualization system as its players compete among themselves to improve their system in order to meet the demand and requirement of the market. Future trend in data visualization system development is towards an automated thematic maps generation aided system. Thematic map can be generated without input of coordinates and manual calculation but just by satellite image or simple image of a certain reservoirs or lakes. This can be done with built-in prediction model generator such as HEA, Fuzzy Logic or Artificial Neural Network in which data can be directly feed into the visualization which subsequently calculate relationship between the output (e.g. Chlorophyta abundance) and other parameters.

REFERENCES

References

Aamand, J., Jorgensen, C.,Arvin, E., and Jensen B. K.(1989).Microbial adaptation to degradation of hydrocarbons in polluted and unpolluted groundwater. *Journal of Contaminant Hydrology,* 4:299-3 12.

Abeliovich, A., and Azov, Y. (1976). Toxicity of ammonia to algae in sewage oxidation ponds. *Applied and environmental microbiology*, 31(6), 801-806.

Aik, Y.C. (2002). The effectiveness of the Asian waterbird census and other related programmes as a tool for waterbird and wetland monitoring in Malaysia. Proceedings of the Workshop on Developing a Proposed Framework for a Wetland Inventory, Assessment and Monitoring system (WIAMS), Kula Lumpur Malaysia. *Wetland International Malaysia*, pp: 92-101.

APHA, AWWA and WPCF(1975). Standard Methods for the Examination of Water and Wastewater 14th edition.American public health association.Washington, pp.125-126.Franson, M.A.H. (1995). American Public Health Association American Water Works Association Water Environment Federation.Methods, 6, 84.

Ariffin, S.J. (1998). Putrajaya and Putrajaya lake development. Proceedings of Seminar on Putrajaya Lake Water Quality Management, Putrajaya Corporation, Selangor Darul Ehsan, pp: 1-30.

Banzhaf, W., Francone, F.D., Keller, R.E., and Nordin, P. (1998). *Genetic programming: an introduction: on the automatic evolution of computer programs and its applications*. San Francisco, CA:Morgan Kaufmann Publishers Inc.Holland, J.H. (1975). Adaptation in Natural and Artificial Systems.*University of Michigan Press*.

Bellinger, E.G. and Sigee, D.C. (2010). References, in Freshwater Algae: Identification and Use as Bioindicators, John Wiley & Sons, Ltd, Chichester, UK.

Briney, A. (2009). Geography. In *Thematic Maps - An overview of thematic maps*. From http://geography.about.com/od/understandmaps/a/thematicmaps.htm.

Cairns, J.Jr., and Dickson, K.L. (1972). *An ecosystematic study in the South River, Virginia*.VirginiaWater Resource Center, Virginia Polytechnic Institute, Blacksburg.Bulletin, pp: 54-104.

Cairns, J.Jr., Lanza, G.R., and Parker, B.C. (1972). Pollution related structural and functional changes in aquatic communities with emphasis on freshwater algae and protozoa. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 124 (5): 79-127.

Cao, H., Recknagel, F., Joo, G. J., and Kim, D. K. (2006). Discovery of predictive rule sets for chlorophyll- a dynamics in the Nakdong River (Korea) by means of the hybrid evolutionary algorithm HEA. *Ecological Informatics*, *1*(1), 43-53.

Cao, H., Recknagel, F., Welk, A., Kim, B., Takamura, N. (2006). Hybrid Evolutionary Algorithm for Rule Set Discovery in Time-Series Data to Forecast and Explain Algal

Population Dynamics in Two Lakes Different in Morphometry and Eutrophication. Ecol Inform 2nd Edition. Springer-Verlag Berlin, Heidelberg, New York, 2006, 330-342.

Cao, H., Yu, J., Kang, L., Yang, H., and Ai, X. (2001). Modeling and prediction for discharge lifetime of battery systems using hybrid evolutionary algorithms. *Computers & chemistry*, *25*(3), 251-259.

Carlson, R.E., (1977). A trophic state index for lakes. *Limnol.Oceanogr*. 22: 361–369.

Chapman, D. V. (Ed.). (1996). Water quality assessments: a guide to the use of biota, sediments and water in environmental monitoring (p. 626). London: E &Fn Spon.Chen Q: A robust fuzzy logic approach to modelling algal biomass. J Hydraulics Res 2004, 42(3): 303–309

Chen, Q., and Mynett, A.E., (2003). Effects of cell size and configuration in cellular automata based prey-predator modeling. *Simulation Modelling Practice and Theory* 11:609-625.

Chen, Q., and Mynett, A.E., (2003).Integration of data mining techniques with heuristic knowledge in a fuzzy logic modelling of eutrophication in Taihu Lake.*Ecological Modeling*. pp. 55-67.

Coesel, P.F.M. and Wardenaar, K. (1990).Growth responses of planktonic desmid species in a temperature—light gradient.*Freshwater Biology*, 23: 551–560.

Coesel, P.F.M. (1983). The significance of desmids as indicators of the trophic status of freshwaters.*Schweiz. Z. Hydrol*., 45: 388-393.

Coesel, P.F.M. (2001). A method for quantifying conservation value in lentic freshwater habitats using desmids as indicator organisms.*Biodiver.Conserv*., 10: 177-187.

Dai, A., Trenberth, K. E., and Karl, T. R. (1998). Global variations in droughts and wet spells: 1900–1995. *Geophysical Research Letters*, *25*(17), 3367-3370.

Drinkwater, M., Rebhan, H., LeTraon, P.Y., Phalippou, L., Cotton, D., Johannessen, J., Ruffini, G., Bahurel, P., Bell, M., Chapron, B., Pinardi, N., Robinson, I., Santoleri, L. and Stammer, D. (2005). The Roadmap for a GMES Operational Oceanography Mission [electronic version], ESA Bulletin, 124, 43-48.

Evans, L. V., and Holligan, M. S. (1972). Correlated light and electron microscope studies on brown algae I. Localization of alginic acid and sulphated polysaccharides in Dictyota.*New Phytologist*, *71*(6), 1161-1172.

Few, S. (2010). Data Visualization for Human Perception. In: *Encyclopedia of Human-Computer Interaction*, Soegaard, Mads and Dam, Rikke Friis (eds.),. Aarhus, Denmark: The Interaction-Design.org Foundation.Available online athttp://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html.

Friedman, V. (2008). Data visualization and infographics. *Graphics, Monday Inspiration*, *14*, 2008.

Gerrath, J.F. (1993). The Biology of Desmids a Decade of Progress. In: *Progress in Phycological Research 9*, Round, F.E. and D.J. Chapman (Eds.). Biopress Ltd., Bristol, pp: 79-192.

Gerrath, J.F. (2005). Conjugating Green Algae and Desmids. In: *Freshwater Algae of North America: Ecology and Classification*, Wehr, J.D. and R.G. Sheath (Eds.). Academic Press, USA., pp: 353-381.

Graham, L. E., and Wilcox, L.W. (2000).Algae. Prentice Hall, Inc. 695, New Jersey, pp: 4.

Grandin, M. (2009). A new report from the UK World Wildlife Foundation reveals the harmful affects of climate change on plankton and fish distribution and species variety in the United Kingdom's seas. . North America: Diversified Business Communications.

Kitsiou, D., and Karydis, M. (2000). Categorical mapping of marine eutrophication based on ecological indices. *Science of the total environment*, *255*(1), 113-127.

Maier, H. R., Dandy, G. C., and Burch, M. D. (1998). Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray, South Australia. *Ecological Modelling*, *105*(2), 257-272.

Hunter, C. L., and Evans, C. W. (1995). Coral reefs in Kaneohe Bay, Hawaii: two centuries of western influence and two decades of data. *Bulletin of Marine Science*, *57*(2), 501-515.

Klemas, V. (2009).The role of remote sensing in predicting anddetermining coastal storm impacts.*Journal of Coastal Research*, 25(6), 1264–1275.

Klemas, V. 2011. Remote sensing of wetlands: case studies comparingpractical techniques.*Journal of Coastal Research*, 27(3), 418–427.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59-69.

Kohonen, T. (2001). Self-organizing maps (Vol. 30). Springer.

Konopka, A., and Brock, T. D. (1978). Effect of temperature on blue-green algae (cyanobacteria) in Lake Mendota. *Applied and Environmental Microbiology*, *36*(4), 572-576.

Koza, J.R. (1992). *Genetic Programming: vol. 1: On the programming of computers by means of natural selection* (Vol. 1). MA: MIT Press.

Koza, J.R. (1994).*Genetic Programming II: Automatic Discovery of Reusable Programs.Cambridge*.Cambridge, MA: MIT Press.

Koza, J. R. (1994). *Genetic Programming II Videotape: The Next Generation* (Vol. 55). J. Noyes (Ed.). Cambridge, MA: MIT Press.

Kutty, G., Ma, L., & Kovacs, J.A. (2001).Characterization of the expression site of the major surface glycoprotein of human-derived Pneumocystis carinii.*Molecular microbiology*, *42*(1), 183-193.

Lehmusluoto, P., and Machbub, B. (1987). National Inventory of *the Major Lakes and Reservoirs in Indonesia: General Limnology*. Helsinki: Helsinki University Press.

Lewis, L.A. and Flechtner, V.R. (2004). Cryptic species of *Scenedesmus* (Chlorophyta) from desert soil communities of western North America. *Journal of Phycology* 40(6): 1127-1137.

Macleod, R. D., and Congalton, R. G. (1998). A quantitative comparison of change-detection algorithms for monitoring eelgrass from remotely sensed data. *Photogrammetric engineering and remote sensing*, *64*(3), 207-216.

Malek, S., Salleh, A., Milow, P., and Baba, M.S.. (2010).A Comparison between Neural Network Based and Fuzzy Logic Models for Chlorophyll-a Estimation, Second International Conference on Computer Engineering and Applications, pp.340-343.

Malek, S., Salleh, A., Milow, P., Baba, M.S., and Sharifah, S. A. (2012) *Applying artificial neural network theory to exploring diatom abundance at tropical Putrajaya Lake, Malaysia. Journal of Freshwater Ecology*, 27 (2). pp. 211-227. ISSN 0270-5060

MarCoast and Aquamar (2012). MarCoast a GMES Service Network. In *The European Water Quality services Network*. Retrieved July 20, 2012, from http://www.marcoast.eu/.

Martin, D.L., Cheyer, A., and Lee, G.L. (1996). Agent development tools for the open agent architecture. In *Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, The Practical Application Company Ltd. pp. 387-404.

McCormick, P.V., and Cairns, J.Jr. (1994).Algae as indicators of environmental change.*Journal of Applied Phycology*, 6(5-6), 509-526.Wehr JD, Descy JP : Use of phytoplankton in large river management. J Phycol 1998, 34: 741–749.

McCourt, R.M. (1995). Green algal phylogeny.*Trends in Ecology and Evolution* 10: 159-163.

Mingyang, L., Yunwei, J., Kumar, S., and Stohlgren, T. J. (2008). Modeling potential habitats for alien species *Dreissena polymorpha* in Continental USA.*Acta Ecologica Sinica, 28*(9), 4253-4258.

Mitchell, M. (1996). An Introduction to Genetic Algorithms.Cambridge: The MIT Press.

Mostert, E.S., and Grobbelaar, J.U. (1987). The influence of nitrogen and phosphorus on algal growth and quality in outdoor mass algal cultures.*Biomass*, 13(4), 219-233.

O'Sullivan, P., and Reynolds, C. S. (Eds.). (2008). *The lakes handbook: limnology and limnetic ecology* (Vol. 1). Wiley.com.

Park, C. C. (1987). Acid rain: rhetoric and reality. London: Mathuen & Co. Ltd.

Perbadanan Putrajaya (1999). Putrajaya Wetlands, Multi-cell approach. pp: 52-63.

Perbadanan Putrajaya (2006). In *Portal Rasmi Perbadanan Putrajaya*. Retrieved July 25, 2012, from www.ppj.gov.my/.

Post, F.H., Nielson, G.M., and Bonneau, G.P. (2002). Data Visualization: The State of the Art. Research paper TU delft.

Post, F. H., Vrolijk, B., Hauser, H., Laramee, R. S., and Doleisch, H. (2003). The state of the art in flow visualisation: Feature extraction and tracking. In *Computer Graphics Forum* (Vol. 22, No. 4, pp. 775-792). Blackwell Publishing, Inc.

Priddle, J., and Happey-Wood, C. M. (1983). Significance of small species of Chlorophyta in freshwater phytoplankton communities with special reference to five Welsh lakes. *The Journal of Ecology*, 793-810.

Principe, J. C., Euliano, N. R., and Lefebvre, W. C. (1999). *Neural and adaptive systems: fundamentals through simulations with CD-ROM*. John Wiley & Sons, Inc..

Prokop, A., Quinn, M.F., Fekri, M., Murad, M., and Ahmed, S.A. (2004).Spectral shifting by dyes to enhance algae growth. Biotechnology and Bioengineering, pp. 1313-1322.

Qua, S., Chen, Q., and Recknagel, F. (2008). Cellular automata based simulation of random versus selective harvesting strategies in predator-prey systems. *Ecological Informatics* 3:252-258.

Ramsar Convention Secretariat (2013).The Ramsar. Convention Manual: a guide to the. Convention on Wetlands (Ramsar,. Iran, 1971), 6th ed.

Recknagel, F., French, M., Harkonen, P., & Yabunaka, K. I. (1997). Artificial neural network approach for modelling and prediction of algal blooms.*Ecological Modelling*, *96*(1), 11-28.

Recknagel, F., Welk, A., Kim, B., & Takamura, N. (2006).Artificial neural network approach to unravel and forecast algal population dynamics of two lakes different in morphometry and eutrophication.In *Ecological informatics* (pp. 325-345).Springer Berlin Heidelberg.

Recknagel, F., Welk, A., Cao, H., Chan, W.S., Talib, A. (2008).Rule-based agents for forecasting algal population dynamics in freshwater lakes discovered by hybrid evolutionary algorithms. Ecological Informatics, pp:46-54.

Reynolds, C. S. (1984). *The ecology of freshwater phytoplankton*. Cambridge University Press.

Rosenblum, L. J. (1994).Scientific Visualization: Advances and challenges. Academic Press.

Ryder, P. (2007). GMES Fast Track Marine Core Service. Strategic Implementation Plan, Final version 24.04.2007 [electronic version].

Ryding, S. O., and Rast, W. (1989).Control of eutrophication of lakes and reservoirs.*Manual the biosphere series, 1*.Sommer, U. (1989).Nutrient status and nutrient competition of phytoplankton in a shallow, hypertrophic lake.*Limnology and Oceanography*, *34*(7), 1162-1173.

Saeed, U., Fatima, K., Gilani, H., Daud, A., and Ashraf, S. (2008, November). Selection of an appropriate classification technique for coastal biomass mapping using high and low resolution dataset. In *Advances in Space Technologies, 2008. ICAST 2008. 2nd International Conference on* (pp. 80-84). IEEE.

Saravi, H.N., Din, Z.B. and Makhlough, A.(2008).Variations in Nutrient Concentration and Phytoplankton Composition at the Euphotic and Aphotic Layers in the Iranian Coastal Waters of the Southern Caspian Sea.*Pakistan Journal of Biological Sciences*, 11: 1179-1193.

Shen, D. S. (2002). Study on limiting factors of water eutrophication of the network of rivers in plain. *Journal of Zhejiang University(Agriculture and Life Sciences)*, *28*(1), 94-97.

Shutes, R.B.E. (2001). Artificial wetlands and water quality improvement.*Environmental International*, 26(5-6): 441-447.

Spencer, D.F., and Lembi, C.A. (1981).Factors Regulating the Spatial Distribution of the Filamentous Alga Pithophora Oedogonia (Chlorophyceae) in an Indiana Lake.*Journal of Phycology*, 17(2), 168-173.

Tomlinson, M. C., Wynne, T. T., and Stumpf, R. P. (2009). An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia brevis*. *Remote Sensing of Environment*, *113*(3), 598-609.

Vanni, M. J., and Temte, J. (1990). Seasonal patterns of grazing and nutrient limitation of phytoplankton in a eutrophic lake. *Limnology and Oceanography*, 35(3), 697-709.Carrillo, P., Reche, I., Sánchez-Castillo, P., and Cruz-Pizarro, L. (1995).Direct and indirect effects of grazing on the phytoplankton seasonal succession in an oligotrophic lake.*Journal of Plankton Research*, *17*(6), 1363-1379.

Wang, X. J., and Liu, R. M. (2005).Spatial analysis and eutrophication assessment for chlorophyll a in Taihu Lake. Environmental Monitoring and Assessment, 101(1-3), 167-174.

Wehr, J.D., and Sheath, R.G.(2003). Freshwater Algae of North America: Ecology and Classification. 2nd Edition., Academic Press, USA., pp: 918.

Wetzel, G.R., Likens, G.E., and Svatora, M. (1996). Limnological analyses 2nd Ed. *Photosynthetica*, *32*(3), 380-380.

Wilson, H., & Recknagel, F. (2001).Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes.*Ecological Modelling*, *146*(1), 69-84.

Zakaria, M., Rajpar, M.N. and Sajap, A.S. (2009).Species Diversity and Feeding Guilds of Birds in Paya Indah Wetland Reserve, Peninsular Malaysia.*International Journal of Zoological Research*, 5: 86-100.

APPENDIX

Appendix A

Script for operating HPC to perform HEA

```
#!/bin/sh

#PBS -V

### Job name
#PBS -N jb2
### Declare jobs non-rerunable
#PBS -r n

### Output files
#PBS -j oe

### Mail to user
#PBS -m ae

### Queue name
#PBS -q hydra

### Request nodes NB THIS IS REQUIRED
#PBS -l nodes=1,walltime=100:00:00

# This job's working directory
echo Working directory is $PBS_O_WORKDIR
cd $PBS_O_WORKDIR
echo Running on host `hostname`
echo Time is `date`

cp /dev/null result.txt
cp /dev/null variable.txt

./newbootstrapevolvesinglerule *.txt
```

Appendix B

Example of Self Organizing Map on water quality data and Chlorophyta abundance.