Predictive modelling of eutrophication and algal bloom formation in tropical lakes

MOHD YUSOFF ISHAK

A thesis submitted for the award of Doctor of Philosophy

2012

Ecoinformatics and Watershed Ecology Laboratory
School of Earth and Environmental Sciences
Environmental Biology Discipline
The University of Adelaide, Australia
CONTENTS

THESIS CONTENTS ..................................................................................................... i
Abstract .......................................................................................................................... v
Declaration ................................................................................................................... vii
Acknowledgements .................................................................................................... viii
List of Tables ................................................................................................................ ix
List of Figures .............................................................................................................. xii

CHAPTER 1 ................................................................................................................. 1
1.0 INTRODUCTION ................................................................................................... 1
1.1 General Introduction ............................................................................................ 1
1.2 The Computational Modelling Approach of This Study ................................. 2
1.3 Objectives and Achievements of This Study ....................................................... 3
1.4 Thesis Overview .................................................................................................. 5

CHAPTER 2 ................................................................................................................. 6
2.0 LITERATURE REVIEW ........................................................................................ 6
2.1 Lake Ecosystem and Lake Models ...................................................................... 7
  2.1.1 Lake Ecosystem ............................................................................................ 7
  2.1.2 Ecological Lake Models ............................................................................... 8
2.2 Data-driven Lake Models .................................................................................. 13
  2.2.1 Artificial Neural Networks (ANN) ............................................................. 14
  2.2.2 Evolutionary Algorithms (EA) ................................................................... 15
  2.2.3 HEA: Hybrid Evolutionary Algorithms ...................................................... 16
2.3 Process-based Lake Models .............................................................................. 17
  2.3.1 The Differential Equation Approach .......................................................... 17
  2.3.2 The Model SALMO-OO ............................................................................. 18
  2.3.3 Structure and Parameter Optimisation of SALMO-PLUS ...................... 22
  2.3.4 Particle Swarm Optimisation (PSO) ........................................................... 23
2.4 DYRESM Model for Lake Putrajaya ................................................................. 25
2.5 Validation Methods ............................................................................................ 26
  2.5.1 Leave-one-out Cross Validation ................................................................. 26
  2.5.2 Bootstrap Procedure .................................................................................... 27
2.6 Models Generic for Lake Ecosystem Categories ............................................ 28
3.6.2 Data Treatment for the HEA ................................................................. 78
3.6.3 Sensitivity Analysis ............................................................................ 79
3.6.4. Generic Predictive Rules for Particular Lake Category ..................... 80

3.7. DYRESM Model .................................................................................. 80

3.8. Fitness evaluation ................................................................................ 81

CHAPTER 4 ................................................................................................ 82
4.0 RESULTS .............................................................................................. 82

4.1 Experimental Results from HEA and the Generic Rules for Tropical Lakes ... 82
  4.1.1 Predictive Chl-a Model for Lake Putrajaya ........................................... 84
  4.1.2 Predictive Algal Biovolume Model for Lake Kenyir ............................. 91
  4.1.3 Predictive Chl-a Model for Lake Penang ............................................. 96
  4.1.4 Predictive Algal Biovolume Model for Lake Penang .......................... 101

4.2 Towards a Generic Forecasting Model for Algal Biovolume in Mesotrophic
  Tropical-stratified Lakes .......................................................................... 104
  4.2.1 Experiment 1 .................................................................................... 105
  4.2.2 Experiment 2 .................................................................................... 109

4.3 Generic Forecasting Model for Algal Biovolume in Mesotrophic Tropical-
  stratified Lakes using Electronically Measured Data ................................. 116
  4.3.1 Experiment with Merged Electronically Measured Data for Lakes Kenyir
    and Penang .............................................................................................. 116
  4.3.2 Experiment with Merged Electronically Measured Data for Lake Putrajaya
    .................................................................................................................. 121

4.4 Model Identification for Different Lake Categories by SALMO PLUS .......... 125

4.5 Mesotrophic tropical-stratified Lakes Kenyir and Penang ......................... 127
  4.5.1 Simulation Results for Lake Kenyir by Means of the Simulation System
    SALMO-PLUS .......................................................................................... 129
  4.5.2 Simulation Results for Lake Penang by Means of the Simulation System
    SALMO-PLUS .......................................................................................... 135

4.6 Dimictic Mesotrophic Saidenbach Reservoir ............................................ 141
  4.6.1 Simulation Results by Means of the Simulation System SALMO-PLUS for
    Saidenbach Reservoir, Germany ............................................................... 141

4.7 Warm-monomictic Hypertrophic Lake .................................................... 147
  4.7.1 Simulation Results by Means of the Simulation SALMO-PLUS for
    Roodeplaat Dam, South Africa ................................................................. 147

4.8 Warm Monomictic and Mesotrophic (South Para Reservoir, South Australia)
    ................................................................................................................. 153
  4.8.1 Simulation results by Means of the Simulation SALMO-PLUS for South
    Para Reservoir, South Australia .............................................................. 153
4.9 Extension of the SALMO-OO Simulation Library by Additional Process Model
................................................................................................................................ 159
4.9.1 Mesotrophic Tropical-stratified (Lake Kenyir)..............................................165
4.9.2 Mesotrophic Tropical-stratified (Lake Penang)............................................168
4.9.3 Dimictic and Mesotrophic (Saidenbach Reservoir, Germany)....................171
4.9.4 Warm Monomictic and Hypertrophic (Roodeplaat Dam, South Africa)......174

CHAPTER 5..................................................................................................... 178
5.0 DISCUSSION AND CONCLUSIONS ............................................................... 178
5.1 General Discussion ...................................................................................... 178
5.2 The SALMO-PLUS Model as a Tool for Tropical Lake Ecosystem Analysis 179
  5.2.1 Parameter Optimisation by SALMO-PLUS ............................................. 180
  5.2.2 Initial Value Identification ...................................................................... 183
5.3 Addition of the New Process Model from Law et al. (2009) into SALMO-OO
   Simulation Library ............................................................................................. 185
  5.3.1 Evaluation of SALMO-OO Results ......................................................... 186
5.4 Data-driven Forecasting Models Developed by HEA ............................... 190
  5.4.1 Input Sensitivity Analyses ....................................................................... 192
  5.4.2 Generic Forecasting Model of Algal Biovolumes for the Lakes Penang and
       Kenyir .............................................................................................................. 197
  5.4.3 Generic Forecasting Model of Algal Biovolumes for the Lakes Penang and
       Kenyir using Electronically Measured Data ............................................... 197
5.5 Combination of Process-based and Data–driven Modelling Techniques ....... 198
5.6 The DYRESM Model for Tropical Lakes ..................................................... 199
5.7 The Limitation of Input Data Set from Tropical Lakes in this Research........ 202
5.8 Closing Gaps on the Tropical Lake Ecosystem Analysis: Role of SALMO-
   PLUS and HEA Models .................................................................................. 204
5.9 Algal Functional Groups Dynamics in Tropical Lake Ecosystem............... 205
5.10 Conclusions .................................................................................................. 206
5.11 Future Works ............................................................................................... 208

APPENDIX A: Manual for preparing input and measured data, uploading data and
   operating the SALMO-PLUS

APPENDIX B: Input files for DYRESM model with a brief description for each file.
Abstract

My original contribution to knowledge is the successful application of two modelling paradigms 1) SALMO-PLUS process based model and 2) HEA data driven model to tropical lakes of different morphometry and trophic status. The application of SALMO-PLUS to tropical lakes involves utilising the SALMO-OO model structure for optimisation. This was followed by multi objective parameter optimisation on selected parameters to seek the optimum parameter values that can model the algal dynamics and state variables fluctuations in the tropical lakes to an acceptable magnitude and peaks.

SALMO-PLUS is another version SALMO-OO with capability to run optimisation by means of particle swarm optimisation (PSO) method. SALMO-OO has been used as a research tool over a number of lakes with different trophic states and mixing conditions to simulate algal succession and respond to ecosystem dynamic. SALMO-OO is driven by process-based differential equations and works by utilizing a library of three phytoplankton growth and three grazing process models.

Evolutionary algorithms (EA) are bio-inspired adaptive methods which mimic processes of biological evolution, natural selection and genetic variation such as cross-over and mutation to develop solutions to complex computational problems (Recknagel et al., 2006). HEA is designed for rule discovery in water quality time-series (Cao et al., 2006b) and is capable of forecasting potential algal population dynamics and outbreaks in water bodies.

The SALMO-PLUS model was applied for simulating the state variables of selected lakes (Lake Kenyir, Lake Penang, Saidenbach Reservoir, Roodeplaat Dam and South Para Reservoir). Measured data from the year 2005 and 1992 were used for Lake Penang and Lake Kenyir respectively. The HEA was applied for predicting the Chl-a and algal biovolume abundance on tropical lakes (Lake Putrajaya, Lake Penang and Lake Kenyir) in Malaysia. This study discusses the application of SALMO-PLUS and HEA towards tropical lakes eutrophication management. The results of application of SALMO-PLUS on tropical lakes are presented, simulating response of the phytoplankton community to fluctuation in nutrient loading, light availability and hydrological aspect in the water bodies. Results of applying HEA on tropical lakes are also interpreted in the context of empirical and causal knowledge on Chl-a and algal biovolumes dynamics under tropical lake water quality conditions by means of rule-based model.

Results for both Lake Kenyir and Penang showed that SALMO-PLUS were able to predict annual average trends not only for chlorophyll-a but also other state variables and algal functional groups. Simulated state variables namely Chl-a, N and P showed good agreement with field observations data for both lakes. Despite the fact that this is the first time SALMO-PLUS been used for tropical lakes and the limited data availability from this region, the simulated values of biological and nutrient state variables match reasonably with measured data. Outcomes from SALMO-PLUS simulation show consistent compliance with algal community assembly obtained from other researchers.
The HEA achieved reasonable accuracy in predicting timing and magnitudes of algal blooms up to 7-days-ahead. The HEA proved to be most efficient in modelling and predicting seasonal dynamics of chlorophyll-a and algal biovolumes. A sensitivity analysis conducted for Lake Penang revealed that algal abundance is not only driven by physical and chemicals characteristics of the water body but also by impact of inorganic substances in the water that contributes to high level of chemical oxygen demand in the water as well.

In addition, this study has successfully implemented a new process model from Law et al. (2009) consisting algal growth, algal grazing, zooplankton growth and zooplankton mortality functions into the SALMO-OO simulation library. Combination of this new process models were tested on dataset from Lake Kenyir, Lake Penang, Saidenbach Reservoir and Roodeplaat Dam within the simulation library to discover the optimal model structures for respective water bodies. Even though the new process model was not selected in complete totality as the optimal model structure for any of the test lakes, the addition has added another alternative for water body simulation in SALMO-OO process library.

Based on these forecasting results, both SALMO-PLUS and HEA have showed potential for utilisation in early warning and strategic control of algal blooms in tropical freshwater lakes. The generic nature of HEA forecast model was also observed when tested for forecasting algal biovolume for merged data of similar lake ecosystem category. Results from merged Lake Kenyir and Lake Penang data showed reasonable accuracy in predicting the timing and magnitudes of algal blooms up to 7-days-ahead. Addition of the new process model from Law et al. (2009) into the SALMO-PLUS simulation library has also expanded the alternative for lake category simulation to give a more comprehensive decision support tool for lake and reservoir management. This study has also affirmed the generality and flexibility of SALMO-PLUS for usage in tropical lakes modelling. SALMO-PLUS was observed to be capable of simulating simultaneous seasonal fluctuations in algal growth and nutrients (phosphate and nitrate) making it valuable for forecasting the impacts of various simulated scenarios for various lake management regimes.
Declaration

I declare that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Libraries, being made available in all forms of media, now or hereafter known.

Mohd Yusoff Ishak
Acknowledgements

In presenting this thesis, I have received generous help from numerous individuals whom I should like to thank.

I would like to express my gratitude to my principal supervisor, A/P Friedrich Recknagel for his patience with me and mentoring throughout my years as a PhD student and for giving me the opportunity to carry out the research. Thanks also to both my second supervisor (Dr. David Lewis) and external supervisor (Prof. Fatimah Md Yusoff) for their support and helpful tips.

I would also like to thank the following people for their guidance and contribution to my PhD project:

To Dr Hongqing Cao, Byron Zhang, Jeffrey Newman and Samuel Martin for their valuable inputs during SALMO-OO and HEA experiments.

To the Malaysian government for their funding of my tuition fee, the Universiti Putra Malaysia and the local authorities for respective water bodies (Putrajaya Corporation, Perbadanan Bekalan Air Pulau Pinang and Tenaga Nasional Berhad) studied in this project as well as Meteorological Department, Institute of Environment and Water Resource Management (IPASA) and Department of Environment Malaysia. I would like to extend thanks to the many people in Malaysia and elsewhere, who so generously contributed to the work presented in this thesis. The extensive results presented herein are due in part to their contributions.

To my fellow postgraduate colleagues; Young-Kil Kim, Grace Wai, Nor Azman Kasan, Pranay Sharma, Philip Northeast, Andrew Barker, Peter Ward and Rebecca Duffield.

Finally, and by no means least, I would like to express my heartfelt thanks to my wonderful wife, Siti Shafikah Md Daud, who has stood by me throughout the stresses and difficulties that being a graduate student brings. Thanks also to my parent and siblings.
List of Tables

Table 1 General characteristics of study lakes .......................................................... 43

Table 2 List of parameters measured *in situ* and *ex situ* in Lake Putrajaya .......... 44

Table 3 List of physico-chemical and biological parameters measured *in situ* and *ex situ* in Lake Kenyir and used for this study ......................................................... 48

Table 4 List of physico-chemical and biological parameters measured *in situ* and *ex situ* in Lake Penang and used for this study ......................................................... 51

Table 5 State variables and processes represented by SALMO-OO (from Recknagel *et al.* (2008a)) ............................................................................................................. 56


Table 7 Three lake categories and their best performing structure of the algal mass balance equations of SALMO-OO (from Recknagel *et al.* (2008a)) ...................... 57

Table 8 Input data required for running the SALMO-OO model ............................... 58

Table 9 Daily input data used for HEA in this study ................................................. 77

Table 10 Parameter settings of HEA for rule model discovery (Cao *et al.*, 2006b) .... 79

Table 11 List of input data required for running the DYRESM model ....................... 81

Table 12 Spectrum of environmental conditions of the lakes chosen for model identification by SALMO-PLUS ................................................................. 125

Table 13 Sources of process models included in the library of SALMO-PLUS ......... 126

Table 14 List of parameters used for multi objective parameter optimisation in SALMO-PLUS ........................................................................................................... 128

Table 15 Expert-based pre-set and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Lake Kenyir ............................................. 129

Table 16 Expert-based pre-set, optimised and interactively fine-tuned parameter values achieved by SALMO-PLUS for Lake Kenyir; X1=diatoms, X2=green algae and X3=blue-green algae .................................................. 130
Table 17 Expert-based pre-set and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Lake Penang ........................................... 135

Table 18 Expert-based pre-set, optimised and interactively fine-tuned parameters values achieved by SALMO-PLUS for Lake Penang; (X1=diatoms, X2=green algae, X3= blue-green)......................................................................................................... 136

Table 19 Expert-based pre-set and interactively fine-tuned initial values for state variables of Saidenbach Reservoir for SALMO-PLUS simulation .................. 142

Table 20 Expert-based pre-set and fine-tuned optimised parameter values achieved by SALMO-PLUS for Saidenbach Reservoir (X1=diatoms, X2=green algae, X3= blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)................................................. 142

Table 21 Reference and interactively fine-tuned initial values for state variables of Roodeplaat Dam for SALMO-PLUS simulation........................................... 148

Table 22 Initial values for parameters and optimised parameter values achieved by SALMO-PLUS for Roodeplaat Dam (X1=diatoms, X2=green algae, X3= blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)................................................. 148

Table 23 Reference and interactively fine-tuned initial values for state variables of South Para Reservoir for SALMO-PLUS simulation................................. 154

Table 24 Initial values for parameters and optimised parameter values achieved by SALMO-PLUS for South Para Reservoir (X1=diatoms, X2=green algae, X3= blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)................................................. 155

Table 25 List of present process model structure in SALMO-OO showing addition of Law et al. (2009) (symbol as model E)................................................................. 159

Table 26 SALMO-OO algal and zooplankton processes formulation.............................. 161

Table 27 List of parameter definitions and unit specific to the new process model with their original values retained for this experiment .................................................... 163

Table 28 Example of a matrix used to perform the test with new process model in SALMO-OO on Saidenbach Reservoir present best structure (B1-D2-A3-D4) ...... 164

Table 29 Four lake categories and their best performing model structures for algal processes identified by SALMO-PLUS .................................................................. 180

Table 30 Comparison on parameters value selected for multi objective parameter optimisation and values after optimisation for Lake Kenyir (X1=diatoms, X2=green algae, X3=blue-green algae) ...................................................................................... 182

Table 31 Comparison on parameters value selected for multi objective parameter optimisation and values after optimisation for Lake Penang. (X1=diatoms, X2=green algae, X3=blue-green algae) ...................................................................................... 183
Table 32 Initial values for diatoms, green algae and blue-green algae in Lake Kenyir and Lake Penang used for this research.................................................................185

Table 33 Summary of process model combinations selected for further assessment from experiment with lakes of different trophic state and mixing conditions. A=Benndorf and Recknagel (1982); B=Park et al. (1974); C=Hongping and Jianyi (2002); D=Arhonditsis and Brett (2005); E=Law et al. (2009).................................187

Table 34 List of process models in the simulation library of SALMO-OO ..............189

Table 35 List of $r^2$ and RMSE values obtained from HEA model experiment........191
List of Figures

Figure 1 Study site: locations of Lake Penang, Lake Putrajaya and Lake Kenyir........42

Figure 2 The description of the study site in Lake Putrajaya (from PJC, pers. comm., 2000). Data from sampling location PLg2 was used for this research .......................46

Figure 3 The description of the study site in Lake Kenyir (from Yusoff (1995)). Data from sampling station number 3 was used for this research........................................49

Figure 4 The description of the study site in Lake Penang (after Makhlough, 2008). Data from sampling location A2 was used for this research. Mengkuang Reservoir is the official name for Lake Penang .................................................................52

Figure 5 Structure diagram of the model SALMO-OO illustrating all the state variables and key processes simulated by the model (from Walter et al. (2001) ........54

Figure 6 PSO using gbest neighbourhood topology showing fully connected neighbourhood relation (from Coelho and Mariani (2009)) ........................................60

Figure 7 Flow chart in PSO approach (from Coelho and Mariani (2009))............62

Figure 8 Conceptual framework of the SALMO-PLUS simulation system (from (Martin & Recknagel, 2010)).......................................................................................63

Figure 9 Snapshot from computer showing the Java version ................................64

Figure 10 Graphical user interface (GUI) for the SALMO-OO simulation library (from : http://ecolinfo.ees.adelaide.edu.au:8080/SALMO-OO/index.html).............65

Figure 11 Visual output of simulation results within the SALMO-OO application (from : http://ecolinfo.ees.adelaide.edu.au:8080/SALMO-OO/index.html)..............66

Figure 12 Flowchart of the hybrid evolutionary algorithm (from Cao et al., 2006a) .76

Figure 13 Conceptual diagram of the hybrid EA for the discovery of predictive rule sets in water quality time-series (Cao, et al., 2006b)..............................................78

Figure 14 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya ............................................................................................................84

Figure 15 Structure of the Chl-a model for Lake Putrajaya a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE-branch ..85

Figure 16 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Putrajaya for the years 2003 to 2008 .................................................................86

Figure 17 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya (experiment 2).................................................................87
Figure 18 Structure of the Chl-a model for Lake Putrajaya (experiment 2) a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE-branch (note: values in the legend box are the range of input data) ........................................ 88

Figure 19 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Putrajaya for the years 2005 to 2008 (experiment 2) ................................................................. 89

Figure 20 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Kenyir .......................................................................................................................... 91

Figure 21 Structure of the model of algal biovolume for Lake Kenyir a) input sensitivity of the THEN-branch b) input sensitivity of the ELSE-branch ..................... 92

Figure 22 Validation of the 7-days-ahead forecasting model for algal biovolume (µm$^3$) in Lake Kenyir for 1992 ............................................................................................. 94

Figure 23 Data division and seasonal distribution by the HEA model for algal biovolume (µm$^3$) in Lake Kenyir for 1992 ................................................................. 94

Figure 24 Seasonal data division by the HEA model for algal biovolume (µm$^3$) showing monsoon seasons in Lake Kenyir for 1992 ......................................................... 95

Figure 25 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Penang .................................................................................................................. 96

Figure 26 Structure of the Chlorophyll-a model for Lake Penang a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE-branch ................................. 97

Figure 27 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Penang for 2005 and 2006 .............................................................................. 99

Figure 28 Data division for the rule set in Lake Penang for 2005 and 2006 showing the input data are divided into two parts depending on the Secchi depth value. ....... 100

Figure 29 Data division seasonal distribution by the model with 7-day-delay input data in Lake Penang for 2005 and 2006 ................................................................. 100

Figure 30 Frequency of input variable selection by HEA during 100 bootstrap runs for algal biovolume prediction for Lake Penang ......................................................... 101

Figure 31 Structure of the algal biovolume model for Lake Penang a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE-branch ..................................... 102

Figure 32 Validation of the 7-days-ahead forecasting model for algal biovolume in Lake Penang for the years 2005 to 2006 ................................................................. 103

Figure 33 Frequency of input variable selection by HEA during 100 bootstrap runs for merged data of lakes Penang and Kenyir ................................................................. 105
Figure 34 Structure of the model for algal biovolume from merged Lake Kenyir and Lake Penang input data a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE branch .................................................. 106

Figure 35 Validation of the 7-days-ahead forecasting model for algal biovolumes in Lake Kenyir for 1992 developed by experiment 1 .............................................................................. 107

Figure 36 Validation of the 7-days-ahead forecasting model for algal biovolumes in Lake Penang for 2005/2006 developed by experiment 1 .......................................................... 108

Figure 37 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Kenyir .................................................................................................................. 109

Figure 38 Structure of the model of algal biovolume for Lake Kenyir a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE branch (note: values in the legend box are the range of input data) ......................... 110

Figure 39 Validation of the 7-days-ahead forecasting model for algal biovolume concentration in Lake Kenyir for the year 1992 ........................................................................... 111

Figure 40 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Penang 2005/2006 ........................................................................................................ 112

Figure 41 Structure of the model of algal biovolume for Lake Penang a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE branch (note: values in the legend box are the range of input data) ......................... 113

Figure 42 Validation of the 7-days-ahead forecasting model for algal biovolume in Lake Penang for 2005 and 2006 ............................................................................................... 114

Figure 43 Frequency of input variable selection by HEA during 100 bootstrap runs for merged electronically measured data of lakes Penang and Kenyir................................. 116

Figure 44 Structure of the model for algal biovolume from merged Lake Kenyir and Lake Penang electronically measured input data a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE branch .......................................................... 117

Figure 45 Validation of the 7-days-ahead forecasting model for algal biovolumes in Lake Kenyir for 1992 developed by experiment using electronically measured input data only ........................................................................................................ 119

Figure 46 Validation of the 7-days-ahead forecasting model for algal biovolumes in Lake Penang for 2005/2006 developed by experiment using electronically measured input data only ........................................................................................................ 119

Figure 47 Frequency of electronically measured input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya ................................................................. 121

Figure 48 Structure of the Chl-a model using electronically measured input data for Lake Putrajaya a) input sensitivity according to the THEN-branch, b) input sensitivity
according to the ELSE-branch (note: values in the legend box are the range of input data).......................................................................................................................................................................................... 122

Figure 49 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Putrajaya for the years 2005 to 2008 using electronically measured input data....... 123

Figure 50a Simulation results of SALMO-PLUS for PO₄-P and NO₃-N concentrations in Lake Kenyir (1992). X-axis in days; ---- = simulated epilimnion; --- = simulated hypolimnion; o = measured epilimnion ................................................................................................................................. 131

Figure 50b Simulation results of SALMO-PLUS for Total phytoplankton and Algal functional groups biomass in Lake Kenyir (1992). X-axis in days; --- = simulated epilimnion; --- = simulated hypolimnion; For algal functional groups simulation, ---- = Diatoms; - - - = Green Algae; x = Blue-green Algae ........................................................................................................ 132

Figure 51a Simulation results of SALMO-PLUS for PO₄-P and Chlorophyll-a concentrations in Lake Penang (2005). X-axis in days; --- = simulated epilimnion; - - - = simulated hypolimnion; o = measured epilimnion ......................................................................................................................................... 137

Figure 51b Simulation results of SALMO-PLUS for Total phytoplankton and Algal functional groups biomass in Lake Penang (2005). X-axis in days; --- = simulated epilimnion; --- = simulated hypolimnion; For algal functional groups simulation, ---- = Diatoms; - - - = Green Algae; x = Blue-green Algae ........................................................................................................ 139

Figure 52a Simulation results of SALMO-PLUS for PO₄-P and Total phytoplankton in Saidenbach Reservoir (1975), X-axis in days; — = simulated epilimnion, --- = simulated hypolimnion, o = measured data.............................................................................................................................................. 143

Figure 52b Simulation results of SALMO-PLUS for zooplankton biomass in Saidenbach Reservoir (1975); X-axis in days; --- = simulated epilimnion, --- = simulated hypolimnion, o = measured data.............................................................................................................................................. 145

Figure 53a Simulation results of SALMO-PLUS for PO₄-P and NO₃-N in Roodeplaat Dam (2003), X-axis in days; --- = simulated epilimnion, --- = simulated hypolimnion, o = measured data.............................................................................................................................................. 149

Figure 53b Simulation results of SALMO-PLUS for Total phytoplankton and Chlorophyll-a concentrations in Roodeplaat Dam (2003) X-axis in days; -- = simulated epilimnion; ---- = simulated hypolimnion; o = measured epilimnion ................................................................. 151

Figure 54a Simulation results of SALMO-PLUS for PO₄-P and O₂ concentrations in South Para Reservoir (2008) X-axis in days. --- = simulated epilimnion; ---- = simulated hypolimnion; o = measured ......................................................................................................................... 156

Figure 54b Simulation results of SALMO-PLUS for Total phytoplankton and zooplankton in South Para Reservoir (2008). X-axis in days --- = simulated epilimnion; ---- = simulated hypolimnion; o = measured ......................................................................................................................... 158
Figure 55 Snapshot from SALMO-OO model selection page showing the new process model from Law et al. (2009), (symbol as model number 2) and the new category for mesotrophic tropical-stratified lake. ................................................................. 160

Figure 56 Lake Kenyir (1992) simulation results from the SALMO-OO optimised model structure and combination with zooplankton mortality process model from Law et al. (2009) for PO₄-P and NO₃-N concentrations. X-axis in days; — = simulated epilimnion; ---- = simulated hypolimnion, o = measured data. For algal functional groups: : — = Diatoms; - - - = Green Algae; x = Blue-green Algae. .......................... 166

Figure 57 Simulation results of SALMO-OO for PO₄-P, Chlorophyll-a, Total phytoplankton and algal functional groups in Lake Penang (2005). X-axis in days. — = simulated epilimnion; ---- = simulated hypolimnion; For algal functional groups simulation, — = Diatoms; - - - = Green Algae; x = Blue-green Algae................. 170

Figure 58 Simulation results of SALMO-OO for PO₄- total phytoplankton biomass, zooplankton and algal functional groups in Saidenbach Reservoir (1975). X-axis in days; — = simulated epilimnion; o = measured data; For algal functional groups simulation, — = Diatoms; - - - = Green Algae; x = Blue-green Algae. .................. 173

Figure 59a Simulation results of SALMO-OO for PO₄-P and NO₃-N concentrations in Roodeplaat Dam (1992). X-axis in days, — = simulated epilimnion; ---- = simulated hypolimnion, o = measured data ................................................................. 176

Figure 59b Simulation results of SALMO-OO for algal functional groups in Roodeplaat Dam (2003). X-axis in days, — = simulated epilimnion; o = measured data. For algal functional groups simulation (bottom, right), — = Diatoms; - - - = Green Algae; x = Blue-green Algae. ........................................................................ 177
CHAPTER 1

1.0 INTRODUCTION

1.1 General Introduction

In many developing countries currently populations and industries are growing fast with increasing demand for water both for drinking, irrigation, hydropower and industries. However, the required water typically cannot be covered by natural freshwater bodies but requires the construction of multipurpose water reservoirs. Whilst these man-made lakes contribute largely to the water supply in these countries, they often lack sustainable management strategies. The construction of a new reservoir can destabilize its surrounding aquatic environment for a number of years and seriously impact aquatic wildlife (Morley, 2007).

Eutrophication is ‘the enrichment of water by nutrients especially compounds of nitrogen and phosphorus, causing an accelerated growth of algae and higher forms of plant life to produce an undesirable disturbance to the balance of organisms and the quality of the water concerned’ (Ansari et al., 2011). Nutrient over-enrichment of water by urban, agricultural and industrial development has promoted the excessive growth of algal and produces serious and far-reaching environmental consequences. Nutrient over-enrichment can be classified into two broad categories: 1) ‘point source’ pollutants originating primarily from inadequate industrial waste, municipal waste and waste-water treatment and 2) ‘non-point source’ pollutants originated from surface and sub-surface run-off from the land surface area. Eutrophication can lead to cyanobacteria blooms that can contaminate freshwater by toxins, anoxia, taste and odour causing high treatment costs in water works, risks for public health and degradation of freshwater ecosystem (Bobbin & Recknagel, 2003). Eutrophication and algal blooms are in particular serious problems in tropical regions, where all metabolic processes are faster and financial resources for implementing adequate monitoring programs often deficient (Sperling et al., 2006).

As Malaysia continues its economic development, there is the threat of environmental degradation which may become worse by global warming. Studies by NAHRIM (2007) have shown that rapid developments of dams and reservoirs in Malaysia have subjected aquatic ecosystems to various levels of stress and possible degradation mainly in terms of water quality and sedimentation. Since Malaysia is already vulnerable to extreme climate events such as typhoons, droughts and floods (Chou, 1992), climate change is expected to exacerbate these vulnerabilities by inducing more weather extremes such as intense precipitation in the wet period and a lack of rainfall in the dry period that would lead to more severe floods and longer droughts (Dueñas, 2007).

Fortunately, the build-up of eutrophication and the formation of algal blooms in water bodies can be measured and calculated by management tools such as models. The use of predictive models can help to minimise the gap on sparse limnological knowledge of tropical water bodies. Modelling techniques can: (1) indicate research requirements for better understanding of ecosystem processes in man-made tropical lakes, and (2) determine eutrophication control measures in man-made tropical lakes.
Previous attempts at applying temperate zone control measures for eutrophication in inter-tropical regions was shown to be largely unsuccessful (Rast & Thornton, 1996).

1.2 The Computational Modelling Approach of This Study

Proactive management of inland waters requires sound ecological knowledge to assess the productivity of lakes, which permits predictions on the future trends under changing conditions due to any perturbation. Predictive modelling based on ecological knowledge and data has been well established in assisting informed management of eutrophication and algal blooms (Ferguson, 1997). Ecological modelling allows either a prediction in terms of future states of the system or elicits greater understanding of how the system functions and the driving forces behind and interactions within the system (Whigham & Fogel, 2003).

Highly dynamic tropical lakes in particular require predictive models to reveal ecological knowledge and provide early warning on imminent water quality threats. Both understanding causal ecological relationship and early warning are needed to establish best management of eutrophication and algal blooms in tropical lakes.

Models

The use of predictive models to quantify the development of algal blooms in relation to changes in an environment was formally recommended in 1990 (NRA, 1990). However, the wide use of ecological models in environmental management started around 1970, when the first eutrophication models emerged and very complex river models were developed (Jørgensen et al., 1996). Models are invaluable tools at studying system dynamics, generalizing discrete observations and predicting future states. Accurate and well validated models are able to provide predictions of the behaviour of dynamically changing systems and provide data and insights that would be difficult or impossible to obtain by ‘traditional’ field and laboratory methods (Howard, 1997). Lake ecological studies utilising models can discover patterns from data sets and serve as valuable tools for researchers and aquatic managers for better understanding and decision making. Thereupon, numerous models have been developed and applied as tools in managing and maintaining quality and quantity of water bodies.

In this study, the data-driven modelling technique HEA (Hybrid Evolutionary Algorithms) (Cao et al., 2006a) and the process-based lake simulation library SALMO-OO (Simulation by means of Analytical Lake MOdels) (Recknagel et al., 2008a) were used for assessing and predicting eutrophication and algal blooms in the three tropical reservoirs in Putrajaya, Kenyir and Penang in Malaysia. An attempt was also made to utilise a one-dimensional hydrodynamic numerical model DYRESM (DYnamic REServoir Simulation Model) which is freely available from the internet (http://www.cwr.uwa.edu.au) for validating the vertical distribution of temperature, salinity and density in Lake Putrajaya.

HEA is used for rule-set discovery and optimisation for forecasting algal population dynamics and growth elucidated from time series data sets (Recknagel, et al., 2006). HEA can also reveal patterns of algal for seasonal and annual succession in lakes and later predict the possibility of algal bloom. SALMO-OO is an object-
oriented implementation of mass balance equation for a typical pelagic food web (diatoms, green algae, blue-green algae and cladocerans) and nutrient cycles (PO$_4$-P, NO$_3$-N, detritus) as well as alternative process models for algal growth and grazing, zooplankton growth and mortality (Recknagel et al., 2008a). This study implemented SALMO-OO to better understand the short and long term ecosystem dynamics of tropical lakes by utilising and embedding all available hydrological and ecological data from the study lakes. The SALMO-OO model can also be applied for simulation and scenario analysis with different management regime for lakes. These two methods are integrated for analysing algal population dynamic and together with water quality to understand the causal-effect relationship of algal bloom in the tropical lakes. Apart from that, this research also attempted to utilise DYRESM for simulating the thermal structure, salinity and density in Lake Putrajaya.

1.3 Objectives and Achievements of This Study

The modelling concepts HEA, SALMO-OO and DYRESM were applied for the first time to three tropical reservoirs in Malaysia in order to:

1. test and validate the models for tropical lakes
2. improve understanding on ecology of tropical lakes and their response to eutrophication management
3. identify best management practice for eutrophication control of the three tropical lakes, and assist in informed decision making

A better understanding will be a key factor in improving the future management plan for the tropical lakes and will provide further assurance for their success. The aim of this study was to provide forecasting tools for eutrophication and algal bloom dynamics of man-made lakes in Malaysia and consequently answer the following questions:

1. Can the models HEA and SALMO-OO be validated for tropical lakes with different morphometry and trophic levels with regards to food web dynamics, nutrient cycles and phytoplankton growth?

Evaluation of the extent and causal relationship of eutrophication continues to attract interest from water experts worldwide. This study investigated eutrophication in tropical lakes by means of HEA and SALMO-OO and demonstrated whether both models can simulate the growth of phytoplankton reasonably well and produce acceptable results. The outcomes from both models have provide additional foundation for the successful implementation of the Catchment Development Management Plan (CDMP) for respective study sites in order to protect the integrity of lake ecosystem.
2. Can a generic model structure of SALMO-OO be found for tropical-stratified mesotrophic lakes?

The key characteristics of SALMO-OO as a generic model that allows the simulation of different lake conditions using the same model structure was tested for tropical lakes. Structure that performs the best in terms of predicting algal growth for tropical lakes conditions was identified by submitting combinations of alternative phytoplankton growth and grazing process models within the simulation library of SALMO-OO for testing in this study. This best structure has confirmed the flexibility and generality of SALMO-OO in predicting algal growth for lakes beyond its conventional domain.

3. Can a generic predictive rule-set be developed by HEA for algal growth in tropical lakes with different morphometry and trophic states?

A generic rule set model is a standard rule which can facilitate model adaptation to a particular lake ecosystem category. The advantage of utilising generic rules are such as reducing the time and money required to develop and maintain several different models for different study sites and can be applied to several different sites with slight adaptation. This study has managed to develop a generic predictive rule-set for algal growth in tropical lakes with different morphometry and trophic by means of HEA. Generic models are also helpful when forecasting for lakes with inadequate data.

4. Can the model DYRESM be validated for dynamic, shallow-polymictic tropical lake such as Lake Putrajaya?

DYRESM was applied to simulate the hydrodynamic of Lake Putrajaya in this study. This model has rarely been applied to tropical water bodies in Malaysia. The main objective for applying DYRESM was to determine if the model could be used as an effective management tool for water quality in Lake Putrajaya. Prediction of thermal structure for Lake Putrajaya by means of DYRESM was expected to provide guidance for lake authority in simulating the changes in thermal regime of the lake in response to climate change and predict its future behaviour.

Considering these promising results, it is important to be cautiously optimistic by pointing out that the lake data of good quality plays a crucial role in the success of these models. It is hoped that this promising results will spur further limnological research, investment and capacity buildings from academic and relevant government research institutions to meet this huge demand of modelling eutrophication and algal bloom formation in tropical lakes. The current scenarios regarding data availability and treatment specific for tropical lakes cases in this research as well as challenges researchers needs to overcome are explained in the discussion and conclusions chapters.
1.4 Thesis Overview

Following this introduction chapter, a literature review is presented in Chapter 2. The literature review presents the background descriptions, categories, approaches and evolution of lake ecological models. Details of approaches related to this study namely data driven (HEA) and differential equation approach (SALMO-OO) are also given. The descriptions on tropical lakes ecosystems with the gaps and advancement made in the field together with recent trends in ecosystem modelling were also highlighted. Chapter 3 presents the field methods and instrumentation used to investigate the limnological processes in Lake Putrajaya, Lake Kenyir and Lake Penang. Details of the models (HEA and SALMO-OO) are also given.

Results from this study are given in Chapter 4. The first section of Chapter 4 presents the results from application of HEA approach to Lake Putrajaya, Lake Kenyir and Lake Penang. Chlorophyll-a and algal biovolume were used as an output for HEA in Lake Putrajaya and Lake Kenyir respectively while both Chlorophyll-a and algal biovolume was used as an output for Lake Penang HEA experiment. This was followed by results from generic forecasting model for algal biovolume in stratified mesotrophic lake generated from merged time series ecological data of Lake Kenyir and Lake Penang.

The second section of Chapter 4 presents the results from application of SALMO-OO to Lake Putrajaya, Lake Kenyir and Lake Penang. These results were based upon successful completion of implementing particle swarm optimisation (PSO) into SALMO-PLUS. Additional results from Sadenbach Reservoir, Roodeplaat Dam and South Para Reservoir were also elaborated. This was followed by results from extension of SALMO-OO simulation library by additional process model to simulate and predict the lake ecosystem processes regarding algal succession and interaction between zooplankton and different algal groups. Results from Lake Kenyir and Lake Penang together with Sadenbach Reservoir and Roodeplaat Dam were presented.

These results are discussed in Chapter 5 with first sections elaborating on modification made to SALMO-OO and SALMO-PLUS for modelling tropical lakes. The success and transferability of HEA model as tools for tropical lake ecosystem analysis are discussed next. This is followed by explanations on the limitations and challenges for modelling tropical lakes. This chapter is concluded with recommendations for future studies.
CHAPTER 2

2.0 LITERATURE REVIEW

The main objective for this research were to 1) validate the models HEA and SALMO-OO for tropical lakes, 2) find a generic model structure of SALMO-OO for tropical lakes and 3) use HEA to develop a generic predictive rule-set for algal growth in tropical lakes and 4) to validate the model DYRESM for tropical lake.

This literature presents on lake ecosystems descriptions and assesses the attempts to model lake ecosystems by means of lake ecological models. Categories and approaches adopted by lake modellers when attempting to model lake ecosystem were also explained. Examples of approaches related to this study namely data driven (HEA), process-based approach (SALMO-OO and DYRESM) are illustrated respectively. The first part of this literature review discusses data driven modelling approaches with special consideration of hybrid evolutionary algorithm (HEA) that allow model synthesis and knowledge discovery. The second part analyses process-based modelling approaches with special consideration of the lake simulation library SALMO-OO that allows novel structure and parameter optimisation for improving model validity. This is followed by descriptions on tropical lakes ecosystems with emphasis of limnological research in Malaysia and gaps observed in the tropical lake research. Recent trends in ecosystem modelling are also presented.

Scientists deal with complex problems of translating environmental and ecological phenomenon into a set of mathematical equations by means of models. The subject of ecological modelling covers a broad series of model such as population dynamics, eutrophication and eco-toxicology. This field has developed rapidly during the last two decades due to three factors (Jørgensen & Bendoricchio, 2001):

1. the development of computer technology, which has enabled us to handle very complex mathematical operating procedures;
2. a wide-ranging understanding of pollution problems, including the knowledge that a complete elimination of pollution is not feasible (zero discharge), but that proper pollution control with the limited economical resources available requires serious consideration of the influence of pollution impacts on ecosystems;
3. our understanding of environmental and ecological problems has increased significantly; in particular, we have gained more knowledge of quantitative relationships in the ecosystems and between ecological properties and environmental factors.

Applications of both modelling approaches to freshwater lakes are then compared in terms of their possibilities and limitations for prediction and elucidation of short-term and long-term dynamics of the plankton community and water quality. The two modelling approaches were expected to provide improved understanding and decision support for eutrophication and algal bloom management of tropical lakes in Malaysia.
The following section outlines briefly the progress of the ecological modelling discipline, with particular interest in eutrophication aspects of tropical lakes. Advances in the models applied for this research are also discussed.

2.1 Lake Ecosystem and Lake Models

2.1.1 Lake Ecosystem

Lake ecosystems are subject to spatial and temporal variability which results in a high degree of uncertainty in relation to phytoplankton assemblages. In terms of the responses of phytoplankton to changing lake environments, deep lakes are slower to respond than shallow ones (Reynolds, 1987). Food web dynamics and nutrient cycling in lakes are basically driven by primary productivity that depends on chemical (nutrients, particularly phosphorus), physical (temperature, underwater light) and biological (composition and abundance of zooplankton) conditions that are also regulated by hydraulic exchanges and surface level fluctuations. Phytoplankton assimilates carbon by photosynthesis depending on seasonality of the ambient nutrient level, temperature and underwater light intensity. Whilst temperature and light conditions are predictable functions of latitude and altitude, nutrients are highly variable in terms of sources, loadings and cycling, and specific for each single lake. Even at optimal physical conditions, the production reaches a ceiling imposed by the supply of carbon or the availability of nutrients (Reynolds, 1999). Plankton community evolution is not simply predictable from the sum of the basic underlying processes (phytoplankton growth, nutrient limitation, grazing), rather it is an emergent property driven by competition, predation, and environmental forcing (Allen & Polimene, 2011). Apart from grazing by zooplankton, physiological death, import, export and sedimentation (sinking) represent further loss processes of phytoplankton biomass in an ecosystem. The development and successive phases of the sciences of inland water encompassing limnology, hydrobiology and freshwater biology is given by Talling (2008).

The morphometry of a lake and its hydrological fluctuations strongly influence variability patterns and community structure of phytoplankton. Therefore, it is difficult to explain why certain species are abundant in a particular lake at a certain time and are replaced by other species at different stages in the development of this lake (Reynolds, 1987). It is possible that a species best suited to exploit resource-saturated environments is likely to be that which arrives first and reproduces fastest with a high reproductive rate rather than being in efficient use of available resources (Reynolds, 1987). The ability of any phytoplankton population to increase depends on its capability to sustain a greater growth rate than loss rate. In certain circumstances the greater growth rates often lead to undesirable concentrations known as blooms. Blooms often occur after a series of hydrological changes, enhanced nutrient inputs and increased thermal stratification that favour proliferation of algal biomass. A classification trophic scheme was proposed by Naumann in 1929 for lake water types based on production of organic matter by phytoplankton as well as primary physical and chemical determinants (Wetzel, 2001). The terms “eutrophic”, “mesotrophic” and “oligotrophic” were proposed for lake water as early as 1907
describing plant growth induced by nutrient supply and provides a classification scheme according to plant production and nutrient availability (Kitsiou & Karydis, 2011). This scheme was modified by OECD (1982) and water bodies are now classified as ultra-oligotrophic, oligotrophic, mesotrophic, eutrophic and hypertrophic based on phosphorus loading, algal densities and water transparency.

An algal bloom is a rapid increase or accumulation in the population of algae in an aquatic system (Wu & Xu, 2011). Algal blooms are globally considered as nuisances (Paerl, 1988; Paerl & Huisman, 2009) and an acute risk for public health after consuming or being exposed to affected lake water (Chorus & Bartram, 1999). The evolution of plankton communities under changing environmental conditions is a complex phenomenon generated from non-linear interactions between abiotic (temperature, pH, light environment, nutrient supply, contaminant exposure etc.) and biotic (physiological responses, predator–prey interactions) components (Allen & Polimene, 2011). Numerous data processing techniques have been proposed for eutrophication assessment (see de Jonge, Elliott, & Orive, 2002; Kitsiou & Karydis, 2011) and studies of algal blooms by means of models are now becoming classical (Truscott & Brindley, 1994). The causality of algal dominance, succession and bloom formation is highly complex and highlights the need for tools to unravel this complexity and make algal blooms events predictable. The complexity makes the phytoplankton dynamic difficult to forecast and the development of computational predictive models a complicated and challenging issue. The essential features of natural ecosystems complexity can be represented in an ecosystem models.

2.1.2 Ecological Lake Models

Many ecosystem models have been developed during the last few decades and used in management of renewable natural resources, conservation biology and assessments of ecological risks posed by nutrients, toxic chemicals and other stressors. Ecological models for lakes can depict the interactions and changes of environmental elements and simulate the dynamics of spatial and temporal patterns in lake ecosystems. These models range from the simplified input-output models to 1-dimensional and complex 3-dimensional models (Imboden, 1974; Ahlgren et al., 1988; Bonnet & Wessen, 2001; Naithani et al., 2007; Wu & Xu, 2011). The models need to reflect the dynamic nature and succession of ecosystems in order to be used for predicting ecosystem behaviour in response to environment, climate and management changes. The most basic ecological models are concerned with the behaviour of a single species, group of species or community and their variation in time while the most complex models attempt to capture the spatial and temporal patterns of an ecosystem at several levels of description (Whigham & Fogel, 2003). The validation of a model is not that it is “true” but that it generates good testable hypothesis relevant to important problems (Levins, 1966).

There are at least three good reasons for modelling aquatic systems, i.e. management, prediction and better understanding of the system (Atanasova et al., 2005). One of the main advantages of applying models lies in predicting future changes (e. g. climate change) or testing management options (e. g nutrient reduction) (Blenckner, 2008). A well-developed lake ecosystem model can be a useful decision
tool for water authorities, provided that it is robust and reliable. van Tongeren (1995) compared the use of regression, ordination and dynamic ecosystem modelling approaches in limnology and proposed to embed regression models into a dynamic ecosystem model by creating ‘nested models’. Gaining knowledge about the future states of lakes will allow managers to conduct the cost and benefit analysis in order to improve water quality to an acceptable level. In addition, models utilised for lake ecological studies can discover patterns in data sets leading to better understanding and decision making (NRA, 1990). This would enhance strategic planning and decision making for natural resources management by attaining the best information possible through these models.

Ecological modelling is crucial for gaining an understanding of ecosystems and achieving their effective management. Bulgakov et al. (2011) reviewed methods of ecological forecast and described models classification based on:

1) the degree of system homogeneity (point or distributed)
2) the period of forecast (short-term, medium-term and long-term)
3) the level of detail and
4) the qualitative or quantitative methods of forecast used

Based on the three criteria of model validity, Levins (1966) distinguishes between three model categories:

**Type-I**: sacrifice generality to realism and precision. These are models that simulate the precise input and responses but lose general applicability. Modeller of this type works by reducing the parameters to those relevant to the short-term behaviour of the organism, make fairly accurate measurements, solve numerically on the computer and end with precise testable predictions applicable to these particular situations.

**Type-II**: sacrifice realism to generality and precision. These are models of limited inputs to isolated processes which do not yield real-world solution. These models involve setting up general equations from which precise results may be obtained and equations from this type of model were found unrealistic. The assumption is their model is analogous to assumptions of frictionless systems or perfect gases. Many of the unrealistic assumptions are expected to cancel each other, that small deviation from realism results in small deviations in the conclusions and that in any case, the way in which nature departs from theory will suggest where further complications will be useful. In short, this model starts with precision and hopes to increase realism.

**Type-III**: sacrifice precision to realism and generality. These models are concerned with the long run qualitative rather than quantitative results and resort to very flexible models which generally assume that, for example, functions are increasing or decreasing instead of specifying the mathematical form of an equation. These models provide generalism through breadth rather than precision which means that the prediction could also be expressed as, for example, inequalities as between tropical and temperate lake species.
Lake ecological models that attempt to capture the spatial and temporal patterns of an aquatic ecosystem at several levels of description revolve around dynamic, interdependent, complex and mostly not completely understood biological processes. Major differences between ecological and lake eutrophication models are related to (Håkanson & Boulion, 2003):

1. target variables (from individual species modelling to total biomass prediction)
2. modelling scales (from daily to annual predictions)
3. modelling structures (from empirical/regression models to approaches based on ordinary and partial differential equations) and
4. driving variables (whether accessed from standard monitoring programs, online data monitoring, climatologically measurements or case specific studies)

Forbes (1887) published the first description of the lake as an ecosystem and the necessity for taking a comprehensive survey of the whole organic complex as a condition to a satisfactory understanding of any part. The same author anticipated future studies on food webs, population and community ecology and the ecosystem concept in the field of classic ecology. A brief description of the origin and evolution of population theory and in particular the prey-predator models was given by Berryman (1992). The history of ecological and environmental modelling started with static models (steady state and regression models) of Lotka-Volterra (the dynamic prey-predator relationship) (Lotka, 1925; Volterra, 1926) and Streeter-Phelps (the oxygen balance in a stream) models in the early 1920s to the dynamic models of today. The Lotka–Volterra equations, also known as the predator–prey equations, are a pair of first-order, non-linear-differential equations frequently used to describe the dynamics of biological systems in which two species interact by being one a predator population and one a prey population. It is characterised by oscillations in the population size of both predator and prey, with the peak of the predator's oscillation lagging slightly behind the peak of the prey's oscillation.

\[
\frac{dx}{dt} = x (\alpha - \beta y) \\
\frac{dy}{dt} = -y (\gamma - \delta x)
\]

(1)

Based on this model, species of predator (y) and prey (x) interact with each other in limited space, where the prey grows at a linear rate \( \alpha \) and gets eaten by the predator at the rate \( \beta \). The predator gains energy by consuming the prey at rate \( \delta \), while decreases through natural death at rate \( \gamma \). Several simplifying assumptions made in the models are:

1. the prey population will grow exponentially when the predator is absent;
2. the predator population will starve in the absence of the prey population (as opposed to switching to another type of prey);
3. predators can consume infinite quantities of prey and
4. there is no environmental complexity (both populations are moving randomly through a homogeneous environment)
Fleming (1939) model that considers the grazing of phytoplankton based on the Volterra equations has set the early works on phytoplankton productivity. Steele (1962) developed a theoretical equation for the photosynthesis-light relation which includes the effects of inhibition in intense light. O'Brien (1974) proposed an a priori model of phytoplankton population growth to simulate the dynamic of phytoplankton-nutrient interactions in lakes where the rate of uptake of nutrients by phytoplankton follows Michaelis-Menten uptake kinetics and this rate of uptake governs the rate of population growth. Riley (as cited in Collins, 1980) included light intensity effects on production as well as factors such as temperature-dependant respiration and nutrient depletion for his phytoplankton productivity model.

While Steele (1962) and Fleming (1939) have successfully predicted phytoplankton population dynamics in specific situation even though they lacked in terms of generality, the model of O'Brien (1974) showed a high degree of generality due to the fact that O'Brien (1974) model is based on a widely observed fact of phytoplankton physiology and does not draw on specific characteristics of an ocean or lake.

On the same time, researches attempt to understand the interactions between major biotic (phytoplankton, zooplankton, fish) and abiotic (nutrient, mainly phosphorus and nitrogen) component of limnetic bodies that causes eutrophication. Lake models that describe eutrophication processes were already developed for environmental management as early as in the seventies (Jørgensen, 2010). The modelling of lake eutrophication started with empirical models relating total phosphorus (TP) and chlorophyll concentrations and input–output models relating TP loading and TP concentration (Mooij et al., 2010).

A key factor in the development of different types of lake model to manage eutrophication was Vollenweider’s (as cited in Håkanson & Boullion, 2003) identification of the simple relationship between sedimentation of phosphorus and water turnover in lakes. Thomann (1977) and DiToro et al. (1975) pioneered the beginning of eutrophication model shift beyond the earlier concepts of modelling biochemical oxygen demand and dissolved oxygen. Due to the limitations of static equilibrium models, for instance to predict response times to management measures and to account for the role of sediments and, later, also food web effects, dynamic models for TP and chlorophyll were developed (Mooij, et al., 2010). More detailed model such as by Bierman et al. (1973) in which temperature effects, cell sinking, cell decomposition, nutrient recycle and predation were included, set the pace for further model developments. Next, processes such as diffusivity and vertical turbulences were incorporated into the models. An overview concerning various early models of lake eutrophication have been provided by Mooij et al. (2010).

Period of 1980s saw further refinement to the models. Ecosystem models started to have increasing number of compartments and more details behaviour of the compartment was included (Radtke & Straskraba, 1980). For example, the practise of coupling phytoplankton growth directly to the external concentration of nutrient has caused problems such as persistent time lags between modelled and observed phytoplankton peaks in both eutrophic and oligotrophic conditions. Due to the ability of phytoplankton to store nutrients in excess of their requirement, the algal population could survive for several weeks after external reserves of nutrients were depleted and
an alga may not respond at once to a change in the external concentrations of nutrients. This problem was solved by separating uptake rate from growth rate, which would permit simulation of storage of surplus nutrients (Collins, 1980). In order to represent lake stratification, models with separate layers of epilimnion, hypolimnion and sediment were also developed (Scavia, 1980).

Numerous models have been developed since and today there are hundreds of ecological models which have been used as tool in research or environmental management (Jørgensen, 1995, 1999, 2005). Field of aquatic ecosystem modelling has significantly improved since the 1970s with the development of more elaborate model structures (Jørgensen, 1995). A comprehensive database of ecological models can be found at http://dino.wiz.uni-kassel.de.html. Breckling et al. (2011) outlines major routes of development leading to the current spectrum of concepts and applications in ecological modelling. Each model used nowadays has been calibrated and verified with real-world monitoring data for the respective watershed and reservoir. The selection of the most appropriate model for a specific case study is usually a matter of which processes are most significant in the modelled situation (Jørgensen et al., 1996).

Apart from differences in complexity, models also differ in approaches. Two fundamentally different modelling approaches can be distinguished (Atanasova et al., 2006):

i. the deductive, knowledge driven approach resulting in **deterministic models**. The deductive method requires expert knowledge to build a process-based model and depends on understanding of the processes controlling the ecological system (Hoffmann, 2006), and

ii. the inductive, **data driven approach** exploring candidate models and match them with measured data resulting in empirical models. The inductive method only uses the information content of measured input and output data of the ecological system to construct a predictive model (Hoffmann, 2006).

Both methods have their own strength and weaknesses. Even though the deductive approach can be more robust since its basis is the important operating mechanism, this approach can be difficult because often we have an incomplete understanding of cause and effect of the system under study. The inductive approach can produce models that are very precise in describing the empirical output data, but they may not generalise or scale well and it can be difficult to extract causality from these models. Thus, in practise both model are used interchangeably to model ecological systems (Hoffmann, 2006).

Deterministic models are in general represented by ordinary differential equations (ODE). They are being applied to lake ecosystems since the 1970s (Jorgensen & Bendorichio, 2001). Inductive lake models such as evolutionary algorithms and neural networks are the results of applying biological-inspired computation to ecological data. These types of model rely heavily on the quality and quantity of measured data available. They have also been demonstrated to be powerful predictive tools for lake ecosystem dynamic simulations (Recknagel, 2002, 2003b).
Starfield et al. (1989) outlines the differences between a conventional deterministic system model and a qualitative rule-based empirical model where, in both conventional and rule-based models, the objective is to simulate how different biotic component of the system respond over time to changes in both the abiotic components and other biotic components. Each component of a deterministic model is represented by a numerical variable. The amount of change in a given variable over some period of time is calculated by difference or differential equation. Examples of deterministic system models are the Lotka-Volterra equations (at the theoretical level).

Ecological modelling application can be broadly classified into three categories (Recknagel, 2003a):

1. The empirical statistical approach
2. The process-based differential equation approach and
3. The empirical data-driven computational approach

This research focuses on the combination of process-based and data-driven approaches.

2.2 Data-driven Lake Models

Evolutionary computation is a discipline of algorithms that makes use of basic principles from natural evolution to evolve solutions to complex computational problems (Whigham et al., 2006). The progression of computer technology with high performance technical computing and biologically-inspired computational approach have greatly facilitated the analysis, synthesis and forecasting of the ecosystem. Bio-inspired algorithms are driven by data on the principles of neuron learning, natural selection or hierarchical inheritance in order to evolve dynamically. This method has been applied in the discipline of ecological informatics and was found to suit well the complex nature of ecological data. The major difference between natural and artificial evolution is the former does not have a predefined goal and is essentially open-ended adaptation process whereas the latter is an optimisation process that attempts to find solutions to predefined problems (Floreano & Mattiussi, 2008).

Biologically-inspired computation techniques such as fuzzy logic, cellular automata, artificial neural networks, adaptive agents and evolutionary algorithms are considered as core concepts of Ecological Informatics (Recknagel, 2002). Ecological informatics is defined as an interdisciplinary framework promoting the use of advanced computational technology for elucidating the principles of information processing at and between all levels of ecosystems (Recknagel, 2003a). The aim of ecological informatics is to address information in various levels in ecological systems in a transcending manner from genes to ecological networks. Evolutionary computation techniques, (e.g., genetic programming) offer opportunities to evolve model structures as well as model parameters explicitly and their results are more comprehensible and easier to interpret than those results generated by the black-box methods (e.g., neural networks) (Yao et al., 2006).
2.2.1 Artificial Neural Networks (ANN)

Artificial neural networks (ANN) have been developed as generalised mathematical models of the concepts and characteristics of biological nervous systems where specific problems is solved by a huge number of interconnected processing elements (called neurones). Neurons are arranged into layers. Each neuron works by receiving signals from the inputs and producing a signal for transmission to all outputs. Each layer represents a non-linear combination of non-linear functions from the previous layer. The first layer interacts with the environment to receive input (input layer) and the final layer (output layer) interacts with the output to present the processed data. Layers between the input and the output are known as hidden layers where they do not interact with the environment. This novel structure of the information processing system holds the key element of artificial neural networks paradigm. Whereas biological neurons are either inhibitory or excitatory and have the same effect on all neurons which the send signals to, artificial neurons can emit both negative and positive signals and thus the same neuron can established both negative and positive synaptic connections with other neurons (Floreano & Mattiussi, 2008). According to Lek and Guégan (1999), an ANN is a ‘black box’ approach which has great capacity in predictive modelling, i.e. all the characters describing the unknown situation must be presented to the trained ANN and the identification (prediction) is then given.

ANNs are inductive techniques that allow extraction of empirical patterns reflected by multivariate non-linear time series data (Recknagel, 2003b). Generally, patterns extraction and trends detection are too complex to be noticed by either humans or other computer techniques. ANNs are capable to overcome this due to their remarkable ability to derive meaning from complicated, abundance or imprecise data and little underlying theory exists. The data, which typically arises through extensive experimentation may be non-linear, non-stationary or chaotic and thus may not be easily modelled. However, ANNs do not require any a priori assumptions about the problem space such as distribution pattern (normal distributions for example) though the addition of such information from the input may help to speed training.

An ANN is configured for a specific problem solving application, such as pattern recognition or data classification through learning process and sample data. In learning by example, ANN is capable of learning how to do tasks based on given training data. Through this initial experience, ANN forms an adaptive learning and creates its own representation of the information it receives during learning time (known as self-organisation). Input data and their corresponding output are needed to train the ANN. Supplied with input data and training method, ANN can be used to perform different tasks depending on the training received. Once properly trained, the network provides a data-driven model which is capable of giving reasonable answers when presented with input vectors that have not been encountered during the training process. The capability to learn from a set of experimental data (e.g. processing conditions and corresponding responses) without actual knowledge of the physical and chemical laws that govern the system has enabled ANNs to predict the output of new independent input data. The key to successfully training an ANN is choosing the right network architecture and training algorithm (Shourian et al., 2008).
Since ANNs are non-linear statistical data modelling tools, application of ANNs in data treatment is especially important in nonlinearities and complex behaviour systems where it can identify and learn correlated patterns between input data sets and corresponding target values. ANNs impose fewer constraints on the functional form of the relationships between input and output variables, making them a logical choice for application when the complexity of the mapping is difficult to anticipate (Shourian et al., 2008). In addition, ANNs are capable of representing both linear and non-linear relationships and are able to learn these relationships directly from data (Anita Talib & Hasan, 2010) and thus are capable of processing problems involving very non-linear and complex data even if the data are imprecise and noisy (Lek & Guégan, 1999). ANNs also was shown to predict the succession, timing and magnitude of algal blooms with reasonable accuracy (Recknagel et al., 1997; Recknagel et al., 2002; Kim et al., 2007a). Therefore, ANNs are ideally suited for modelling ecological data which are known to be very complex and often non-linear.

ANNs can be classified by their method of learning or training. In supervised learning, the network is taught to recognise output or target data while in unsupervised learning performs clustering of data into similar groups based on the attributes or feature of the inputs (Anita Talib, 2006).

ANNs have a wide range of applicability and various types of neural networks such as auto-associative memory, optimization, data reduction and generalization have been developed in order to cater for different kinds of problems. ANNs have been applied to problems ranging from predicting cation exchange capacity in soil properties (Tang et al., 2009), identifying morphometric features in geological information (Ehsani et al., 2010), determination of effectiveness in neuroscience treatment (Leighty et al., 2008) as well as in the management of water and waste water treatment process (Khataee & Kasiri, 2010). The type of network preferred by any researchers depends on the nature of the problem to be solved. Among the most popular ANNs are (i) multi-layer feed-forward neural networks trained by back propagation algorithm, i.e. back propagation network (BPN), and (ii) Kohonen self-organizing mapping, i.e. Kohonen network (SOM).

2.2.2 Evolutionary Algorithms (EA)

Evolutionary algorithms (EA) are bio-inspired adaptive methods which mimic processes of biological evolution, natural selection and genetic variation such as cross-over and mutation to develop solutions to complex computational problems (Recknagel et al., 2006). There are several types of evolutionary algorithms used to evolve solutions to the target problem; the three most common according to Hoffmann (2006) are genetic algorithm (GA), evolutionary strategies (ES) and genetic programming (GP). These algorithms are a class of non-deterministic (derivative-free), stochastic, iterative search techniques that emulate some of the principles of Darwin evolution (selection and reproduction of the fittest individuals, with some introduced variation during the reproduction step) (Hoffmann, 2006).
The essential principles and component required for a simple GA program and could also apply for the implementations of all EA (with some modification) are as follows;

i. population of individual solutions
ii. fitness function to evaluate the quality of each solution
iii. selection mechanism for choosing some individuals to reproduce
iv. operator for rearranging and changing the information content of those individuals chosen to reproduce
v. termination criterion

The use of EA was shown to be capable of finding good solutions to novel problems in complex environment – requirements that suit the concepts of ecology and ecological modelling (Whigham & Fogel, 2003). In EA, a model can also be expressed in the form of rules.

2.2.3 HEA: Hybrid Evolutionary Algorithms

This study uses a hybrid EA (HEA) from Cao et al (2006a) that differs from and improves on the EA approaches by using genetic programming to generate and optimise the structure of the rule set and later using a general genetic algorithm to optimise the random parameters in the rule set. HEA is designed for rule discovery in water quality time-series (Cao et al., 2006b) and is capable of forecasting potential algal population dynamics and outbreaks in water bodies. In the HEA, a model is expressed in the form of IF-THEN-ELSE.

HEA works by simulating a number of simplified processes occurring in life and implemented in artificial media such as a supercomputer. It uses genetic programming to generate and optimize structure of rule sets and a genetic algorithm (GA) to optimize the parameters of a rule set (Cao et al., 2006b). HEA is capable of forecasting potential algal population dynamics and outbreaks in water bodies by providing alternative approaches to problem solving - where solutions of the problem are evolved rather than the problems being solved directly.

In this study, HEA was used to search for suitable representation of a problem solution by means of genetic operators and the principle of ‘survival of the fittest’. During each run, the program will assess each output by means of ‘fitness cases’. Only the fitter results are selected for recombination by using the genetic operator such as crossover and mutation to create the next generation. These will be consequently repeated for generations until the criteria for termination of the program has been met. This is followed with parameter optimization of the rule set by means of genetic algorithm. The optimisation of constants within process models of chlorophyll-a concentration for freshwater ecosystems, particularly in lake environments has significantly improves the quality of the results on the unseen data (Whigham et al., 2006).
The complexity of rule sets can be controlled by the maximum tree depth and the rule set size. Rules discovered by HEA have the IF-THEN-ELSE structure and allow imbedding complex function synthesises from various predefined arithmetic operator (Cao et al., 2006b). Fitness of the solution is assessed using the Root Mean Square Error (RMSE). Sensitivity analyses are then applied for each rule to assess the influence of changes in input on the output variable. Examination on both THEN and ELSE separately gives specific information about algal growth on respective branch. The model output includes a list of the variables used within each run, the frequency of each input parameter being used and the RMSE as well as r-square values.

In the framework of the present study, HEA was applied to forecast algal population dynamic in the study sites. HEA has successfully been applied to shallow, eutrophic Lake Kasumigaura (Japan) and deep, mesotrophic Lake Soyang (Korea) where rule sets that can forecast for 7-days ahead seasonal abundance of blue-green algae and diatom populations with relatively high accuracy in these lakes were also been discovered (Cao et al., 2006b). It has also been successfully applied in the hypertrophic Nakdong River (Cao et al., 2006a), temperate, warm-monomictic and hypertrophic lakes in South Africa (Recknagel et al., 2008b), two eutrophic Dutch lakes Veluwemeer and Wolderwijd (Talib, 2008) and warm monomictic and mesotrophic lakes in Australia (Welk et al., 2008). In addition to the forecasting capabilities, these rule set models are also explanatory for the relationships between physical, chemical variables and the abundances of algal populations.

2.3. Process-based Lake Models

Process-based lake model ranges from simple to complex and detailed mathematical models based on representation of basic physiological principals of phytoplankton growth and other interactions. Processes such as photosynthesis, respiration and mortality form basic of such models and include usage of differential equations. In process-based models, the mathematical representations separate out all the processes and use hydrological inputs to results in a time series of outputs.

2.3.1 The Differential Equation Approach

Differential equation (DE) models focus on the representation of ecological processes by mathematical equations. These differential equations are generally known as rates in ecological systems. The DE is used to describe the relationship between pools and fluxes of variables that change over time and space.

An ordinary differential equation (ODE) is a differential equation in which the dependent variable is a function of a single independent variable. A partial differential equation (PDE) is a differential equation in which the unknown function is a function of multiple independent variables and the equation involves its partial derivatives. An example of a phytoplankton biomass differential equation to simulate growth and grazing processes is given as follows:
\[
dPHT/dt = \text{phytoplankton photosynthesis} - \text{phytoplankton respiration} - \text{phytoplankton grazing}
\]

This equation relates that the change in phytoplankton biomass (PHT) over time (t) depends on the phytoplankton photosynthesis, phytoplankton respiration and grazing by zooplankton respectively, where:

- Phytoplankton photosynthesis depends on the growth rate (termed as \( U_{\text{max}} \)) and is limited by nutrient (normally nitrate and phosphate) availability, light and temperature.
- Grazing by zooplankton is based on respiration rates, grazing rates and temperature limitations.

Differential equations is used whenever a deterministic relationship involving some continuously varying quantities (modelled by functions) and their rates of change in space and/or time (expressed as derivatives) is known or postulated. This is normally gained from laboratory or site experiment as well as from literature. In order for the biomass of phytoplankton to increase over time, phytoplankton photosynthesis rate must be higher than the respiration and grazing processes.

An increasing number of state variables couple with more non-linear interactions between nutrients and biota has increased the complexity of these models. These kinds of models are now a popular choice for decision support tools especially for lake management.

2.3.2 The Model SALMO-OO

SALMO (Recknagel & Benndorf, 1982) was designed using the FORTRAN IV programming language and validated as being generic for a range of lake properties. Current version of SALMO-OO is generic for non-shallow lakes with maximum depth more than five meters. The advance of computational technology has augment to the current state of the art knowledge in ecological modelling including SALMO model. However, as model becomes bigger, they tend to be more complex and rigid in comparison with the flexible and complex characteristic of ecosystems. Cetin et al. (2005) has strengthen the SALMO by implementing a more generic simulation library for lakes by taking advantage of object-oriented design and Java programming in producing SALMO-OO model with SALMO as the core. This has increased the generality, realism and flexibility of SALMO-OO for simulations lakes of different trophic states. The design of object-oriented programming languages such as JAVA has largely inspired by organising principles in ecology, such as hierarchy and taxonomy (Recknagel & Cao, 2009). SALMO-OO model is now a comprehensive decision support tool for lake and reservoir management which can be used to support the ranking of alternative management scenarios and to base decisions on understanding and expert knowledge in limnological area. The SALMO-OO model in its original form simulated a variety of lake conditions very well, whilst still maintaining an acceptable level of accuracy (Cetin, 2007).
The SALMO-OO has four process models in the library adopted from Park et al. (1974), Hongping and Jianyi (2002) and Arhonditsis and Brett (2005) respectively for phytoplankton growth and grazing as well as zooplankton growth and mortality in addition to the original model. More than twenty set of data for freshwater lakes in Australia and overseas are available in the database. Summary of the state variables and process equation provided by SALMO-OO to simulate pelagic food web dynamics and nutrient cycles is given in Cetin (2007). Each of the state variables is described by an ODE which is solved by the fourth order Runge-Kutta method. SALMO-OO is used for long term scenario analysis where general overview of the ecological processes such as long-term eutrophication trend and built-up can be observed and detected. It has been proven to be useful tool for integrated decision making in eutrophication control of lakes and reservoirs (Recknagel et al., 1985).

The growth of phytoplankton is resource-limited and therefore, it is possible to developed models resembling the interactions of phytoplankton-resource. One simple model describing the relationship between the growth rate of an algal species and the availability of a limiting resource is the Monod model (Tilman et al., 1982). Monod (1949) described the growth of bacteria in steady-state chemostat culture as a rectangular-hyperbolic (Michaelis–Menten type) function of the concentration of the limiting external nutrient, thus providing a simple description of resource-limited growth. A basic assumption governing the use of this model is that the growth rate of an alga is dependent solely on the concentration of a particular limiting nutrient (Goldman & Carpenter, 1974). Individual species of phytoplankton have known a maximum performance that is subject to growth opportunities, water temperature and an adequate supply of light and nutrients. Thus, the core ODE in SALMO-OO addresses the sensitivity of growth performances of named phytoplankton species to water temperature, saturating light and the spectrum of nutritional requirements supply.

Phytoplankton growth rates in SALMO-OO are calculated with a temperature-dependent maximum rate times a reduction factor for nutrient and light limitation. The basic logistic equations consider the following:

**Temperature**

Temperature has long been considered a primary factor in determining phytoplankton succession because of its effects on the maximum growth rates and physiology of resource utilization (Tilman et al., 1982). Many factors can be related to the occurrence and dominance of cyanobacteria, but there is a general consensus that temperature is crucial for the growth and abundance of these organisms (Paerl & Huisman, 2009).

**Nutrient**

A balanced quantity of nutrients that is required for phytoplankton growth can be explained in two ways: In Liebig’s Law of the Minimum, the yield of a plant population is limited by the nutrient that is least available relative to the plant’s requirement for that nutrient (Buitenhuis & Geider, 2010). In Blackman’s Law, when a biological process is constrained by a number of separate factors, the rate of the process is limited by the pace of the ‘slowest factor’. This law is implicit in some models of phytoplankton growth where the growth rate is treated as the minimum of
the dependence of growth on nutrient concentrations and light intensity, modulated by
the temperature dependence of the light and nutrient saturated growth rate (Buitenhuis & Geider, 2010). The relationship between phosphorus concentration and chlorophyll
suggests, that phosphorus and at times nitrogen and silicon are limiting nutrient
(Tilman et al., 1982).

Nutrient limitation for phytoplankton growth can be modelled using either the
Droop model or the Monod Model. Both models have their strength and weaknesses
and the major difference between the two is that Droop model of phytoplankton
growth relates growth rates to internal nutrient concentrations (intracellular cell quota)
while the Monod model relates phytoplankton growth to external concentrations of
dissolved nutrients (Hecky & Kilham, 1988). The use of Monod model is preferred
because it directly relates a phytoplankton population level response to measurable
environmental parameters such as limiting nutrient concentrations (Sommer, 1991).
Monod model differ, but is still based upon Michaelis-Menten limiting steps. Thus,
the equations for both models function are identical.

When a single nutrient is limiting, the Monod-Michaelis-Menten formulation
which is supported by empirical evidences and has a theoretical basis is often used
(Scavia, 1980; Poggiale et al., 2010). This Monod-Michaelis-Menten formulation is a
function of the external nutrient concentrations and is constructed from two
parameters: a nutrient-saturated growth rate and a half-saturation constant. This
formulation does not consider internal nutrient quota (as in Droop model). However,
in situations where more than one resource can be limiting, or when species
abundances shift, or when the physical–chemical environment is dynamic, a single
Monod model will likely fail to describe algal growth kinetics over the whole possible
parameter space (Rhee & Gotham, 1981). Therefore, most ecosystem models employ
either multiplicative or threshold formulations, which are both empirical functions for
multi limiting nutrients (Poggiale et al., 2010) and more sophisticated models
consider both the external and intracellular nutrient concentrations.

Light
During the daylight period, it is not certain that there will always be sufficient
light to saturate growth rate. Below the water surface, the attenuation and scatter of
down-welling light have the effect of a sharp vertical fall in light intensity and except
in very shallow lakes, a corresponding decline with increasing depth is the likelihood
of light levels being able to saturate growth rates.

Given that phytoplankton may be entrained simultaneously in turbulent
circulations and that they are carried through this light gradient, it is natural that they
experience fluctuating light levels. Even in quite shallow lakes, it is inevitable that
vertical mixing takes them to depths where the light does not saturate photosynthesis.

Growth rate does not depend on instantaneous saturation of photosynthesis but
is a function of the light-independent growth process being supplied with sufficient
photosynthetic materials to sustain growth over the division cycle. If vertical
entrainment between higher and lower levels of incident light results in this
requirement being no longer satisfied, then growth rate is proportionately reduced.
**Mortality**
Phytoplankton population dynamics are also subject to ongoing mortalities, especially as a consequence of sinking out to deeper water and/or of consumption by pelagic grazers. Loss rates of phytoplankton may become so high that they can greatly exceed the rates of recruitment to growing populations. In this case, the whole population simultaneously decline.

In his review Jørgensen (2010) confirms the general trend in modelling where wider experiences in incorporating additional different details that are often useful in specific case is continuing. This comes as a compliment to the effort of overcoming the limitation of previous modelling and further improvements in models such as: parameter optimisation (Cao *et al*., 2008), hybridised models (Whigham & Recknagel, 2001), object oriented models (Cetin, 2007) and individual-based models application. All these models have the aims to create more flexible, realistic models with a greater explanatory and predictive power. There has also been an increasing interest from modelling the impact of climate changes on lakes (Jørgensen, 2010).

One of the recommendations on SALMO-OO made by Zhang (2006) was to extend the model library. Consequently, this study proposed to include another process model for algal growth, algal grazing, zooplankton growth and zooplankton mortality into the SALMO-OO library as part of the enhancement to the model.

During models’ evaluation, their parameters are set at optimal values in order to generate a good model prediction. Various different approaches have been used to overcome the parameter optimisation problem in ecological models. Common methods applied to obtain an optimum model parameter set is by manually changing one or a few parameters stepwise until the model's prediction is reasonably close to the measured data. The most basic approaches have been the iterative calibration of the most important parameters by trial and error coupled with best-judgement tuning (Scavia, 1980) and individually adjusting parameters to the maximum or minimum of their assigned ranges whilst keeping all other parameters at their assigned means (Schladow & Hamilton, 1997). Global optimisation technique by Swartz and Bremermann (1975) is among the earliest work on parameter optimisation for mathematical models of differential equations in biology. Apart from being highly subjective, time consuming and resource intensive, the trial-and-error procedure relies on the modeller’s experience and prior knowledge of the lake under study. Manual optimisation for sensitive parameters such as phytoplankton respiration rate and maximum zooplankton grazing rate by Cetin (2007) does not require the design and programming of an optimization algorithm or goodness function and is not computationally intensive. However, this is highly subjective, time-consuming and practically impossible to achieve for some large and complex models. Law *et al*. (2009) has demonstrated the capability of Bayesian model calibration in efficiently estimating the parameter values of aquatic biogeochemical models. Nonetheless, this arguably artificial intelligence technique (Chen *et al*., 2008) is computationally expensive. With an increase in model complexity, alternative methods are required in order to solve optimisation problem and determine optimal parameter values.

Not all optimisation method can improve models’ performance. Radtke and Straskraba (1980) found that the use of dynamic optimisation procedures for an algal
population model has given rise to several problems. This has led researchers to develop bio-inspired computational techniques for solving complex problems. Knowledge-based or artificial intelligence techniques such as evolutionary and genetic algorithms as well as swarm intelligence are used increasingly as alternatives to more classical techniques to model environmental systems (Chen et al., 2008).

2.3.3 Structure and Parameter Optimisation of SALMO-PLUS

The SALMO-PLUS stand-alone as applied in this research integrates SALMO-OO as the simulation module with a particle swarm optimisation (PSO) algorithm as an optimisation tool for optimum simulation of nutrient state variables and phytoplankton dynamic in lake ecosystem. This is done by optimising targeted parameter in search of their respective optimal values. Simulation-optimisation methods linking a detailed simulation model with a heuristic or population-based evolutionary algorithm are becoming increasingly attractive for solving optimisation models (Shourian et al., 2008). PSO is a population-based swarm intelligence algorithm driven by the simulation of a social psychological metaphor instead of the survival of the fittest individual (Coelho & Mariani, 2009). PSO searches the problem space in parallel with a population of candidate solutions, similar to the approach adopted by evolutionary algorithm. However, whereas the search in evolutionary algorithm is driven by competition among candidate solutions, the search in PSO is driven by cooperation (Floreano & Mattius, 2008).

Changes of the parameters from their default values are based on optimisation of a goal function that can describe the fitness under changing lake environmental conditions. The goal function of optimisation in this research is to minimise the divergence of the calculated state variables from the measured data. The problem of parameter estimation has been solved in this research with approaches ranging from iterative trial and error estimations to the use of particle swarm optimisation. Indeed, parameters in models are constantly varied to account for adaptations and shifts in the species composition for specific lakes.

Optimisation aims at finding the best (i.e., optimal) solution from some sets of available alternatives to solve a problem. This research has determined optimal values for eleven biochemical parameters, which have been identified as most influential through a global sensitivity analysis. Only parameters affecting key processes ratio of phytoplankton such as nutrient uptake (nitrate and phosphate), growth (temperature, light) and zooplankton grazing and growth (temperature, mortality) were considered for optimisation. Eleven parameters specific and related to algal growth targeted for optimisation in this research are listed in Table 14.

The particle swarm optimisation is proposed as an alternative to genetic algorithm (GA) due to the following comparison (Liu et al., 2005):

1) Particle swarm optimisation has memory where knowledge of good solutions is retained by all particles whereas in GA, previous knowledge of the problem is destroyed once the population changes.
2) Particle swarm optimisation has constructive cooperation between particles and particles in the swarm share information between them.

3) Particle swarm optimisation is a simple concept, can easily be programmed and neither encoding nor decoding processes are needed as they are in GA.

In single objective optimisation, the optimal solution is clearly defined. Application of PSO to single objective optimisation tasks has been found to be fast and reliable, often converging to global optimal solutions within a few steps (Kennedy & Eberhart, 2001). By contrast, optimal solution in multi objective optimisation needs to consider all objectives concurrently. Application of multi objective optimisation is common in cases where two or more sometimes competing and/or incommensurable objective functions have to be minimised simultaneously (Parsopoulos & Vrahatis, 2002). A practical solution to a multi-objective parameter optimisation is to investigate a set of solutions, each of which satisfies the objectives at an acceptable level without being dominated by any other solution. The trade-off achieving optimisation give rise to numerous optimal solutions alternatives (Parsopoulos & Vrahatis, 2002).

The two general approaches to multiple-objective optimisation as discussed by Konak et al. (2006) are 1) using a combination of the individual objective functions into a single composite function or move all but one objective to the constraint set and 2) to determine an entire Pareto optimal solution set or a representative subset. This research uses the later approach. A Pareto optimal set is a set of solutions that are non-dominated with respect to each other in the solution space. Pareto optimal solution sets are often preferred to single solutions because they can be practical when considering real-life problems since the final solution of the decision-maker is always a trade-off (Konak et al., 2006). However, identifying the entire Pareto optimal set and proof of the solution optimality are both computationally expensive. Therefore, the practical solution by looking for a set of solution that represents the closest to Pareto optimal was adopted here.

2.3.4 Particle Swarm Optimisation (PSO)

Modern heuristic optimisation techniques such as hybrid evolutionary algorithms and artificial neural network have been given much attention by many researchers due to their ability to find an almost global optimal solution, as an alternative to the conventional mathematical approaches (Coelho & Mariani, 2009). One of these modern heuristic optimisation paradigms is the Particle Swarm Optimization. The PSO is a simulation of simplified social ecosystem and capable to be used as an optimizer (Kennedy & Eberhart, 1995). The original intent of PSO was to mimic biological techniques and graphically simulate the choreography of a bird’s flock or fish school to help solving computational problems. PSO is a very simple concept and paradigm of a few lines of computer code and requires only primitive mathematical operators and is computationally inexpensive in terms both memory requirements and speed (Kennedy & Eberhart, 1995). This technique is inspired by the collective social behaviours of simple individuals of biological systems interacting dynamically within their environment and each other. PSO is one of the two algorithms under Swarm Intelligence (SI). The other algorithm is known as ant
colony optimisation (ACO). SI is a form of agent-based modelling inspired by colonies of social animals such as ants and bees (Denby & Le Hégarat-Mascle, 2003) or school of fish. As a population based stochastic optimization computational technique, PSO has a number of similarities with other evolutionary computation techniques such as GA. However, PSO is a population-based swarm intelligence evolutionary algorithm driven by the simulation of a social psychological behaviour instead of the survival of the fittest individual (Coelho & Mariani, 2009). The optimisation objective function or goal function is to minimize the sum of the squares of the weighted difference between actual measurements and simulated results within the constraints of parameter ranges (Swayne et al., 2010).

As a population-based evolutionary algorithm, PSO is initialized and operates with a population of random solutions and searches for optimum by updating generations. However, PSO is different from GA in the sense that the former has no evolution operators such as crossover and mutation compared to the latter. In PSO, the population dynamics simulate the behaviour of a “birds’ flock”, where social sharing of information takes place and individuals profit from the discoveries and previous experience of all other companions during the search for food (Parsopoulos & Vrahatis, 2002). The potential solutions in PSO are called particles where it searches the problem space by following the current optimum particles and solutions that can be represented as a point in an n-dimensional solution space. Unlike the most of the evolutionary algorithms, each potential solution (individual) in PSO is also associated with a randomized velocity, and the particles are then “flown” through the problem space (Coelho & Mariani, 2009) where a number of particles are randomly set into motion through this space. Therefore, a wider area of problem space is covered by particles where different searching paths of different particles increase the chance of PSO finding the global optimum solution.

According to Coelho and Mariani (2009), each particle in PSO keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) it has achieved so far and this value is called pbest while another “best” value that is tracked by the global version of the particle swarm optimizer is the overall best value and its location obtained so far by any particle in the population is known as gbest. Particles observe the ‘fitness’ of themselves and their neighbours at each iteration and ‘emulate’ their successful neighbours (defined as those whose current position represents a better solution to the problem than theirs) by moving towards them. In doing so, an approach for grouping particles into competing, semi-autonomous flocks can be applied or all the particles can work together into a single global flock instead. The performance of each particle is measured according to a fitness function. In optimisation problems, the fitness function is usually the goal function under consideration (Parsopoulos & Vrahatis, 2002).

PSO is an extremely simple algorithm that seems to be effective for optimizing a wide range of functions (Kennedy & Eberhart, 1995). It has the advantages of easy implementation and few parameters to adjust compared to GA. Details of PSO is given in Shourian et al (2008). This study utilised PSO in the SALMO-PLUS (Martin & Recknagel, 2010) to offer solutions to the parameter optimisation by minimising the difference between the models measured and predicted data and maximizing the r-squared values.
The performance of PSO greatly depends on its parameters and proper control of global exploration and local exploitation that assist in finding the optimum solution. This is another aspect to be considered for future research in SALMO-PLUS. Liu et al. (2005) proposed to use the inertia weight of PSO that controls the impact of previous velocity on the current one in order to balance between exploration and exploitation and achieve trade-off between the two. The proper control of the inertia weight is very important to find the optimum solution accurately and efficiently.

Regardless of the optimization method used, model parameter optimisation requires a measure for the "goodness" of model prediction, i.e. for how well the model produces the desired prediction. In this case, r-squared and RMSE (root mean squared error) values are used. R-squared ranges from zero to one, with zero indicating that calculated outputs of the proposed model do not match measured outputs and one indicating perfect prediction. On the contrary, lower values of RMSE indicate better fit. RMSE is a good measure of how accurately the model predicts the response and is the most important validation criterion especially when the main purpose of the model is prediction.

2.4 DYRESM Model for Lake Putrajaya

The hydrodynamic numerical model DYRESM was developed by Imberger et al. (1978) and is based on fundamental energy, mass and momentum equations describing the one-dimensional structure detected in small to medium lakes and reservoirs. DYRESM is freely available via webpage http://www.cwr.uwa.edu.au/

DYRESM simulate water bodies by dividing into horizontal Lagrangian layers which are expandable and contractible to accommodate change in volume due to inflow, outflow and evaporation. The Lagrangian formulation avoids the need to calculate vertical velocities, greatly decreasing computational time and minimising numerical diffusion (Hamilton & Schladow, 1997). The thickness of the layers also changes when vertical movement accommodates the bathymetry of the water body (Lewis, 2004). These layer structures can change rapidly between isothermal and stratified conditions to reflect seasonal behaviour in lakes.

DYRESM predictions are driven by volume changes produced by inflow and mixing and are dependent on the thickness of the horizontal layers to detect changes in vertical density stratification (Imberger et al., 1978). The thermal profile of Lake Putrajaya was simulated by DYRESM by utilising vertical water temperature, salinity and density with horizontal Lagrangian layers that vary in thickness and number. DYRESM assumes that the density stratification occurring in lakes inhibits vertical motion. In contrast, horizontal density variations are defused by lateral and longitudinal convection at a faster rate than vertical advection. Therefore, 1-dimensionality is met. Small to medium sized water bodies often exhibit a one-dimensional structure that can be adequately modelled with first order balances of energy, momentum and mass which are determined by vertical property variation (Lewis, 2004).
2.5 Validation Methods

The next step in a forecasting study is the testing of the accuracy of the predictions. Simulated results must be treated with caution and it must be checked whether the predicted values are sufficiently accurate for the prediction of the observed values. A general class of methods known as resampling procedures is available for this purpose. These include the bootstrap procedure; leave one out cross validation and the jack-knife procedure. Only the two most common methods of accuracy estimation (cross-validation and bootstrap) are discussed in the following.

2.5.1 Leave-one-out Cross Validation

In the field of bioinspired-computation, the accuracy of predictions from data that was not used for training can be obtained by means of cross-validation method. The idea is to randomly split a sample into two subsamples and to use one subsample as the training sample (for model building) and the other one as the test sample (for model validation). This process is repeated for a number of times where partitioning of the data set is the key element of cross-validation. The three most common partitioning in cross-validation are:

1) leave-one-out cross-validation
2) leave-many-out cross-validation and
3) K-fold cross-validation

Details of these three methods for cross-validation are given by Shtatland et al. (2004). A requirement of cross-validation is that the data set should be of a sufficient size for being split into two parts each large enough to apply a cross-validation procedure (Adèr et al., 2008). In leave-one-out cross-validation, a single observation from the original sample is left out at a time for validation while the remaining observations are used as the training data. This process is repeated until each observation in the sample is used once as the validation data. Thus, in leave-one-out cross-validation method, the learning algorithm is trained multiple times, using all but one of the training set data points. Cross-validation uses the leave-one-out training samples to ensure that the training samples do not contain the test point (Efron & Tibshirani, 1997). For K-fold cross-validation, data are split into K subsets with each is held out in turn as the validation set. Therefore, leave-one-out could be said as the degenerate case of K-fold cross-validation, where K is chosen as the total number of examples (a common choice for K-fold cross validation is K=10).

The advantage of leave-one-out cross validation is that it does not waste data where in training, all but one of the points are used. Therefore, the resulting regression or model is essentially the same as if they had been trained on all the data points. However, this could also be a disadvantage for this method where the large number of repetitive training process is computationally very expensive. A fundamental difference between cross-validation and bootstrapping is that the bootstrap re-samples the available data at random with replacement, whereas cross-validation does it without replacement (Shtatland et al., 2004).
2.5.2 Bootstrap Procedure

Bootstrap is a powerful technique that permits the variability in a random quantity to be assessed using just the data at hand (McLachlan, 1992). The procedure of bootstrap requires sampling within a sample approach where sampling depends on the original sample. In this research, the algorithm of HEA randomly picks a certain given percentage of data for training and testing respectively. The bootstrap method assumes that the sample is a valid representative of the population and the sub samples come from the same distribution of the population. This means that data collected for a single experiment are utilised to simulate what the results might be if the experiment was repeated numerously with a new sample of data. Each sample is assumed to be drawn independently from the other samples.

There are several cases for which a bootstrap procedure is recommended (Adèr et al., 2008):

- **When the theoretical distribution of a statistic of interest is complicated or unknown.** Since the bootstrapping procedure is distribution-independent it provides an indirect method to assess the properties of the distribution underlying the sample and the parameters of interest that are derived from this distribution.

- **When the sample size is insufficient for straightforward statistical inference.** If the underlying distribution is well-known, bootstrapping provides a way to account for the distortions caused by the specific sample that may not be fully representative of the population.

- **When power calculations have to be performed and a small pilot sample is available.** Most power and sample size calculations are heavily dependent on the standard deviation of the statistic of interest. If the estimate used is incorrect, the required sample size will also be wrong. One method to get an impression of the variation of the statistic is to use a small pilot sample and perform bootstrapping on it to get impression of the variance.

Besides the advantage of providing more accurate estimates for prediction error, the bootstrap replication also provide a direct assessment of variability for estimated parameters in the prediction rule (Efron & Tibshirani, 1997). Increasing computational power available has increase the number of bootstrap samples recommended in literature as well. More samples could be submitted given more powerful computers are available. However, increasing the number of samples can only reduce the effects of random sampling errors arising from a bootstrap procedure itself and may not increase the amount of information in the original data.

Bootstrap procedures can also be used when the information about the distribution of the statistic of interest is known beforehand. If this is the case, the
sampling procedure may be adapted accordingly to suit the distribution assumptions. Bootstrap method was used over leave one out for HEA experiment in this study.

2.6 Models Generic for Lake Ecosystem Categories

The rule-based models for algal population dynamics should ideally be applicable to more than one lake and be generic to a certain extent. However, lake ecosystem dynamics posed diverse structures and functions which are induced by climate, morphometry and trophic state. Thus, it is nearly impossible to create a single model that is valid for any lake in the world. The successful expansion of SALMO-OO simulation scope to include prediction on tropical lake requires an accurate simulation of key state variables such as phosphate concentration, phytoplankton biomass and algal functional groups as well as biomass of herbivorous zooplankton. This requires optimisation of parameters in SALMO-OO model to better suit respective local conditions of lakes under study.

Recknagel et al. (2008c) defined lake categories by trophic state and circulation type, assuming that circulation type reflects climate conditions and morphometry to some extent while the trophic state indicates habitat properties and community structures. Due to the similarities in key lake characteristics, it can be assumed that water bodies within a similar lake ecosystem category encounter similar ecological behaviours, health issues and management options (Welk et al., 2008).

2.7 Tropical Lake Ecosystem

A comprehensive understanding of the metabolic processes supporting phytoplankton growth has been well developed. These include the relationship between: photosynthesis and underwater light penetration (Ganf, 1974a; Lewis, 1974; Longhi & Beisner, 2009; Elliott et al., 2010), the requirement and rates of uptake of essential nutrients (Lewis, 1974; Longhi & Beisner, 2009; Elliott et al., 2010) and the effect of temperature gradients on the distribution of phytoplankton into more defined layers (Robarts & Zohary, 1987; Longhi & Beisner, 2009; Elliott et al., 2010). The temporal cyclic patterns of phytoplankton composition are well-known in temperate regions, where temperature, stratification, light, and nutrients availability exhibit recurrent seasonal dynamics (George et al., 2010). In contrast, tropical lakes showed minimal seasonal variation in solar radiation and water temperature, but seasonal patterns of phytoplankton communities are not negligible (Talling, 1966, 1987). According to Tundisi (1990), most reservoirs in tropical latitudes follow a significantly different path from the reservoirs in temperate zone, with regards to effects of forcing functions, decomposition rates and seasonal circulation periods. The lake variety, especially high temperatures, intense illumination and prolonged growing season lend special interest to the tropical water bodies (Talling, 1965).

Our present understanding of lake ecosystem structure and function derives primarily from research conducted in northern temperate regions (Ewel & Fontaine, 1983). Tropical reservoirs are poorly studied, but their numbers are steadily growing.
(Talling & Lemoalle, 1998). Indeed, limnology has been predominantly a discipline of high latitude with the earliest limnological studies (e.g. Vollenweider, 1975) and most of the world’s extensively studied lakes and reservoirs being concentrated in the temperate zones (Beadle, 1981). This is despite the fact that lakes in tropical regions differ markedly from one another in size, basin shape, local climate and biota. Tropical climate involves a relatively high temperature, high rainfall and humidity with small seasonal fluctuations. Also in the tropic, seasonal changes are apt to be less well marked (Beadle, 1981).

Among the lakes located on equatorial, large lakes in Africa and South America received much more attention compared to the same water bodies in the South East Asia region (see Beadle, 1981; Vincent et al., 1986; Talling, 1987; Moreau et al., 1993; Gunkel, 2000; Msomphora, 2005; Figueredo & Giani, 2009) where Malaysia is located. This is largely due to the fact that lakes in that area are much bigger than in any other tropical region. However, limnological research on tropical lakes of South East Asia region is nothing new. The earliest scientific record on tropical waters of South East Asia region was the German Sunda expedition in 1928 which examined a series of lakes in Java, Sumatra and Bali (Thienemann, 1954). Thienemann (1954) observed that a different set of organic life from temperate region was found around the land of tropical area, while the tropical freshwater fauna greatly resembles temperate areas, apart from fish, higher crayfish and some other categories of animals.

Due to the aforementioned facts, certain definitions of temperate climate might be difficult to apply to tropical lakes for the reason that some lakes do not fit easily into any classification system that exists (Horne & Goldman, 1994). Because of the fundamental differences between temperate and warm-water tropical, the temperate lake models are not applicable without modifications to the majority of water bodies of the region (Salas & Martino, 1991). Local scientist in tropical regions have developed a considerable understanding of the ecosystems of tropical lakes, however, their information is not widely available in international publications generally due to logistical issues in conducting research and capability concerns.

Environmental conditions influenced by tropical latitude can be divided into three types: solar radiation input, geostrophic influence and air-mass circulation (Talling & Lemoalle, 1998). Different latitudes on earth receive different amounts of solar radiations. Energy-related variables such as air temperature, day length and visible incident radiation are highly correlated with latitude and altitude (Brylinsky & Mann, 1973). These energy-related variables are the key factors in determining a region’s climate where according to the Lambert-Beer’s law, the irradiance of a single wavelength will be attenuated exponentially in a homogenous medium in relation to the quantity and quality of the attenuating components within the path length (Loiselle et al., 2007).

Any sunlight that falls on water penetrates to varying depths depending on the intensity of radiation, the amount of surface reflection and the transparency of water at the time. Even in transparent water, only 36% of the sunlight which enters the water penetrates to about one meter and the light fades away quickly with increasing depth (Ewusie, 1980). The solar radiation sustains the ecology of a water body by heating the water body resulting in the differences in temperature of water mass,
mixing and chemical composition alterations. This will affect the suitability of habitats to the temperature tolerances of aquatic life. The solar radiation also provides light for photosynthesis by plants, either the microscopic algae or the macrophytes.

The combination of both intensive solar heating and high moisture produces a persistent zone of convection where generous cloud formation and frequent rainfall occur. The major characteristic of lakes in tropical region is the continually high temperature. Tropical waters are naturally warmer than temperate waters and they rarely freeze at any time of the year. One of the reasons is that higher latitudes receive less heat than lower latitude areas nearer the equator. This phenomenon is also due to the difference in hypolimnetic temperature between temperate and tropical lakes and the variable duration of stratification in tropical lakes (Talling & Lemoalle, 1998). Small temperature differences at high temperature can generate the same stability in a lake as much as greater temperature difference did at lower temperatures (Wetzel & Likens, 2000). Thus, the physical importance of temperature profile is not the temperature itself but the density difference produced by the thermal difference. However, water currents and prevailing winds cause an up-welling of sub-surface waters and help maintain the cooling effects in the water and break down the thermocline by bringing colder waters from lower depths to the surface (Talling, 2001).

Evaporation, precipitation and average temperatures characterise the sub-climate at the local level. The tropics share with other latitudinal belts the prevalence of three environmental cycles of fixed periods – the diel (day-night), lunar and annual (Talling & Lemoalle, 1998). The diel cycle of thermal stratification, is more marked in many tropical than in temperate lakes. Thermal stratification often interacts with biological activity by associating with density structure to condition and influence the depth-time distribution of induced chemical changes in the water (for example of dissolved oxygen, carbon dioxide and pH) and some planktonic populations (especially blue-green algae or cyanobacteria) (Talling, 2001). In terms of oxygen, tropical eutrophic lakes often have year-round depressed oxygen levels near the bottoms (Horne & Goldman, 1994).

In addition, the eutrophication process in the tropics and the temperate region is probably different since lakes in the tropics are limited in nitrogen rather than phosphorus (Nilssen, 1984). The differences in the structure of the trophic cascade have been hypothesized to cause rare occurrences of a clear-water phase in tropical lakes (Leeuwen et al., 2007). Many tropical fish species are omnivorous and predate on zooplankton as well. They constantly reproduce planktivorous individuals during the year round thus, reducing the likelihood of large zooplankton causing a clear water-water phase in the region.

The general impression from studies of tropical lakes is that large diurnal changes in irradiance, temperature, pH values and oxygen saturation are more influential on the lakes ecology than any seasonal changes that occur (Ganf, 1974b). The on-going monitoring at Lake Putrajaya indicates that the local climate shows little seasonal change and the population densities remain fairly constant from day to day. The seasonal constancy of the flora and fauna of tropical lakes is in contrast with the pronounced seasonal effects in temperate lakes. In addition, the constant water inflow
and its close proximity to the equator also prevent seasonality in Lake Putrajaya compared to temperate lakes.

Consequently, ecological indices developed for temperate lakes may not truly be applicable to tropical lakes such as in Malaysia. A simplified total phosphorus model was developed by Salas and Martino (1991) based on minimum data collection of 40 lakes and reservoirs in Latin America. Their studies were based on research by different investigators of diverse laboratories of the region and produced a model to predict the trophic response of lakes/reservoirs to changes in their total phosphorus concentrations based on the following relationships of mass balance equation:

\[ P_\lambda = \frac{L(P)}{Z(1/T_\omega + K_s)} \]  

Where:
- \( P_\lambda \) = total phosphorus (mg/l)
- \( L(P) \) = areal total phosphorus loading rate (gm\(^{-2}\)yr\(^{-1}\))
- \( Z \) = average depth of lake (m)
- \( T_\omega \) = detention time (yr)
- \( K_s \) = overall loss rate of total phosphorus (yr\(^{-1}\))

The difficulty of using the above equation is that the overall loss rate, \( K_s \), is neither readily known nor can it be measured in direct experimental way. Hence there is a need to calculate the \( K_s \) term by applying the other knowing parameters of the equation first and later estimate correlation with more readily obtainable parameters which upon substitution permits the use of the above equation as a predictive tool. As the estimation correlation requires non-available local data, it was decided not to use the Salas and Martino (1991) index for the purpose of current research. Furthermore, the index by Salas and Martino (1991) should be treated with care since most of the water bodies selected in their research is reservoirs with high sedimentation rate, had detention times of less than one year and showed markedly seasonality inherent in the location of the water bodies. As expected, studies on the trophic state of tropical lakes in Brazil by Petrucio et al. (2006) considering the Salas and Martino (1991) and Carlson (1977) indexes found that the former index was more appropriate to determine the trophic level of the region. However, Petrucio et al. (2006) found considerably varied trophic state for their study sites by using the Salas and Martino (1991) index compared to Carlson (1977) index. The two notable aspects in which tropical and large lakes differ from temperate large lakes are hydrology and lake circulation, which in turn affect nutrient cycles, algal production and fish production (GLRA, 2002).

As tropical areas are accustomed to high annual temperatures, one might assume tropical lakes are not vulnerable to the effects of global warming. However, studies on Lake Tanganyika in Central Africa found a rise in the local temperature and an alteration in the nutrient balance (Connor, 2003). Profound effects of global warming on a tropical lake were also demonstrated by Verburg et al. (2003). They demonstrated how increased water density gradients slowed down vertical mixing and reduced primary production in deeper parts of the lake. A growing concern in ecology over the last decade has been to understand how climate change might affect freshwater ecosystems and the possible increases in the abundance of Cyanobacteria species (see Paerl & Huisman, 2009). The impact of regional effects of global climate change on aquatic ecosystem functions and services can be larger than that of local
anthropogenic activity (O'Reilly et al., 2003). A rise in surface-water temperature will increase the stability of water columns in parallel with the regional warming pattern. Coupled with a regional decrease in wind velocity, these will contribute to less mixing thus, reducing deep-water nutrient up-welling and entrainment into surface waters. This is another interesting area to investigate since up-welling is a very important phenomenon in the cycling of nutrients in the waters of some tropical areas (Ewusie, 1980).

Although many studies on the effects of water quality changes on the biological, physico-chemical and hydrological properties of temperate lakes have been conducted in the last century, research on tropical lake ecosystem responses to anthropogenic pollution and local climate dynamics is still due (e.g. Vollenweider, 1975; Carlson, 1977; Wetzel, 1983; Marshall & Peters, 1989; Hejzlar et al., 2006). In the equatorial region, surface water is so abundant as to be a major natural resource and yet been only partially exploited. However, economic and social activities tend to aggravate difficulties arising from the natural fluctuations in the hydrological regime and the human impact on water resources increases with population growth and the spread of urbanization (Bonell et al., 1993).

Tropical inland waters are less frequently researched partly due to rare records of studies conducted and less published data in international journals. Among the tropics part, the South East Asian region attracts less interest compared to the South America area (Calijuri et al., 2002; Petrucio, et al., 2006) and Central Africa (Talling, 2001; O'Reilly et al., 2003; Verburg et al., 2003). Talling and Lemoalle (1998) produced a comprehensive review of tropical inland waters where according to Lewis (1987) information on tropical waters is so diffuse that it is difficult to use. However, the review by Talling and Lemoalle (1998) made little mention of limnological studies in Malaysia. In general, tropical examples especially from South East Asia region are scanty and to generalise is often insecure. Nevertheless, that should not prevent consideration being given to the potential long-term impacts of climate change on water resources (Rast & Thornton, 1996).

### 2.8 Limnological Research in Malaysia

Generally, research on limnological aspects of tropical lakes in Malaysia are sporadic even though the earliest limnological studies on a river in Malaysia was conducted in 1970s by Bishop(1973). Reports on eutrophication of freshwater ecosystems in Malaysia began a little bit later (see Ho, 1971; Arumugam & Furtado, 1980; Qududs, 1980; Arumugam & Furtado, 1981; Kah-Wai & Ali, 1998). Ho (1971) studied the primary productivity of algal community of Lake Ampang while Kumano (1978) studied freshwater dinoflagellates from natural lake known as Lake Bera. Preliminary studies on eutrophication of three lakes near Kuala Lumpur were done by Sulaiman et al. (1991).

Toward the 1990's, more emphasise was given on specific sub-discipline approaches such as studies on vertical and seasonal distribution of phytoplankton in Lake Kenyir (Yusoff et al., 1998). A study on long term phytoplankton changes in a small tropical reservoir was initiated by Anton (1994). Shuhaimi-Othman et al. (2007)
studied limnological aspects of Lake Chini which is another one of the two natural lakes in Peninsular Malaysia.

However, as far as this author is concerned, only a few scientific studies have been published internationally on man-made lakes ecosystem in Malaysia (see Arumugam & Furtado, 1980; Yusoff & Shark, 1987; Yusoff et al., 1997; Yusoff, et al., 1998) although some naturally occurring lakes have been investigated (see Kumano, 1978; Furtado & Mori, 1982; Shuhaimi-Othman et al., 2007; Shuhaimi-Othman et al., 2008). Indeed, the number of institutions primarily concerned with the rivers and lakes in the tropics is very small and major research teams are infrequent (Payne, 1986). However, cases of severe algal bloom in man-made tropical lakes are well known. Thornton (1982) reported a severe bloom in Lake Mcllwaine of Zimbabwe that had altered the fish community of the lake and caused the drinking water treatment to become expensive.

Efforts by relevant authorities to assess the status of eutrophication of water bodies in Malaysia are continuing despite the limited published work on these study areas and an increasing number of such water bodies. NAHRIM (2007) has categorised 90 lakes and reservoirs in Malaysia with 61% for water supply and irrigation, 39% for other purposes (hydropower, flood control, recreational and silt retention). However, only 30 out of the 90 lakes were extensively studied. This is surprising since man-made lakes dominate the Malaysian lentic environment. Ho (1996) reported there are only a few natural lakes in Malaysia and 54 man-made impoundments with their water surface areas range from 10 ha to 37,000 ha. A study by NAHRIM (2007) concluded that there are more than 90 major water bodies in Malaysia—many of which are for water supply, irrigation, hydropower generation, flood control and sediment traps. The current number of man-made water bodies is expected to rise over the next few years to meet the increasing demand for water and can easily reach thousands if small-sized water bodies i.e. lake and ponds are included.

2.9 Past Modelling Effort on Lakes in Malaysia

This section summarise the research efforts and limnological management at study sites in Malaysia. Efforts to predict potential eutrophication in water bodies by means of models are rare in Malaysia. Water quality monitoring in Malaysia involves rivers that are important water sources (Hight & Ferrier, 2006) and groundwater sources in selected area where conventional water supply is uneconomic. The situation in Malaysia is as what was reported by Jørgensen (2010) that limited amounts of data in developing countries often make it difficult to construct reliable lake models in spite of the urgent need for good lake models to aid in environmental management. One of the reasons could be that very few natural lakes exist in Malaysia and are located far from the populated area. There are only two natural lakes in Peninsular Malaysia (Lake Chini and Lake Bera). Generally, man-made lakes or reservoirs dominate the Malaysian lentic environment, with the majority being built either for irrigation or water supply (NAHRIM, 2007). These man-made water bodies are heavily managed and thus are less subject to natural ecosystem dynamics and anthropogenic influences.
However, the scarcity of modelling applications to predict algal blooms is in contrast with the numerous cases of algal blooms in Malaysia. Only one modelling approach utilizing Artificial Neural Network (ANN) to predict the water quality index in Langat River was established by Malaysian researcher (Hafizan et al., 2004). Kassim et al. (1998) applied the DYRESM model for studying profiles of temperature, dissolved oxygen, iron and manganese at Terip Reservoir for 304 days. Reviews of all modelling efforts indicate that models such as SALMO-OO and HEA have never been applied to lakes in Malaysia before.

2.9.1 Lake Putrajaya

There was only one modelling effort conducted by a consultant for the local authority managing Lake Putrajaya by means of the MIKE 11 model. This dynamic, one-dimensional model was utilised for a study on the design, management and operation of Lake Putrajaya. MIKE 11 contains modules for run-off simulations, hydrodynamics, flood forecasting, transport and dilution of dissolved substances, sediment transport and river morphology as well as various water quality processes. However, the generalised system of MIKE 11 is more suitable for one-dimensional flows like in rivers and estuaries (see Stronska & Borowicz, 1999). Jeong et al. (2003) found that a river mechanistic model has failed to describe uncertainty and complexity of reservoir-like ecological dynamics of Nakdong River in Korea.

Presently, there is no other ecological modelling research conducted at Lake Putrajaya apart from routine physical, chemical and biological monitoring by the local authority. Review on all previous modelling efforts indicates that this proposed research has never been conducted in Lake Putrajaya before.

Management of Lake Putrajaya

Lake Putrajaya is among the best managed water bodies in Malaysia. Situated in the city of Putrajaya, the lake management authority is responsible for management decisions and activities.

Potential threat in Lake Putrajaya

Even though Lake Putrajaya has so far been well managed, potential threat from an increasing human population and in-lake traffic are the most significant aspect to be taken care of. The increasing and on-going land development around the lake also increases the lake’s vulnerability to siltation and anthropogenic pollution and thus, potential eutrophication.

2.9.2 Lake Kenyir

All the models used in this study have never been applied to Lake Kenyir before. This study would be the first attempt to model the Lake Kenyir ecosystem dynamics.
Management of Lake Kenyir

The Terengganu Tengah Development Authority (KETENGAH) is a state statutory body tasked with the mission to increase tourist visit and develop Lake Kenyir as an eco-tourism area. In doing so, KETENGAH is embarking on plan to turn this South-East Asia’s largest man-made lake into a major tourist attraction in Malaysia. One of the plans is to turn Lake Kenyir into a duty-free zone. This has been helped with additional allocations from the federal government. However, the management in Lake Kenyir is quite complex - as with other water bodies in Malaysia-where visitor need permission from the state Wildlife Protection and National Parks Department (Perhilitan) prior to making the visit due to the fact that Lake Kenyir borders the National Park of Malaysia, whereas areas surrounding the hydropower power generation belongs to the national power company.

Potential threat in Lake Kenyir

Recreational fisheries is fast growing in Malaysia with about 15% increase per year (Yusoff et al., 2006) and plans to turn Lake Kenyir into fishing heaven will definitely increase the number of people visiting the lake. In order to rectify the disparity in the agro-food import sector, the Ministry of Agriculture has developed Aquaculture Industrial Zone (ZIA) plans for the East Coast Economic Region (ECER) to promote aquaculture in the area (MOA, 2007). This project will turn 2,800 ha of Lake Kenyir into the biggest aquaculture area in Malaysia and the government will provide 60,000 fish-cages at the cost of MYR50 million for local entrepreneurs involved in the project. An output of 2,000 tonnes a year from this zone will cater mainly for export market. However, a preliminary survey by the local university indicated that the E. coli levels in the water had have a negative effect on the fish population and that bacteria was traced from the sewage in the lake (Choong, 2010).

Presently, there are 30 to 40 boathouses on Lake Kenyir for the convenience of tourists and each can accommodate 10 to 15 passengers at a time. These boathouses had been allowed to operate there for the past 10 years without a proper sewerage system and nearly all of them discharge waste from the toilets directly into the lake. With no proper planning on sanitation and waste management, use of bigger house-boats to cater for increasing numbers of anglers in the future will increase the potential of water pollution and eutrophication in the lake.

Lake Kenyir faces more threats than just untreated sewage discharges. The boathouses have no proper sanitation where everything is flushed into the lake including food leftovers while water is drawn from the lake for cooking and washing. Over-fishing is very evident and there is inadequate enforcement by park rangers on the limit to the size or number or species of fish that are landed and for live fish being taken out.

2.9.3 Lake Penang

This study would be the first attempt to model the Lake Penang ecosystem dynamics. All the models used in this study have never been applied to Lake Penang ecosystem dynamics before.
Management of Lake Penang

The state of Penang in Malaysia is made up of the island of Penang (or Pulau Pinang) and a part of the mainland, Seberang Perai. The island part of this state which has sandy beaches and clear blue seas are famous as a major tourist destination where millions of tourists arrive each year. The island has traditionally been promoted as the ‘Pearl of the Orient’ in tourism industry and earned substantial tourist revenues in the past. This has led to hundreds of hotels being built over the years and caused a tremendous demand for water supply industry. These tourism industries as well as the industrial Free Trade Zone area depended on the Lake Penang for water supply.

Potential threat in Lake Penang

Despite the high rainfall in Malaysia, water shortage is already rampant in many states. Penang receives rainfall unevenly throughout the year causing droughts at times while at other times causing severe flooding (PBAPP, pers. comm. 2010). According to Chan (2006), Penang is still affected by unpredictable water supply quality although the state has privatized its water supply to a public listed company. Residents of Penang use an average of 285 litres of water daily which is the highest compared with people in other states in Malaysia. The high water usage by the public could be due to the lowest tariff compared to other states. This high usage has put additional constraints for the water supply in the state.

The state of Penang gets more than 80% of its water supply from River Muda, which flows from the state of Kedah where it has no jurisdiction (PBAPP, pers. comm. 2010). The water from River Muda flows into River Dua from which it is pumped for use in Penang. A proposal by the state of Kedah to allow logging operations in the water catchment of River Muda has also threatened the future of water supply for Penang (Hashim et al., 2010).

Apart from an eutrophication threat, the state also has experienced bitter incidences related to water resource such as the spillage of 2,700 litres of diesel into canal leading to River Dua Water Treatment Plant (Johari & Rohani, 2004) which together with Lake Penang form part of potable water supply for Penang state in Malaysia. Agricultural activities on the steep slopes of the watershed may contribute to significant nutrient loads into the Lake Penang. These threats have made the state government to increase the physical boundary height of Lake Penang and negotiate for an inter-state transfer of water from neighbouring states as a short to medium term plan for securing water resources.

2.10 The Gaps on Tropical Lakes Modelling

The nature of nutrient and phytoplankton dynamics in tropical lakes needs to be taken into account by any proposed modelling attempt. However, since the majority of ecological models were developed for temperate lakes, applications of these models for tropical lake ecosystems are rare. The capability of these models in predicting algal growth and nutrient dynamics in the tropical areas may be limited. Lampert and Sommer (1997) reported that the difference between temperate and
tropical conditions has made the temperate based lake seasonal succession model of phytoplankton and zooplankton applicability to tropical lakes questionable since it has not been adequately examined. According to Ortiz et al. (2006), the OECD criteria (OECD, 1982) to determine the trophic status of lakes are largely inapplicable to tropical lakes because N has been found as the limiting nutrient for algal growth in Latin American lakes in contrast with P-limitation in those of OECD study. Thus, eutrophication control in Latin America lakes using OECD predictive equations is recommended to be used with utmost cautions.

Even though there is a foundation of understanding of the mechanisms by which phytoplankton communities grow, there remains a hiatus between the basic scientific knowledge of eutrophication processes in tropical lakes and the ability to apply this modelling approach to the practical requirements of water authorities and legislators for decision making. On top of this, model applications for environmental management issues on a local scale seem to have languished with the pertinent papers comprise a smaller portion of the published modelling literature and receive lower citations (Arhonditsis et al., 2006).

Numerous studies have been documented on the subject of tropical and subtropical lake limnology (see Tundisi, 1990; Salas & Martino, 1991; Figueredo & Giani, 2009). Tundisi (1990) identified basic characteristics of tropical and subtropical reservoirs which necessitate modelling effort. However, the modelling for eutrophication risk assessments in tropical lake ecosystem are limited to studies based principal component analysis (Parinet et al., 2004), simplified mathematical models (Salas & Martino, 1991), simplified nutrient-algae dynamics models (Jaffe, 1988), TP-chlorophyll empirical models (Huszar et al., 2006), multivariate correlations (Sarawuth et al., 2008), neural network based and fuzzy logic models (Malek et al., 2010). No complex model representing tropical lakes ecosystems was found in the literature.

2.10.1 Data Availability and Quality

Despite the increasing attention on limnological studies of tropical lakes, the scope on phytoplankton and nutrient dynamic remains very limited and the lack of long-term environmental data serves to compound the ambiguity surrounding our knowledge of process dynamics in the water bodies. Models capability in predicting measured or observed data is of paramount importance for data-driven and process-based models such as HEA and SALMO-OO. Ideally, data of good quality should be comprehensive, having long historical sequence and measured on constant time frame. However, in Malaysia, chances of acquiring generic datasets suitable for a variety of modelling exercises are very limited. Reasons being, that data sampling is tedious, time consuming and cost prohibitive to maintain for a long time frame. Logistical and financial aspects constitute most to the issues regarding data collection, availability and uncertainty in the tropic.

Scarcity of information is a problem which is typically encountered by agencies with responsibilities for managing lake water quality at a regional level (Walker, 1982). Indeed, the main difficulties encountered in modelling the tropical
water bodies in tropics and especially Malaysia are the limited data recorded and the high variability of parameters chosen by respective researchers. The lack of suitable data not only increases the assumptions within the model but limits the confidence of the modelling results (van Tongeren, 1995). Apart from a good quality of reliable input data, imposing the correct descriptions of the components and processes for any lakes functioning is essential for ecological models to perform well. As Jørgensen and Bendoricchio (2001) pointed out, modellers must also be concerned with the right description of the system properties where too little research has been done in this direction. Lack of suitable data contributed to limited confidence on the forecasted results or model findings and force assumption to be made. Hence, there is an immediate need to close the gap on sparse limnological research conducted on tropical water bodies in Malaysia.

2.10.2 Research Efforts

On regional basis, Komatsu et al. (2007) works had shown that global warming influences trophic lake conditions, further promoting algal growth and changing the aquatic ecosystems in Japan. However, the majority of limnological studies on water bodies in Malaysia are yet to gain attention from international publications. Although there has been a significant amount of lake ecological research conducted in Malaysia, most of it is in the form of short-term, small scale studies on limited ecological components. This has made local expert and agencies to publish their findings in local or regional journals only. Much research especially from postdoctoral studies are not been published on the local journals at all. Failure of local modelling research to feature in international publications could also be due to the fact that data availability is scarce in Malaysia.

There are significant gaps on information of water bodies in Malaysia due to poor coordination between research efforts among experts as well as among authorities. Lack of local expert and suitable agency to monitor on long term basis are other popular reasons. It is a common thing for research in Malaysia to be conducted within a year or two time frames only and on ad-hoc basis. In order to address problems and challenges in limnological research in Malaysia, the National Hydraulic Research Institute of Malaysia (NAHRIM) (2007) is currently coordinating cooperation between researchers and water authorities and also to reduce redundancy in research fields.

2.10.3 Malaysian Scenario

Intensive lake monitoring studies are not feasible in Malaysian context due to the large numbers of lakes which must be considered. With limited studies on selected water bodies, gaps in nutrients causal and effects in relation to the hydrology and ecology of tropical lakes in Malaysia are still unclear to the researchers and should be studied and adequately addressed. Evaluation on the existing literature has identified that the availability of reliable bathymetric information and biological data as among the limitations to the utilisation of process-based ecological models for
lakes in Malaysia. Because of these limitations, the lake health status is subject to uncertainty. Filling these data gaps will allow the use of models to better predict ecosystem response in the area.

However, all is not lost for tropical lake modelling study in Malaysia. Indeed, a substantial knowledge gained through the limnological studies in the tropical region has added up to the current knowledge of lakes in Malaysia. However, despite numerous monitoring conducted in limnological aspect of Malaysian water bodies for the half century in an effort to understand the lake ecosystem, no phytoplankton or nutrient dynamic modelling study has ever been applied to any of the three lakes in this study.

The consequences of harmful algal blooms entail high cost incurred by the local authorities. The removal of planktonic biomass form in water storage reservoir for potability, efficient hydropower generation and safe water contact activities is costing a lot of money (Ho, 1996; Omran, 2011). On top of that, the uneven distribution of rainfall and water resources renders some periods and parts of Malaysia too dry and others too wet (Weng et al., 2006).

Therefore, gaps in knowledge of the scenario analysis and eutrophication prediction of the tropical lake ecosystem still exist and are yet to be addressed. This can affect the utilization of sound ecological models because profound knowledge on the lake ecosystem, the problem, the ecological components and the reactions of the ecosystem on the system level are the necessary basis. When progressive nutrient enrichment appears to be the cause of the additional plant operating cost, managers need to know what causes the problem and what corrective action is effective. This is where ecosystem model for phytoplankton dynamic is sought after.

2.11 Recent Trends in Ecosystem Modelling

Most of the recent effort on ecosystem modelling is to increase the realism of ecosystem models by expending the description of phytoplankton (in temperate lakes) and focusing on measurement of changes in nutrient concentrations (in sub-tropical lakes) (Fragoso et al., 2008). Maar et al. (2011) constructed a common ecological model (ERGOM) for different water bodies (North Sea and Baltic Sea). This approach is similar to this study where a generic rule was derived from experiments on two separate lakes having similar trophic level and climatic conditions. A new approach called ‘structural dynamic modelling which allowed for varying numbers of variables and a model structure that could change during simulation time was developed by Nielsen and Jørgensen (2011). Recent advance in two-dimensional mathematical ecological models showed further separation of algae into three groups and incorporation of detailed zooplankton sub models (e.g. Zhang et al., 2008).

After the object-oriented implementation of the model SALMO (Benndorf & Recknagel, 1982) by Zhang (2006), the model continues to become subject of various enhancement studies. Cetin (2007) has incorporated into SALMO-OO, alternative process representations for algal growth and grazing as well as zooplankton growth
and mortality adopted from Park et al., (1974), Arhonditsis & Brett (2005) and Hongping & Jianyi (2002). Additions of process models into the SALMO-OO library is also one of the contributions of this study.

Cao et al. (2008) integrates EA into SALMO-OO in a multi-objective approach to determine optimum temperature or phosphate functions for selected parameters of the model and specific for certain lake categories. A hybrid approach was implemented by integrating EA into the complex ODE. The EA works by evolving temperature and phosphate functions for parameters related to photosynthesis, respiration and zooplankton grazing for the three algal groups (diatoms, green algae and blue-green algae) simulated in SALMO-OO. Results showed that the simulation accuracy of SALMO-OO on seasonal biomass dynamics of the algal groups has improved significantly after replacing the constant parameters with optimum functions.

Optimisation of crucial parameters for specific climate and environmental conditions of lake from different categories are crucial in reducing inaccuracy and uncertainty to models when constant parameter values for lakes of different trophic status and climatic conditions were applied. Manual calibration of SALMO-OO parameters by tuning the values within their observed literature ranges as performed by Cetin (2007) are tiresome in view of 128 constant parameters which are causally related to ecological, chemical and transport processes in lakes that needs to be considered. These values that are either from experimental and field experiment or adopted from the literature can be optimised automatically by means of PSO to suit different lakes environment. This study is a novel attempt intended to apply PSO for optimising selected crucial parameters in the SALMO-OO model. This model is called SALMO-PLUS. The advantage of PSO in this study is that both model structure and parameter can be optimised for better results in SALMO-PLUS.

2.12 Summary

The role of ecological modelling has been recognised among the expert. Characteristics of lake ecosystems in terms of species interaction, system process and flux dynamics can be predicted by ecological models. Highly complex and nonlinear ecological time series data can be successfully unravelled and predicted by means of biologically inspired computation such as HEA. Process-based models such as SALMO-OO can help in describing physical and chemicals processes influencing population dynamics in water bodies and illustrate the effects of nutrient fluctuations on the lake ecosystem functions.

Tropical lakes in this study are meant to cater for various and different purposes but share the same increasing eutrophication threat and competing user-demands. These lakes require predictive computational modelling techniques for improved understanding on its ecosystem dynamics.
The first part of this study is to test whether HEA can elucidate algal growth dynamics in tropical water bodies by means of rule discovery based on ecological time series data input. The predictive and explanatory models generated by HEA can be used for proactive management and decision support tools. The second part will be to assess whether SALMO-OO can predict the algal dynamic and simulate trophic conditions for study lakes. The last part will be to investigate whether DYRESM can simulate the thermal profile, salinity and density for Lake Putrajaya. A single ‘right’ approach does not exist and instead, multiple modelling approaches, applied concurrently to a given problem, can help develop an integrative view on the functioning of lake ecosystems (Mooij et al., 2010). Output from these models can be used to analysis management approach for better decision making and management plans in the study lakes.
CHAPTER 3

3.0 MATERIALS AND METHODS

3.1. Study Sites and Data Acquisition

The study sites in this research cover three different but significant uses of tropical water body in Malaysia. All the three tropical water bodies in this research are man-made with various inputs of management measures for the purpose of day-to-day operation. Lake Kenyir is the deepest lake among the three and is being used for hydropower generation and flood retention purposes. Lake Penang was constructed for drinking water strategic storage as well as for flood mitigation uses. Lake Putrajaya is the latest shallow, man-made lake in Malaysia that was constructed for recreational uses and aesthetic functions. Study site locations of Lake Penang, Lake Putrajaya and Lake Kenyir are shown in Figure 1.

![Study site: locations of Lake Penang, Lake Putrajaya and Lake Kenyir](image)

**Figure 1** Study site: locations of Lake Penang, Lake Putrajaya and Lake Kenyir

Lake Putrajaya and Lake Kenyir are type of reservoir formed by the accumulation of flowing water behind a constructed dam (river reservoir) while Lake Penang was created by pumping water into an artificial impoundment formed by a
manmade embankment around its entire perimeter to create a fully bunded dam (storage reservoir).

The climate of Malaysia is characterised by a humid tropical climate with uniform average daily temperatures of 21-32°C and high humidity averaging about 85% (Meteorological Department, 2007). Equatorial areas such as Malaysia are characterised by high humidity and temperature and abundant rainfall (Ho, 1996). However, there is a slight but definite increasing warming trend in the mean annual temperature records in Malaysia (Ahmed & Jahi, 2004; Salmah et al., 2007) consistent with global temperature trends.

Generally, rainfall pattern is affected by the north-east (November-March) and south-west (June-August) monsoons which bring heavy rainfall. Periods for the months of April until May and from September until October are expected to experience less rain due to changes in monsoonal winds.

Due to small seasonal variation in incoming solar radiation, the annual difference in day length is very minimal and a day length of 12.30 hours year round is regular. It can be assumed that Malaysia has abundant sunshine and solar radiation but a totally clear sky during the whole day is uncommon. General characteristics of Lake Putrajaya, Lake Kenyir and Lake Penang are given in Table 1.

Table 1 General characteristics of study lakes

<table>
<thead>
<tr>
<th>Item</th>
<th>Lake Putrajaya</th>
<th>Lake Kenyir</th>
<th>Lake Penang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of impoundment</td>
<td>2002</td>
<td>1986</td>
<td>1984</td>
</tr>
<tr>
<td>Surface area (km²)</td>
<td>4</td>
<td>369</td>
<td>1.74</td>
</tr>
<tr>
<td>Storage volume (m³)</td>
<td>26.5 million</td>
<td>13.6 billion</td>
<td>23.6 million</td>
</tr>
<tr>
<td>Maximum depth (meters)</td>
<td>14</td>
<td>145</td>
<td>44</td>
</tr>
<tr>
<td>Mean depth (meters)</td>
<td>6.6</td>
<td>37</td>
<td>17.9</td>
</tr>
<tr>
<td>Water residence time</td>
<td>132 days</td>
<td>n.a</td>
<td>n.a</td>
</tr>
<tr>
<td>Catchments area (km²)</td>
<td>51</td>
<td>380</td>
<td>3.9</td>
</tr>
<tr>
<td>Lake shoreline stretches</td>
<td>34.0 km</td>
<td>n.a</td>
<td>n.a</td>
</tr>
<tr>
<td>Elevation (a.s.l.)</td>
<td>31</td>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>Water level</td>
<td>RL 21 meters</td>
<td>n.a</td>
<td>43.3m a.s.l</td>
</tr>
</tbody>
</table>

n.a = not available

3.1.1 Limnological Time Series Data of Lake Putrajaya

Monitoring of water quality, hydrological and meteorological data in Lake Putrajaya for the purpose of this research was conducted for a one-year period beginning from April 2008 with weekly and monthly sampling. A monitoring site with the deepest depth and situated just before the lake outflow was chosen as the sampling site for this study (see station PLg2 in Figure 2).
The Hydrolab multi-probe data logger was used for measurements of temperature, pH, conductivity, salinity and dissolved oxygen for this water research. Sampling procedures and analytical methods for the ex situ measured parameters in Table 2 were adopted from the manual published by the APHA (1995).

Meteorological data such as rainfall, wind speed, humidity, solar radiation and percentage of possible hours of sunlight were obtained from the weather station closest to the sampling site. Only wind speed data was obtained from the Meteorological Department weather station located about 40 kilometres from sampling site. Meteorological data used as inputs into the DYRESM model were given in daily time-step. The meteorological data was also used to compliment the biological and physico-chemical input data in order to investigate the phytoplankton dynamics in the three lakes by means of HEA. Hydrological data such as daily flow volumes from tributaries into the lake were obtained online via a telemetric sampling station at each tributary’s arm.

**Table 2** List of parameters measured *in situ* and *ex situ* in Lake Putrajaya

<table>
<thead>
<tr>
<th>Measured parameters</th>
<th>Unit</th>
<th>Test Method</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>in situ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water Temperature</td>
<td>ºC</td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Secchi depth</td>
<td>m</td>
<td>Secchi disc</td>
<td>weekly</td>
</tr>
<tr>
<td>pH</td>
<td></td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>mg/l</td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Turbidity</td>
<td>NTU</td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Conductivity</td>
<td>µS cm⁻¹</td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Total Dissolved Solids</td>
<td>mg/l</td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td>Salinity</td>
<td></td>
<td>Hydrolab multi-probes</td>
<td>weekly</td>
</tr>
<tr>
<td><strong>Ex situ</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ammoniacal-nitrogen</td>
<td>mg/l</td>
<td>APHA 4500-NH₃ G</td>
<td>monthly</td>
</tr>
<tr>
<td>Nitrate-Nitrogen</td>
<td>mg/l</td>
<td>APHA 4500-NO₃⁻ H</td>
<td>monthly</td>
</tr>
<tr>
<td>Phosphate-Phosphorus</td>
<td>mg/l</td>
<td>APHA 4500-P C</td>
<td>monthly</td>
</tr>
<tr>
<td>Biochemical Oxygen Demand</td>
<td>mg/l</td>
<td>APHA 5210 B</td>
<td>monthly</td>
</tr>
<tr>
<td>Chemical Oxygen Demand</td>
<td>mg/l</td>
<td>APHA 5220 C</td>
<td>monthly</td>
</tr>
<tr>
<td>Total Suspended Solids</td>
<td>mg/l</td>
<td>APHA 2540 D</td>
<td>monthly</td>
</tr>
<tr>
<td>Chlorophyll-α</td>
<td>µg/l</td>
<td>APHA 10200 H (3)</td>
<td>monthly</td>
</tr>
</tbody>
</table>
Vertical water temperature, salinity and dissolved oxygen profiles in Lake Putrajaya were recorded at 1-m intervals at station PLg2. This was measured by manually casting the probe according to the 1-meter depth specification during the diurnal sampling. Thermocline depths were defined as the top of the 1-m measurement interval at which the greatest temperature gradient occurred. Stream water temperatures were estimated from wetland water quality data.

Historical data collected prior to the sampling period for this study was obtained from Alam Sekitar Malaysia Pty. Ltd. This certified laboratory also applied methods provided by APHA (1995). Historical data from Lake Putrajaya was used for modelling experiments by means of HEA.

**Data pre-processing**

The limnological time series data from Lake Putrajaya has been sampled by a team of researchers since 1999. Data pre-processing was done to get a clean data, free from missing values by means of linear interpolation. Incident of inconsistency in the data set was treated by cautious removal of the identified sample among the variables. Time series data plot was used in order to confirm the historical pattern and current trends in the data sets and to check for any anomalies.

Lake Putrajaya is situated in Putrajaya City at latitude 2° 55’ 11” and longitude 101° 41’ 24” (Figure 2). The lake was among the first completed projects in the development of Putrajaya City as new administrative centre for the federal government of Malaysia and serve as the main features of this new city. Thus, the lake is increasingly scrutinised by stakeholders due to this prime location.

Lake Putrajaya was constructed by damming and inundating the 400 hectares valleys of small streams (namely River Chuau and River Bisa) creating a shallow tropical lake which is previously rubber estates with topography ranges between 8 to 152 meters. Physical constructions started in 1997 in two phases, completed in May 1998 and the lake was fully inundated by January 1999. Both phases were physically connected from the 26th of August 2002. A year round water supply for this lake comes from the wetland tributaries on the upstream as well as from the two monsoons season’s characteristic of this area. Annual rainfall average is 2,839 mm while annual evaporation average is 1,720 mm. This lake is categorised as mesotrophic shallow-polymictic tropical lake in this research.

The catchment area of the lake is 51 square kilometers and most of the upstream areas were mostly undeveloped when the lake was constructed. The total volume of the whole lake is 23.5 million cubic meters and the water depth is within the range of 3.0 to 14.0 meters. Minimal seasonal fluctuations of its water level characterised this lake, resulted in a mean water residence time of 132 days. According to studies by NAHRIM (2007), the lake was classified as medium in terms of eutrophication status.

The lake has been primarily designed to enhance the aesthetic appeal of Putrajaya City and to be used for sport and recreation as well as a tourist attraction (Ho, 2006). Apart from its multi-functional usage, Lake Putrajaya is also popular as
recreational spot in a waterfront city with an average of 300,000 foreign tourists visits annually (BERNAMA, 2006).

The catchment area of Lake Putrajaya lies within the jurisdiction of three separate local authorities, thus results in a complicated management issues due to different development priority of different authorities. About 60% of the lake water flow from the wetland and the remaining 40% is the direct discharge from bordering promenade area. Among others, catchment area consists of a public golf course, a university hostel and an agricultural research centre. Certain activities within the catchment areas are known to apply agricultural supplies that sporadically cause algal bloom in the drainage system leading towards the lake.

Current status
A system of wetlands was built on the upstream part of the lake to treat surface runoff caused by agricultural activities and to counter threats from future developments upstream. Presently, the wetlands have managed to filter and contain algal bloom, preventing it from spreading into Lake Putrajaya. However, major changes had happen in the catchments since the completion of the lake that include the development of upstream areas into a commercial area as well as increase in agriculture practises. All these changes are observed to alter the hydrology of the

Figure 2 The description of the study site in Lake Putrajaya (from PJC, pers. comm., 2000). Data from sampling location PLg2 was used for this research

NOTE:
This figure/table/image has been removed to comply with copyright regulations. It is included in the print copy of the thesis held by the University of Adelaide Library.
wetland ecosystem and eventually Lake Putrajaya itself. The overall capacity of the wetlands to filter the incoming pollutants brought by the increasing run-off and incoming pollutant concentration are also in doubt. Owing to these scenarios, Putrajaya Wetlands as the first man-made wetlands in Malaysia have attracted much scientific attention to understand the hydrology and ecology of the ecosystem (Sim et al., 2008). Incidents of overflowing and faulty sewage treatment plants located in the city and neighbouring hostel were also reported to cause sporadic algal bloom in drains leading to the lake (PJC, pers. com. 2010).

Ironically, no similar interest was observed for the lake. Future capacity of the wetland to withstand continuing occurrence of pollution is doubtful as it is acknowledged that lake ecosystems are known to evolve and driven by exogenous forces rather than existing permanently and in isolation (Recknagel, 2003a).

This research will utilize HEA model in investigating the eutrophication processes with respect to phytoplankton growth and nutrient cycling in Lake Putrajaya. In addition, this research will use SALMO-OO models to allow simulation and prediction of the nutrient loadings and algal functional groups dynamics in Lake Putrajaya. Application of both models will allow swift assessment of water quality and trophic status in the lake.

3.1.2 Lake Kenyir

Limnological time series data for 1992 was monthly collected in Lake Kenyir by Yusoff and Lock (1995) and was used for HEA and SALMO-OO models in this study. Data from the deepest monitoring site and situated closest to the dam outflow was chosen for this study (see station 3 in Figure 3). Water temperatures and dissolved oxygen profiles in Lake Kenyir were sampled using YSI dissolved oxygen meter equipped with a thermistor. Hydrological and meteorological data for this study was obtained separately from weather station owned by the hydropower generation authority in Lake Kenyir.

Sampling procedures and analytical methods for the ex situ measured parameters in Table 3 were adopted from the manual published by the APHA (1995).
Table 3 List of physico-chemical and biological parameters measured in situ and ex situ in Lake Kenyir and used for this study

<table>
<thead>
<tr>
<th>Measured parameters</th>
<th>Unit</th>
<th>Test Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>in situ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water Temperature</td>
<td>ºC</td>
<td>YSI meter</td>
</tr>
<tr>
<td>Secchi depth</td>
<td>m</td>
<td>Secchi disc</td>
</tr>
<tr>
<td>pH</td>
<td>-</td>
<td>YSI meter</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>mg/l</td>
<td>YSI meter</td>
</tr>
<tr>
<td><strong>Ex situ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nitrate-Nitrogen</td>
<td>mg/l</td>
<td>APHA 4500-NO$_3^-$ H</td>
</tr>
<tr>
<td>Phosphate-Phosphorus</td>
<td>mg/l</td>
<td>APHA 4500-P C</td>
</tr>
<tr>
<td>Chlorophyll-α</td>
<td>µg/l</td>
<td>APHA 10200 H (3)</td>
</tr>
</tbody>
</table>

Located in the east coast of Peninsular Malaysia (4° 47’ to 5° 15’N and 102° 32’ to 102° 55’ E) at an elevation of 145 m above sea level, Lake Kenyir was formed when steep rainforest valleys were flooded behind a dam across the River Terengganu (see Figure 3). The dense hilly tropical rainforest inside the catchment area for Lake Kenyir is drained by more than 40 rivers. There are more than 340 small islands spread out in water catchments area of this multipurpose hydroelectric power and flood mitigation scheme. The construction of Kenyir hydroelectric dam was started in 1978. When completed in 1985, it had created Lake Kenyir which is the biggest man-made lake in Southeast Asia region. A power station complex at this lake is supplying 100 MW (megawatts) power units of electricity for entire Peninsular Malaysia. The north and north-eastern part of this reservoir is fairly flat palm oil plantation area while the rest includes the tropical rainforest reserve area as well as part of a national park.

The area has a humid tropical climate dominated by the north-east monsoon (November-January) and influenced very slightly by the southwest monsoon (May-October). Lake Kenyir generally experiences heavy rain from the north-east monsoon. Temperatures are relatively constant throughout the year otherwise it is always hot and dry season. The submerged tropical rainforest during the early stage of the dam construction generates a considerable amount of organic matter for decomposition which causes nutrient enrichment in the reservoir (Yusoff et al.,1995). The lake has a complex morphometry and its water is enriched by decomposition of drowned forest biomass but strong thermal stratification ensures that most nutrients are trapped in the anoxic hypolimnion (Yusoff et al.,1995).

Lake Kenyir is a deep strongly stratified mesotrophic tropical lake with surface temperature ranging from 24.2°C to 32.0°C and bottom temperature ranging from 28.0°C to 24.0°C (Yusoff & Lock, 1995). As a result, nutrient rich anoxic
waters are locked up in hypolimnion and are not available for utilization in the surface water (Yusoff et al., 1995). The ratio of catchment to the water surface area of Lake Kenyir is expecting to exert moderate influence in terms of catchment to water body effect. However, high rainfall volume is expected to increase suspended and dissolved materials entering the lake.

Figure 3 The description of the study site in Lake Kenyir (from Yusoff (1995)). Data from sampling station number 3 was used for this research

Current status

Although the lake was constructed for the generation of hydroelectric power, it has evolved to become a multipurpose reservoir supporting small scale fisheries, aquaculture, flood control, transportation, recreation and ecotourism. Numerous infrastructure developments and expansion in and around the lake to cater for increasing tourists without proper planning may increase the potential of eutrophication. In addition, the state government, with federal funding is proposing to turn Lake Kenyir into a duty-free zone (Choong, 2010). A study by local university revealed that the level of E. coli in the lake had been rising and that it was at a worrying level. The main threat is pollution from logging, chalets, boathouses and uncontrolled terrain development. On top of that, there is a golf course right next to the lake.
User conflict typically arise when a series of reservoir are operated for flood control, hydropower generation and recreation including fish and wildlife propagation (Ford & Thornton, 1992). On one hand, flood control requires water storage during periods of high flow, followed by quick release of this water to warrant satisfactory storage is available for future high flow events. On the other hand, water for hydropower generation is stored at optimum storage capacity for a certain period to increase the hydrostatic pressure on the turbines for greater power generation efficiency and to supply adequate water for peaking power generation during those high demand period. In terms of recreation or inland fisheries management, the reservoir pool elevation is adjusted by maintaining water level at certain level to induce fish spawning. However, storing nutrient-enriched flood waters may limit water-contact recreational activities at the lake and cause water quality problems. On top of that, frequent fluctuations during any of these options may impact recreational activities, erode lake shorelines as well as affect fish spawning patterns.

This research will utilize HEA model and SALMO-OO model in investigating the potential eutrophication built up processes with respect to phytoplankton growth and nutrient cycling in Lake Kenyir.

3.1.3 Lake Penang

Limnological time series data from Lake Penang was monthly collected from August 2005 to July 2006 by Makhlough (2008) and was used for HEA and SALMO-OO models in this study. Data from the deepest monitoring site was chosen for this study (see station A2 in Figure 4).

The physico-chemical and biological parameters in Table 4 were collected monthly from the water surface to the bottom in 5 meter intervals over the period from August 2005 to July 2006. The sampling period comprised two different tropical seasons (the dry and the rainy season).

Meteorological data such as rainfall and solar radiation and hydrological data such as inflow and outflow volumes were obtained from the water authority managing the lake.

Lake Penang is officially known as Mengkuang Dam. However the name Lake Penang is used for the purpose of this research. Lake Penang is a reservoir located at latitude N 05’ 21’’ 57” and longitude E 100’ 29’’ 52” in the state of Penang in Malaysia (Figure 4). The catchments area of this lake is predominantly old rubber trees emergent among secondary forests. Constructed in 1982, the reservoir has a height of 24.7 meter. This medium-deep-stratified tropical lake is shaped like a broad bowl set within a range of high hills approximately 160 meters above sea level but is close to population area. The dam is curved downstream to correspond with the topography, in particular a low saddle on the right abutment, but also to enhance the bowl shaped reservoir. The spillway and draw-off work are combined in a single intake tower, directly upstream of the reservoir. The draw-off work comprised of four intakes at different levels of the intake tower (i.e. 21.9, 28.3, 33.5 and 38.7 meter) and
water draw-off is either done via 28.3 or 33.5 meter level pipes (PBAPP, pers. com. 2010).

Table 4 List of physico-chemical and biological parameters measured in situ and ex situ in Lake Penang and used for this study

<table>
<thead>
<tr>
<th>Measured parameters</th>
<th>Unit</th>
<th>Test Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>in situ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water Temperature</td>
<td>ºC</td>
<td>YSI meter</td>
</tr>
<tr>
<td>pH</td>
<td></td>
<td>YSI meter</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>mg/l</td>
<td>YSI meter</td>
</tr>
<tr>
<td>Secchi depth</td>
<td>m</td>
<td>Secchi disc</td>
</tr>
<tr>
<td><strong>ex situ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate-Phosphorus</td>
<td>mg/l</td>
<td>APHA</td>
</tr>
<tr>
<td>Nitrate-Nitrogen</td>
<td>mg/l</td>
<td>APHA</td>
</tr>
<tr>
<td>Chlorophyll-a</td>
<td>µg/l</td>
<td>APHA</td>
</tr>
<tr>
<td>Ammoniacal-nitrogen</td>
<td>mg/l</td>
<td>APHA</td>
</tr>
<tr>
<td>Biochemical Oxygen Demand</td>
<td>mg/l</td>
<td>APHA</td>
</tr>
<tr>
<td>Chemical Oxygen Demand</td>
<td>mg/l</td>
<td>APHA</td>
</tr>
</tbody>
</table>

Lake Penang was constructed as part of the Mengkuang Pumped Storage Scheme to provide an additional 370,000 m³/day of raw water for supply to Penang state and the surrounding areas (Taylor *et al.*, 1992). The Mengkuang Pumped Storage Scheme operates by obtaining river water supply from the River Muda just upstream of tidal barrage. The water from the river is pumped into the dam for storage via pipeline during the rainy season between April and December at about 440 million litres a day. During the dry seasons (from January to March) water from this reservoir is pumped back to the River Dua Treatment Plants (WTP) located at 13 kilometres downhill through the same pipeline at a maximum capacity of 1.114 billion litres a day, to make up for the low water supply from the river.

River Dua WTP is currently the largest WTP in Penang state. It is able to produce 770 million litres per day (MLD). About 500 MLD of the production is distributed to mainland area (Seberang Perai) and the balance of 270 MLD is distributed to Penang Island. The working capacity of River Dua WTP was increased to a total of 750 ML water per day to meet the state’s water need up to the year 2010 (Ho, 1996). However, the population of Penang is projected to increase from 1.48 million in 2010 to 2.42 million in the year 2050 (DOE, 2009). To be able to cater for growing population, agricultural and industrial demands until 2020, Lake Penang has undergone a number of developments and the current expansion project would increase the current capacity of 23.6 billion litres to 78 million cubic metres. It
is anticipated that the expansion of Lake Penang would also minimise the impact of flooding in the northern part of Penang and the neighbouring state of Kedah and alleviate the effect of drought seasons in the state itself (DOE, 2009). This scheme also functions as flood mitigation measures by pumping excess wet season flow from River Muda and stores it for use in the critical dry season especially from the months of February to May every year.

Figure 4 The description of the study site in Lake Penang (after Makhlough, 2008). Data from sampling location A2 was used for this research. Mengkuang Reservoir is the official name for Lake Penang.

Current status
Presently, Lake Penang is the biggest lake in term of capacity for the state of Penang and supplies enough water for the whole state population. The reservoir area has been declared a National Park by the Government of Malaysia (Taylor et al., 1992). The biggest threat for the dam is encroachment into water catchments areas through agricultural activities, camping activities, illegal burning and squatters. Since the state of Penang gets 80% of its raw water from the River Muda which flows through agricultural areas before pouring into WTP for treatment, serious attention is needed to prevent any incident of water quality deterioration especially eutrophication.
This research will utilize HEA model in investigating potential algal bloom and SALMO-OO model in studying the nutrient dynamic and phytoplankton functional groups fluctuation in Lake Penang.

3.2 SALMO-OO (Simulation by means of an Analytical Lake MOdel)

The lake simulation library SALMO-OO is the object-oriented implementation of the model SALMO (Benndorf & Recknagel, 1982; Recknagel & Benndorf, 1982) by the JAVA programming language (Zhang, 2006; Recknagel et al., 2008a; Recknagel et al., 2008b;). As a result, SALMO has been rebuilt into a portable, user-friendly and easily adapted software package, called SALMO-OO (Cetin, 2007; Recknagel et al. 2008a,b).

SALMO-OO is a system of ordinary differential equations (ODE) that represent key ecological processes in order to determine mass balances of inorganic nutrients (nitrate and phosphate), detritus, oxygen, zooplankton and phytoplankton.

Through continuous improvements over the past decade, SALMO-OO is now applicable for non-shallow lakes (maximum depth > 5m) and able to simulate daily dynamics of phytoplankton, zooplankton, oxygen, phosphates, nitrate and detritus concentrations in response to lake-specific data of water temperature, solar radiation and nutrient loadings measured for one particular year (Zhang, 2006). A broad range of water bodies with diverse morphometry (stratified temperate lakes with ice cover in winter to thermally stratified Mediterranean lakes in summer) and trophic levels (from oligotrophic to hypertrophic conditions) has been successfully simulated by this model.

This dynamic water quality model for freshwater lake was originally developed by Benndorf and Recknagel (1982) and Recknagel and Benndorf (1982) and was used for scientific and management purposes of lacustrine resources beginning from temperate European lakes and now broadly expanded to support eutrophication management in Australia (e.g. Walter et al., 2001). The model SALMO-OO has also been successfully applied to lakes in South Africa (Recknagel et al., 2008b) and South Korea (Recknagel et al., 1995). Detailed descriptions and explanations about SALMO-OO are given by Cetin (2007) and Newman (2008) and all the following terms are based on the two authors.

The two-layer model of SALMO-OO is capable of simulating the epilimnion and hypolimnion of stratified water bodies by assuming that the concentrations of biomass or matter are homogenously distributed within the water column or layer. This system of ODE is numerically solved using the fourth order Runge-Kutta method. The model structure of SALMO-OO supports a variety of options for the simulation of scenarios, either of theoretical nature or management relevance (Walter et al., 2001). Figure 5 shows the simplified structure of the model SALMO-OO with six rectangles representing the state variables. These state variables represent the left-hand parts of an ODE.
3.2.1 The Overview of SALMO-OO

Figure 5 Structure diagram of the model SALMO-OO illustrating all the state variables and key processes simulated by the model (from Walter et al. (2001))
Changes in state variables are determined by the process equations (i.e. growth or mortality) which in turn depend on respective rate variables and represented by the circle shapes. The curved arrows connecting the state variable and the rate variables represent the causal relationships between the two. The incoming straight arrows indicate the increasing of the mass from a source (i.e. zooplankton growth) and the outgoing straight arrows indicate the decreasing mass to a sink (i.e. zooplankton mortality). The straight arrows on the extreme left represent the inputs while the outputs are represented by arrows on the extreme right.

The mass balance of all state variables include influx, outflux and vertical flux during stratification. Nutrient (nitrate and phosphate) mass balances include consumption by the three algal groups (diatoms, green and blue-green algae), sediment release under anaerobic conditions as well as remineralisation resulting from algal grazing and zooplankton mortality. Effects of fish are considered indirectly in the zooplankton mortality rate. Co-precipitation, nitrification and denitrification processes are included as well. SALMO-OO has 128 constant parameters that were defined and specified for numerous process equations. Among the driving variables in this model are:

1. Nutrient (nitrate, phosphate) concentration of streams entering the lake g N/m³; mg P/m³)
2. Volume of water inflow and outflow (m³/d)
3. Water temperature of mixed volume and stratified layers (°C)
4. Mean mixing and maximum water depth (m)
5. Total volume of water body (m³)
6. Incident solar radiation (J/cm².d)

The daily simulation of algal growth begins with the calculation of the mean underwater light intensity from photosynthetic active solar radiation by considering light extinction over lake depth using Lambert-Beers law. The six state variables in SALMO-OO model are represented in Table 5:

Processes simulated for phytoplankton include growth, zooplankton grazing, import and export, exchange between layers during stratification and sedimentation. Phytoplankton state variables are divided into three different ODEs for the simulation of three functional groups:

1. Diatoms
2. Green algae
3. Blue-green algae

Processes simulated for zooplankton include growth, mortality, import, export and migration. The zooplankton state variable represents herbivorous zooplankton and the mortality rate of zooplankton reflects predation by fish and zooplankton. The simulation of the oxygen budget includes import, export, exchange of oxygen between the epilimnion and hypolimnion, oxygen production by phytoplankton and consumption by zooplankton.
Table 5 State variables and processes represented by SALMO-OO (from Recknagel et al. (2008a))

<table>
<thead>
<tr>
<th>SALMO-OO processes</th>
<th>State variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phosphate (mg/m³)</td>
</tr>
<tr>
<td></td>
<td>d P/dt</td>
</tr>
<tr>
<td>Influx</td>
<td>(INF*P)/V</td>
</tr>
<tr>
<td>Outflow</td>
<td>(P*O)/V</td>
</tr>
<tr>
<td>Vertical flux</td>
<td>(Pc,h * VFLUXh,0)/V</td>
</tr>
<tr>
<td>between epi- and hypolimnion</td>
<td>Vh,0</td>
</tr>
<tr>
<td>Production</td>
<td>PSR</td>
</tr>
<tr>
<td>Sediment release</td>
<td>Z*(ZMO*RATN)</td>
</tr>
<tr>
<td>Remineralization</td>
<td>Z*(ZMO*RATN)</td>
</tr>
<tr>
<td>Denitrification</td>
<td>Growth</td>
</tr>
<tr>
<td></td>
<td>AGRO[i]*A[i]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Tab. 6 represents the alternative process models for algal photosynthesis, algal respiration, algal grazing, zooplankton growth and zooplankton mortality that the simulation library of SALMO-OO contains. The Tab 7 indicates the best performing model structures of SALMO-OO identifies for three different lake categories.
Table 6 Process library for algal growth and grazing, zooplankton growth and mortality with AGRO\textsubscript{i,A}, AGRA\textsubscript{i,A}, ZGRO\textsubscript{i,A}, ZMO\textsubscript{i,A}, adopted from Benndorf and Recknagel (1982) AGRO\textsubscript{i,B}, AGRA\textsubscript{i,B}, ZGRO\textsubscript{i,B}, ZMO\textsubscript{i,B}, adopted from Park et al. (1974) AGRO\textsubscript{i,C}, AGRA\textsubscript{i,C}, ZGRO\textsubscript{i,C}, ZMO\textsubscript{i,C}, adopted from Hongping and Jianyi (2002) and AGRO\textsubscript{i,D}, AGRA\textsubscript{i,D}, ZGRO\textsubscript{i,D}, ZMO\textsubscript{i,D}, adopted from Arhonditsis and Brett (2005) (from Recknagel et al. (2008a))

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algal growth</strong></td>
<td>AGRO\textsubscript{i,A} = PHO\textsubscript{[i,A]} - RA\textsubscript{i,A}</td>
<td>AGRO\textsubscript{i,B} = PHO\textsubscript{[i,B]} - RA\textsubscript{i,B}</td>
<td>AGRO\textsubscript{i,C} = PHO\textsubscript{[i,C]} - RA\textsubscript{i,C}</td>
<td>AGRO\textsubscript{i,D} = PHO\textsubscript{[i,D]} - RA\textsubscript{i,D}</td>
</tr>
<tr>
<td><strong>Pho. synthesis</strong></td>
<td>PH\textsubscript{[i]} = (PHOPT\textsubscript{MAX,A} - (EA\textsubscript{C,A} / PHOPT\textsubscript{MAX,A}) \times (PH\textsubscript{[i]} / PHOPT\textsubscript{MAX,A})</td>
<td>PH\textsubscript{[i]} = (PHOPT\textsubscript{MAX,A} - (EA\textsubscript{C,A} / PHOPT\textsubscript{MAX,A}) \times (PH\textsubscript{[i]} / PHOPT\textsubscript{MAX,A})</td>
<td>PH\textsubscript{[i]} = (PHOPT\textsubscript{MAX,A} - (EA\textsubscript{C,A} / PHOPT\textsubscript{MAX,A}) \times (PH\textsubscript{[i]} / PHOPT\textsubscript{MAX,A})</td>
<td>PH\textsubscript{[i]} = (PHOPT\textsubscript{MAX,A} - (EA\textsubscript{C,A} / PHOPT\textsubscript{MAX,A}) \times (PH\textsubscript{[i]} / PHOPT\textsubscript{MAX,A})</td>
</tr>
<tr>
<td><strong>Zoo. growth</strong></td>
<td>ZGRO\textsubscript{i,A} = GMAX\textsubscript{[i,A]} / (KAG\textsubscript{[i,A]} + 1)</td>
<td>ZGRO\textsubscript{i,B} = GMAX\textsubscript{[i,B]} / (KAG\textsubscript{[i,B]} + 1)</td>
<td>ZGRO\textsubscript{i,C} = GMAX\textsubscript{[i,C]} / (KAG\textsubscript{[i,C]} + 1)</td>
<td>ZGRO\textsubscript{i,D} = GMAX\textsubscript{[i,D]} / (KAG\textsubscript{[i,D]} + 1)</td>
</tr>
<tr>
<td><strong>Zoo. mortality</strong></td>
<td>ZMO\textsubscript{i,C} = KIN\textsubscript{[i,C]} + N\textsubscript{[i,C]}</td>
<td>ZMO\textsubscript{i,B} = KIN\textsubscript{[i,B]} + N\textsubscript{[i,B]}</td>
<td>ZMO\textsubscript{i,C} = KIN\textsubscript{[i,C]} + N\textsubscript{[i,C]}</td>
<td>ZMO\textsubscript{i,D} = KIN\textsubscript{[i,D]} + N\textsubscript{[i,D]}</td>
</tr>
</tbody>
</table>

Table 7 Three lake categories and their best performing structure of the algal mass balance equations of SALMO-OO (from Recknagel et al. (2008a))

<table>
<thead>
<tr>
<th>Lake categories</th>
<th>Lake examples</th>
<th>Best performing algal mass balance equations: (da_0^i/dt)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Warm monomictic and hypertrophic</strong></td>
<td>Hartbeespoort (South Africa)</td>
<td>+AGRO\textsubscript{i,D}</td>
</tr>
<tr>
<td><strong>Ditric and euxinotic</strong></td>
<td>Kilpoor (South Africa)</td>
<td>-AGRO\textsubscript{i,B}</td>
</tr>
<tr>
<td><strong>Diatom and mesotrophic</strong></td>
<td>Buotzen (Germany)</td>
<td>+AGRO\textsubscript{i,C}</td>
</tr>
<tr>
<td><strong>Anoxic</strong></td>
<td>Aventes (Germany)</td>
<td>-AGRO\textsubscript{i,A}</td>
</tr>
</tbody>
</table>

- 57 -
3.2.2 Data Requirements for the Model SALMO-OO

SALMO-OO uses measured time series of physico-chemical, hydrological and meteorological data that were interpolated into daily values using linear interpolation method in MS Excel to get a complete 365 days of data. The daily input values for one complete year were then averaged to produce a 10-day steps format applicable for SALMO-OO.

The beginning and the duration of thermal stratification also need to be determined. For the tropical lakes Kenyir and Penang, thermal stratification lasted for 12 months. Calculations of total volume, surface area and mixing depths were performed by means of measured water level data, hypsographic curve and temperature profiles supplied by the local water authorities. Data pre-processing was done in order to clean the data and gain better understanding on the nature of the data with regards to SALMO-OO model application. Initial values for state variables and control parameters such as timing of seasons (for non-tropical lakes) were also determined. The input data required to run the SALMO-OO are given in Table 8.

The SALMO-OO output from data was compared with measured data for phytoplankton, nutrient (phosphate and nitrate) concentrations, oxygen and algal biovolume of respective lakes.

Table 8 Input data required for running the SALMO-OO model

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definitions</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Volume of the total water body</td>
<td>m$^3$</td>
</tr>
<tr>
<td>VE</td>
<td>Volume of the epilimnion</td>
<td>m$^3$</td>
</tr>
<tr>
<td>VH</td>
<td>Volume of the epilimnion</td>
<td>m$^3$</td>
</tr>
<tr>
<td>ZMIXREAL</td>
<td>Real mixing depth</td>
<td>m</td>
</tr>
<tr>
<td>ZMIX</td>
<td>Mean depth of the mixed layer</td>
<td>m</td>
</tr>
<tr>
<td>ZHM</td>
<td>Mean depth of the hypolimnion</td>
<td>m</td>
</tr>
<tr>
<td>QIN</td>
<td>Inflow of water from the mixed layer</td>
<td>m$^3$/d</td>
</tr>
<tr>
<td>QHIN</td>
<td>Inflow of water from the hypolimnion</td>
<td>m$^3$/d</td>
</tr>
<tr>
<td>QOUT</td>
<td>Outflow of water from the mixed layer</td>
<td>m$^3$/d</td>
</tr>
<tr>
<td>QHOUT</td>
<td>Outflow of water from the hypolimnion</td>
<td>m$^3$/d</td>
</tr>
<tr>
<td>SRF</td>
<td>Factor for strong rain events (highest ratio between water flow of two consecutive days)</td>
<td>unitless</td>
</tr>
<tr>
<td>I</td>
<td>Incident solar radiation (photosynthetic active radiation)</td>
<td>J/cm$^2$.d</td>
</tr>
<tr>
<td>T</td>
<td>Water temperature in the mixed layer</td>
<td>°C</td>
</tr>
<tr>
<td>TH</td>
<td>Water temperature in the hypolimnion</td>
<td>°C</td>
</tr>
<tr>
<td>PIN</td>
<td>Phosphate concentration of the inflowing water to the mixed layer</td>
<td>mg/m$^3$</td>
</tr>
<tr>
<td>NIN</td>
<td>Concentration of inorganic nitrogen of the inflow to the mixed layer</td>
<td>g/m$^3$</td>
</tr>
<tr>
<td>POMIN</td>
<td>Concentration of particulate organic matter in the inflowing water to the mixed layer</td>
<td>g/m$^3$</td>
</tr>
</tbody>
</table>
3.3 SALMO-PLUS

One of the purposes of this research is to find a model structure of SALMO-OO that performs best for tropical stratified, mesotrophic lakes. This was conducted by searching for the optimal model structure and parameter optimisation by means of SALMO-PLUS. SALMO-PLUS (Martin & Recknagel, 2010) has the capability to optimise the model structure and parameters values by particle swarm optimisation (PSO), and run sensitivity analysis. It is expected that SALMO-PLUS will discover the most suitable model structure and optimum parameters for a specific lake and identify a generic model structure and optimum parameters for a specific lake category.

Parameters to be optimised were chosen from sensitivity analysis where most sensitive parameters were ranked. Eleven parameters were selected for multi objective optimisations that reflect the most influencing parameters to the model. The range was set to 10 to 20% of the initial parameter values. The selection of parameter values, which are optimal in minimisation the error between measured and calculated state variables were adopted as the goal function for the optimisation. The optimisation process itself consists of measuring the objective function for a set of parameter values and then finding the parameter value that minimises goal function (see section 4.5).

These parameters were later subjected to expert fine-tuning according to the range of experimentally determined values reported in the literature and based on the experience and knowledge understanding of model behaviour. List of parameters chosen for optimisation and comparison of their respective values are shown and discussed in the results chapter (Chapter 4).

3.3.1 Particle Swarm Optimisation Method for Model Structure and Parameter Optimisation in SALMO-PLUS

In order to find the optimum model structure and parameter values for SALMO-PLUS during the simulation, Martin & Recknagel (2010) has integrated the PSO into SALMO-PLUS. The open source Java version of JSwarm-PSO package used for this research was obtained from http://jswarm-pso.sourceforge.net. JSwarm-PSO is designed to require minimum effort to use while maintaining its high modularity.

As a population based stochastic optimization approach, PSO operates with a population of random solutions and searches for optimum by updating generations. During the optimization process, the parameter's value may be changed in accordance to its type within an interval, specified by lower and upper bounds. Discrete parameter is represented by a finite set of decisions with essential direction: the parameter influences the objective like a numeric parameter but can take values from the specified set only. It begins at a lower bound and increments by a step size up to an upper bound.
According to Shourian et al. (2008) and Coelho and Mariani (2009), each particle in PSO algorithm is a candidate solution equivalent to a point in a D-dimensional space; hence the $i$th particle’s position can be represented as $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$. Each particle keeps track of its coordinates and flies through the search space, depending on two important positions:

$$p_i = (p_{i1}, p_{i2}, \ldots, p_{iD}),$$

the best position (fitness) of current particle has found so far ($p_{best}$) and

$$p_g = (p_{g1}, p_{g2}, \ldots, p_{gD}),$$

the global best position identified in the entire population ($g_{best}$). $g_{best}$ is tracked by the global version of the particle swarm optimiser and is the overall best value and location obtained so far by any particle in the population. The $g_{best}$ version of PSO is adopted for this research.

The velocity of each particle flying toward its $p_{best}$ and $g_{best}$ locations (global version of PSO) in each time step was set to the default value as in JSwarm-PSO package. The magnitude of the velocity for each particle in the next iteration and position for each particle in the search space may be controlled by users by means of constriction factor in the equations. Acceleration is weighted by random terms, with separate random numbers being generated for acceleration toward $p_{best}$ and $g_{best}$ locations, respectively. These control and maintain particles exploration capability during future iterations and prevent premature convergence of the algorithm while attracting particles towards the $p_{best}$ and $g_{best}$ positions at the same time.

The PSO algorithm starts with a set of randomly generated solutions and the swarm is updated using equations that control velocity and positioning of particles in every iteration. This process is repeated until no further improvement is obtained for the objective function value. When each particle has all the other particles as neighbours, this implies that the global best particle-position for all particles is identical as shown is Fig. 6.

![PSO Algorithm](image)

**Figure 6** PSO using $g_{best}$ neighbourhood topology showing fully connected neighbourhood relation (from Coelho and Mariani (2009))
A description on how to write the optimisation in Java version of \textit{jswarm-pso} is given as follows:

- Create your own \textbf{Fitness function}. It must derive from \texttt{jswarm\_pso.FitnessFunction} and custom \texttt{evaluate} method must be created.

\begin{verbatim}
import jswarm_pso.FitnessFunction;
public class MyFitnessFunction extends FitnessFunction {
    public double evaluate(double position[]) {
        return position[0] + position[1];
    }
}
\end{verbatim}

- Create your own particle derived from \texttt{jswarm\_pso.Particle}

\begin{verbatim}
import jswarm_pso.Particle;
public class MyParticle extends Particle {
    // Create a 2-dimentional particle
    public MyParticle() {
    }
}
\end{verbatim}

- Create a \textbf{Swarm} and \textit{evolve} for a few iterations (use your favourite \texttt{stop criteria})

\begin{verbatim}
// Create a swarm (using 'MyParticle' as sample particle
// and 'MyFitnessFunction' as finess function)
Swarm swarm = new Swarm(Swarm.DEFAULT_NUMBER_OF_PARTICLES
    , new MyParticle()
    , new MyFitnessFunction());
// Set position (and velocity) constraints.
// i.e.: where to look for solutions
swarm.setMaxPosition(1);
swarm.setMinPosition(0);
// Optimize a few times
for( int i = 0; i < 20; i++ ) swarm.evolve();
// Print en results
System.out.println(swarm.toStringStats())
\end{verbatim}

Note: default number and values as adopted in JPSOswarm.java is maintained for this research.

The following steps adopted from Coelho and Mariani (2009) and Fig. 7 gives the procedure for implementing the global version of PSO.

(i) Initialize population of particles with random positions and velocities in the \( n \) dimensional problem space using a uniform probability distribution function.

(ii) Evaluate the fitness value of each particle.
(iii) Compare each particle’s fitness with the particle’s $p_{best}$. If the current value is better than $p_{best}$, then set the $p_{best}$ value equal to the current value and the $p_{best}$ location equal to the current location in $n$-dimensional space.

(iv) Compare the fitness with the population’s overall previous best. If the current value is better than $g_{best}$, then reset $g_{best}$ to the current particle’s array index and value.

(v) Change the velocity and position of the particle accordingly.

Figure 7: Flowchart in PSO approach (from Coelho and Mariani (2009))

Figure 8 shows the flowchart for model structure and parameter optimisation in SALMO-PLUS. Lake specific model simulation was initialised combining input variables, parameters and model structure. All process model combination of the SALMO-OO library was tested for lake specific before an optimum model structure is chosen. This is followed by sensitivity analysis of selected or constant parameters from the rank of most sensitive parameter by means of multi objective parameter
optimisation. The values from this parameter optimisation were adopted together with the optimum model structure for simulation.

Figure 8 Conceptual framework of the SALMO-PLUS simulation system (from (Martin & Recknagel, 2010))

3.3.2 Running SALMO-PLUS

Operating SALMO-PLUS requires Java version 1.6.0_20. Fig.9 shows a snapshot from computer on the information about Java version. Inspecting the Java version on computer can be done by the following steps:
1. In Windows, open a DOS Command Prompt (also known as command window) and enter the following command:
   `java -version` and press enter,
2. The following output will appear where the Java version is listed:

   ![Command Prompt](image)

   **Figure 9** Snapshot from computer showing the Java version

   A manual for preparing input and measured data, uploading data and operating the SALMO-PLUS is given in Appendix A

   The output from SALMO-PLUS simulations for each study site data were compared to measured data for phytoplankton biomass, nutrient (phosphate and nitrate) concentrations and algal functional groups. In cases where no phytoplankton functional group data were available from study sites, phytoplankton succession theory for different trophic conditions was adopted from Reynolds (1984) to determine the extent of SALMO-PLUS simulation on algal functional groups.

### 3.4 Running SALMO-OO

The application of SALMO-OO for different lake categories requires numerous water quality parameters. Figure 10 shows a graphical user interface (GUI) in SALMO-OO to test different combinations of alternative growth and grazing models for a selected lake conditions. Drop down menus for selecting lake data sets, year of data, alternative growth and grazing functions and scenario selection options were included in the GUI.
User may choose any combinations of lake, year data and growth and grazing function and scenario selection to begin with. Another category which is known as ‘tropical stratified and mesotrophic’ has been created specifically for this study. The simulation results of state variable outputs are automatically shown after simulation has been completed within the SALMO-OO application. Measured data (if available) is shown in red dot within the respective predicted output graph for comparison purposes (see Fig. 11). The resulted RMSE and $r^2$ values for state variable outputs with measured data are shown on the same graphs as well.
3.4.1 Addition of Process Models from Law *et al.* (2009) to the Library of SALMO-OO

The search for new process models for the SALMO-OO library focused on phytoplankton models that were of the ODEs form. This is to ensure that the potential new process models were compatible with the model structure from present SALMO-OO. The Microsoft Office Excel was used to test potential new algal growth, algal grazing, zooplankton growth and zooplankton mortality models for addition into SALMO-OO library. The development, testing and running of ODE for potential additional models were also made using the same program. Test with Microsoft Office Excel on the new process model from Law *et al.* (2009) showed results of reasonable order of magnitude as expectation and within the realistic range for each function category. The criteria been tested were phytoplankton growth and grazing as well as zooplankton growth and mortality functions.

It was also crucial that the potential models chosen for consideration showed their capability to simulate phytoplankton and zooplankton seasonal dynamics with a reasonable range and order of magnitude. The fundamental part is to reproduce the functions in MS Excel formulations where inadequate documentation of the many potential models in the literature was always unfavourable to this author during the initial stage of experiment. Three phytoplankton functional groups (i.e. diatoms, green algae and blue-green algae) were tested for each potential model.

Figure 11 Visual output of simulation results within the SALMO-OO application (from: http://ecolinfo.ees.adelaide.edu.au:8080/SALMO-OO/index.html)
Investigations on the literature for potential process models were conducted where different combinations of growth equations and growth-limiting functions regarding nutrients, light and water temperature were deliberated. Potential models were also tested for comprehensive different combinations of growth functions within the SALMO-OO simulation library.

### 3.4.2 Evaluation Criteria for Potential Models

Numerous algal population models are basically similar in structure but they differ in the mathematical representation of the key processes such as growth and grazing. Investigation on the scientific literature for models describing phytoplankton dynamics in water bodies were made following the criteria set by Cetin (2007). Several key aspects are simplified in the following:

1. The phytoplankton models that were of the form of ordinary differential equations
2. The phytoplankton models chosen should be able to simulate phytoplankton seasonal dynamics reasonably well
3. The models chosen are adequately documented to allow proper simulation in the Microsoft Office Excel
4. The ability of the model to properly simulate phytoplankton functional groups dynamic

#### Testing for new model library

All the 128 parameters in SALMO-OO model are kept constant during the simulation for the new proposed process model. The new process model was also analysed for new parameters which are not available in SALMO-OO model. In cases where new parameters were observed in the proposed new process model, investigations on literature for parameter values were carried out to find the suitable parameter ranges for those parameters. However, further examination shows that even though these new parameter were originally having different names in their scientific publication, there are similar by definitions and only differ by their specific values. No new parameters which are not found in SALMO-OO were observed in the new process model from Law et al. (2009). Therefore, only notation/naming standard amendment was done in order to standardise these parameters according to the SALMO-OO convention. Parameters from Law et al. (2009) with same meaning but hold different values when compared to SALMO-OO parameters are listed in Table 27 in Chapter 4 with their original values retained during this experiment.

The output from SALMO-OO simulations using new process model from Law et al. (2009) for each study site data were compared to measured data for phytoplankton biomass, nutrient (phosphate and nitrate) concentrations and algal functional groups. In cases where no phytoplankton functional group data were available from study sites, phytoplankton succession theory for different trophic conditions was adopted from Reynolds (1984) to determine the extent of SALMO-OO simulation on algae functional groups.
A phytoplankton growth and grazing and zooplankton growth and mortality model by Law et al. (2009) was selected for addition into the current SALMO-OO simulation library. This new model was a result of modelling analysis of the plankton variability in Lake Washington over a 35-year period (1964-1998) to examine the relative importance of the ecological mechanism underlying plankton seasonal variability in the lake. It considers the interplay among four state variables (phosphate, phytoplankton, zooplankton and detritus) and was founded upon an intermediate complexity plankton model that is used to reproduce the limiting nutrient (phosphate)-phytoplankton-zooplankton-detritus (particulate phosphorus) dynamic in the lake. The spatial segmentation of the model consists of three compartments representing the epilimnion, thermocline (mesolimnion) and hypolimnion of the lake. The simulation of this model was solved numerically using the fourth order Runge-Kutta method with a time step of 1 day. The main objective of this selected model was to unravel the temporal evolution of the ecological mechanisms (direct and indirect pathways) that drive the planktonic patterns and to elucidate the structural changes in lake functioning induced from nutrient loading and climate variability in Lake Washington. Bayesian calibration scheme was used to parameterise this model. After a period of increasing amount of secondary sewage inflow from 1941 to 1963, this point nutrient source was eventually eliminated in 1968 and Lake Washington is now regarded as successfully recovered from eutrophication.

This new library extension was adopted due to its capability to unravel the temporal evolution of the ecological mechanisms (direct and indirect pathways) that drive the planktonic patterns and to elucidate the structural changes in lake functioning induced from nutrient loading and climate variability (Law et al., 2009). It is also hypothesize that this new sub-model would enhance the SALMO-OO model’s simulation capabilities particularly regarding zooplankton mortality in relation to planktivorous fish’s effect especially for tropical lakes. The zooplankton losses due to consumption by higher predators in this new process model are modelled using a sigmoid closure term which represents a ‘switch-able’ type of predator behaviour controlled by a prey threshold concentration.

Another reason for choosing this new sub-model for addition into SALMO-OO library was because it has been proven to be successfully applied to Lake Washington. The Lake Washington is one example of a successful lake restoration effort and now categorised as mesotrophic lake (Smith as cited in Horn, 2003). Located in the State of Washington, U. S. A. adjacent to the city of Seattle, Lake Washington presently has an area of 88km² and a mean depth of 33m. The residence time of water in the lake (i.e the ratio of the lake volume to the inflow of freshwater) is about 3.4 years. The lake is monomictic and overturning generally occurs from the end of November until April of May (Laws, 2000).

The object-oriented programming language Java was used to create an additional alternative model inside the SALMO-OO library. The new equations for the library were translated into the same basic code structure in SALMO-OO for phytoplankton growth, phytoplankton grazing, zooplankton growth and zooplankton mortality respectively. The equations consisting mainly inputs for the biomass of
each algae functional group \((A_i)\) phosphate (P), nitrate (N) and parameter index as well as measured input such as reference temperature, water temperature and solar radiation. A loop function was created to allow the calculation of growth and grazing for each phytoplankton functional group with links to the measured input data.

### 3.5.1 Phytoplankton Process Model in Law et al. (2009)

Phytoplankton growth in Law et al. (2009) is calculated as results of photosynthesis minus respiration per unit biomass.

\[
AGRO_{i,T,E,H} = (PHO_{i,T,E,H} - RA_{i,T,E,H}) \times A_{i,T,E,H} \tag{3}
\]

where:
- \(AGRO_{i,T,E,H}\) = phytoplankton growth process in total mixed layer, epilimnion; hypolimnion
- \(PHO_{i,T,E,H}\) = phytoplankton photosynthesis rate
- \(RA_{i,T,E,H}\) = phytoplankton respiration rate
- \(A_{i,T,E,H}\) = phytoplankton biomass

Generally, photosynthesis is limited by nutrients (phosphate and nitrate), light and water temperature. However, Law et al. (2009) is modelling a system that is phosphate limited throughout the annual cycle; so nitrate was not considered in the formulation.

\[
PHO_{i,T,E,H} = PHOMAX_i \times PHOL_i \times PHOT_i \times PHOP_i \times A_i \tag{4}
\]

where
- \(PHOT_i\) = growth modified by temperature function
- \(PHOL_i\) = growth limited by light function
- \(PHOP_i\) = growth limited by dissolved inorganic phosphorus
- \(PHOMAX_i\) = maximum photosynthesis rate \((\text{d}^{-1})\)
- \(A_i\) = phytoplankton biomass

The dependence of the planktonic process on temperature is modelled by \(PHOT_i\) function that resembles a Gaussian-like probability curve. The equations used to represent the effects of temperature on phytoplankton metabolism in Law et al. (2009) were different from Arhonditsis and Brett (2005) (another alternative process model available in the SALMO-OO library) where the former used the Gaussian approach in replacing the exponential approach adopted by the latter. Gaussian curve is similar to normal curve or bell curve. The formula used to calculate temperature limitation is based on the function of reference temperature and water temperature and is given as follows:
$\text{PHOT}_i = e^{kt(T_{\text{Temp}} - T_{\text{Temp ref}})^2}$  \hspace{1cm} (5)

where:
- $\text{PHOT}_i$ = growth modified by temperature function
- $kt$ = effect of temperature on phytoplankton processes
- $T_{\text{Temp ref}}$ = reference temperature
- $T_{\text{Temp}}$ = water temperature

The phosphate equation in SALMO-OO model library considers the phytoplankton uptake, the gains due to zooplankton excretion/predation, the bacteria mediated mineralization of detritus and the net diffusive fluxes between adjacent compartments.

The function used to calculate the nutrient limitation in Law et al. (2009) is based on a parsimonious model that only considers the basic processes underlying the Lake Washington dynamics. Law et al. (2009) is modelling a system that is phosphate limited throughout the annual cycle; so nitrate was not essential in this particular model. Phosphate limitation function from Law et al. (2009) is modelled using the Michaelis-Menten kinetics which is similar to other process model in the library of SALMO-OO. However, Law et al. (2009) emphasizes solely on phosphate as the limiting nutrient, due to the nature of nutrient loading in their study lake which is Lake Washington, U.S.A. Phosphate limitation on phytoplankton growth for this selected model is given as follows;

$$\text{PHOP}_i = \frac{\text{PO}_4}{K_P + \text{PO}_4} \hspace{1cm} (6)$$

where:
- $\text{PHOP}_i$ = growth limited by dissolved inorganic phosphate
- $\text{PO}_4$ = dissolved inorganic phosphorus (mg/m$^3$)
- $K_P$ = half-saturation constant for PO$_4$ uptake by algae

The light dependence of phytoplankton growth is modelled using Steele’s equation with Beer’s law to scale photosynthetically active radiation to depth. This slightly differs from Michaelis-Menten kinetics which was used by other model in the SALMO-OO library. Both background and chlorophyll-a attenuation are considered in the extinction coefficient of this equation. The function used to calculate the light limitation is given as follows:

$$\text{PHOL}_i = 2.718FD \cdot \left( e^{-\alpha_1} - e^{-\alpha_0} / kei*Hi \right) \hspace{1cm} (7)$$

where:
- $\text{PHOL}_i$ = growth modified by light limitation function
- 2.718 = Euler's number
- $FD$ = is an intermediate function with formula: 0.15sin(2π(t/365-0.22))+0.51
- $\alpha_0$ = an intermediate variable with formula: (Ia/Ib)*$e^{-kei*Hi}$
- $\alpha_1$ = an intermediate variable with formula: (Ia/Ib)*$e^{-kei*Hi}$
\( \alpha^1 \) = an intermediate variable with formula: \( (Ia/Is)*e^{-keiH2} \)

\( kei \) = an intermediate function with formula: \( kei = kb + kcAi / 50 \)

\( Hi \) = depth of the \( i \) spatial compartment

\( H_1 \) and \( H_2 \) = distance of upper and lower bound of \( i \) compartment from the surface of the water body.

\( KL \) = light extinction coefficient due to Chl a (m\(^2\)mg\(^{-1}\))

\( KB \) = background light extinction coefficient (m\(^{-1}\))

\( Ai \) = phytoplankton biomass

\( t \) = Julian day

\( Is \) = half saturation light intensity

\( Ia \) = daily solar radiation at the surface of the water body

\( \pi = 2.314 \)

Note:

\( Ia \) = is given formula by \( Ia = 220*\sin(2\pi(t/365-0.23))+290 \)

Phytoplankton respiration in Law et al. (2009) is given as follows:

\[
RA_{i_{T,E,H}} = \text{RATOPT}_i \times \text{PHOT}_i \times Ai
\]  \hspace{1cm} (8)

where:

\( RA_{i_{T,E,H}} \) = phytoplankton respiration

\( \text{RATOPT}_i \) = phytoplankton respiration rate, day\(^{-1}\)

\( \text{PHOT}_i \) = growth modified by temperature function – similar as given in the temperature equation calculation section

\( Ai \) = phytoplankton biomass

Respiration is limited by temperature and maximum respiration rate and is modelled as an exponential function where \( kt \) is the effect of temperature on phytoplankton processes. The respiration function described by Hongping and Jianyi (2002) and Arhonditsis and Brett (2005) (both are another alternative in the library of SALMO-OO) also use exponential rate albeit with different parameter values.

The algal grazing by zooplankton is considered as key ecological process determining loss of algal biomass and is generally similarly formulated for each of the SALMO-OO library models. Algal grazing from Law et al. (2009) is given as follows:

\[
G_{\text{max}} \times (Ai \times P/Cphyto)^2 \times PHTz \times Z
\]
\[
KG^2 + (Ai \times P/Cphyto)^2 + PFD \times D_i^2
\]  \hspace{1cm} (9)

where:

\( G_{\text{max}} \) = maximum zooplankton grazing rate (day\(^{-1}\))

\( KG \) = zooplankton grazing half saturation coefficient, mgPm\(^{-3}\)

\( PFD \) = relative zooplankton preference for detritus compared to phytoplankton

\( P/Cphyto \) = phosphorus to carbon ratio for phytoplankton; 0.015 mg P (mgC\(^{-1}\))

\( D_i \) = detritus biomass
$Z_i = \text{zooplankton biomass}$

and

$\text{PHT}_z = \text{intermediate function with formula: } e^{KT_z(temp_i - temp_{ref})^2}$

where $KT_z = \text{effect of temperature on zooplankton processes (C}^{-2})$

$temp_{ref} = \text{reference temperature}$

$temp_i = \text{water temperature}$

### 3.5.2 Zooplankton Process Model in Law et al. (2009)

In Law et al. (2009), zooplankton feeding upon food sources is formulated using the Holling Type III function that is the ‘S’ shape graph or sigmoid. Zooplankton growth process model is given as follows:

$$Feff_i \times G_{max} \times ((A_i \times P/C_{phyto})^2 + \omega \times D_i^2) \times \text{PHT}_z \times Z$$

(10)

where:

$Feff_i = \text{zooplankton assimilation efficiency (day}^{-1})$

$G_{max} = \text{maximum zooplankton grazing rate (day}^{-1})$

$P/C_{phyto} = \text{phosphorus to carbon ratio for phytoplankton; } 0.015 \text{ mg P (mgC)}^{-1}$

$KG = \text{zooplankton grazing half saturation coefficient, mgPm}^{-3}$

$\text{PFD} = \text{relative zooplankton preference for detritus compared to phytoplankton}$

$Z_i = \text{zooplankton biomass}$

$D_i = \text{detritus biomass}$

$A_i = \text{phytoplankton biomass}$

and

$\text{PHT}_z = \text{intermediate function with formula: } e^{KT_z(temp_i - temp_{ref})^2}$

where $KT_z = \text{effect of temperature on zooplankton processes (C}^{-2})$

$temp_{ref} = \text{reference temperature}$

$temp_i = \text{water temperature}$

Law et al. (2009) formulation for zooplankton mortality process model includes a temperature limitation functions that regulates the zooplankton mortality and is given as follows:

$$d \times (Z_i^3 / \text{pred}^2 + Z_i^2) \times \text{PHTZ}$$

(11)

where:

$d = \text{zooplankton mortality rate (day}^{-1})$

$Z_i = \text{zooplankton biomass}$

$\text{pred} = \text{half-saturation constant for predation (mgCm}^3)$

and

$\text{PHTZ} = \text{intermediate function with formula: } e^{KT_Z(temp_i - temp_{ref})^2}$

where $KT_Z = \text{effect of temperature on zooplankton processes (C}^{-2})$
The dependence of the zooplankton process on temperature is modelled by PHTZ function that resembles a Gaussian-like probability curve. The formula for temperature function for zooplankton mortality is given as follows:

\[ \text{PHTZ} = e^{K_TZ (\text{temp}_i - \text{temp}_{\text{ref}})^2} \]  \hspace{1cm} (12)

where:  
\( K_TZ \) = effect of temperature on zooplankton processes, \( ^\circ \text{C}^{-2} \)  
\( \text{temp}_{\text{ref}} \) = reference temperature  
\( \text{temp}_i \) = water temperature

### 3.5.3 Differences Between Process Model in Law et al. (2009) and SALMO-OO Simulation Library

**Differences in the photosynthesis formulation**

All photosynthesis formulation in SALMO-OO is limited by nutrients, light and temperature. However, different approaches were used by different authors in SALMO-OO model as listed in the following:

1. **The multiplicative function** (example: Hongping and Jianyi, 2002)  
   This function is the maximum photosynthesis rate multiplied by sub-optimal conditions, which generally includes limitations due to light intensity, water temperature and different nutrient concentrations.

   \[ \text{PHO}_{iTEH} = \text{PHOMAX}_i \times \text{PHOL}_i \times \text{PHOT}_i \times \text{PHOP}_i \times \text{PHON}_i \]  \hspace{1cm} (13)

2. **The minimum function** (example: Arhonditsis and Brett, 2005)  
   (also known as Liebig’s Law of the Minimum construct)  
   The minimum function is used to a lesser degree than the multiplicative function. In many cases, the minimum function is used to determine the limiting nutrient only rather than single limiting resource except for some case with light intensity.

   \[ \text{PHO}_{iTEH} = \text{PHOMAX}_i \times \text{PHOL}_i \times \text{PHOT}_i \times \min (\text{PHOP}_i, \text{PHON}_i) \times \min (\text{PHOP}_i, \text{PHON}_i) \]  \hspace{1cm} (14)

The nutrient limitation calculation by Benndorf and Recknagel (1982) is based on Michaelis-Menten kinetics with a threshold function that determines the limiting nutrient based on the N:P ratio. Function to determine which nutrient is limiting phytoplankton growth is given by the following:

\[ \text{PHON}_i = \text{PHOP}_i; \text{ if } N/P \geq 0.0072 \]  
\[ \text{PHON}_i = \text{PHON}_i; \text{ if } N/P \leq 0.0072 \]  \hspace{1cm} (15)
3. The mean resistance construct (example: Park et al., 1974)

\[
\text{PHO}_{i,T,E,H} = \text{PHOMAX}_i \times U_t \times \text{PHOT}_i
\]  

(16)

Where: \( U_t = \frac{N}{\sum [1/f(U_t)]} \)

Therefore: \( U_t = \frac{3}{((f(\text{PHOL}_i) + f(\text{PHOP}_i) + f(\text{PHON}_i)))} \)

4. The Law et al. (2009) is modelling a system that is phosphate limited throughout the annual cycle. Therefore, nitrate was not considered in the formulation that is given as follows;

\[
\text{PHOi}_{t,T,E,H} = \text{PHOMAXi} \times \text{PHOL}_i \times \text{PHOT}_i \times \text{PHOP}_i
\]

(17)

**Nutrient function**

As mentioned earlier, approaches adopted by equations in the SALMO-OO simulation library regarding phosphorus and nitrogen limitation for phytoplankton biomass growth differed from Law et al. (2009). The original SALMO-OO considers nitrate and phosphate contribution with a threshold function that determines the limiting nutrient based on the N:P ratio. Thus the inclusion of Law et al. (2009) process model into the library of SALMO-OO was expected to offer another potential scenario of lakes where only phosphorus is the limiting nutrient. While they differ in terms of which nutrient is limiting, the nutrient limitation calculation in both SALMO-OO library and the new process model from Law et al. (2009) are based on Michaelis-Menten kinetics.

Arhonditsis and Brett (2005) uses minimum function based on Liebig’s Law of the Minimum to determine which nutrient is the most limiting for phytoplankton growth. However, within their lake model, silicon and carbon are also considered as nutrients affecting phytoplankton metabolism and thus are also included in the minimum function, whereas within SALMO-OO simulation library, only phosphate and nitrogen state variables are considered (Cetin, 2007). It should be noted that nutrient limitation formulation by Hongping and Jianyi (2002) also consider phosphate only for nutrient equation. However, the complete equation structure for phytoplankton growth in Hongping and Jianyi (2002) model was found to be different from Law et al. (2009). The Law et al. (2009) adopted a parsimonious model that only considers the basic processes underlying the Lake Washington dynamics as the function used to calculate the nutrient limitation. This function caters for a lake ecosystem that is phosphate limited throughout the annual cycle and nitrate was not considered essential. Equation for phosphate as the most limiting nutrient for phytoplankton growth in Law et al. (2009) is given as follows;
\[ \text{PHOP}_i = \frac{\text{PO4}_i}{(\text{KP}_i + \text{PO4}_i)} \]  

(18)

where:
PHOP\(_i\) = growth limited by dissolved inorganic phosphate
PO4\(_i\) = dissolved inorganic phosphorus (mg/m\(^3\))
KP\(_i\) = half-saturation constant for PO4 uptake by algae

**Temperature function**

The difference between the present process models in the SALMO-OO library and Law *et al.* (2009) model is that, in SALMO-OO library, the dependence of algal growth on temperature is included as an optimum function, where growth increases linearly until an optimum temperature is reached and then plateaus to a more constant rate. Whereas, the dependence of the planktonic process in Law *et al.* (2009) on temperature is modelled by temperature function (PHOT\(_i\)) that resembles a Gaussian-like probability curve. The equations used to represent the effects of temperature on phytoplankton metabolism in Law *et al.* (2009) were different from Arhonditsis and Brett (2005) where the former used the Gaussian approach in replacing the exponential approach adopted by the latter. Gaussian curve is similar to normal curve or bell curve. The formula used to calculate temperature limitation is based on the function of reference temperature and water temperature and is given as follows:

\[ \text{PHOT}_i = e^{-kt(T_{\text{temp}} - T_{\text{ref}})^2} \]  

(19)

where:
PHOT\(_i\) = growth modified by temperature function
kt = effect of temperature on phytoplankton processes
Treft = reference temperature
Tempi = water temperature

### 3.6 HEA (Hybrid Evolutionary Algorithms)

This study is meant to utilize the power of HEA to investigate and discover the dynamics of algal population in tropical lake ecosystems. The HEA was designed to uncover predictive rules in ecological time-series data (Cao *et al*., 2006a) by combining genetic programming to generate and optimise the structure of rules and genetic algorithm to optimise parameters of rules (Recknagel *et al*., 2006). Therefore, HEA was used to search for suitable representation of a problem solution by means of genetic operators and the principle of ‘survival of the fittest’. During each run, the program will assess each output by means of ‘fitness cases’. Only the fitter results are selected for recombination by using the genetic operator such as crossover and mutation to create the next generation. These resulting rules will be consequently repeated for generations using genetic operators until the criteria for termination of the program have been met and the fittest rule has been determined. This is followed by parameter optimization of the rule set by means of genetic algorithm. According to Whigham *et al*., (2006), the optimisation of constants within process models of
chlorophyll-a concentration for freshwater ecosystems, particularly in lake environments has significantly improves the quality of the results on the unseen data. Figure 12 illustrate the flowchart of the hybrid evolutionary algorithm.

![Flowchart of the hybrid evolutionary algorithm](image)

**Figure 12** Flowchart of the hybrid evolutionary algorithm (from Cao et al., 2006a).

Time series data from lake ecosystems parameters were used to forecast algal blooms as well as seasonal succession of algal species in the distant future such as 7-days ahead. Prediction of both the timing and magnitude of an algal species were forecasted by a set of rules. The complexity of HEA rule sets can be controlled by the maximum tree depth and the rule set size. Rules discovered by HEA have the IF-THEN-ELSE structure and allow imbedding complex function syntheses from various predefined arithmetic operator (Cao, et al., 2006b). The model output includes a list of the variables used within each run, the frequency of each input parameter being used and the RMSE (root mean square error) as well as r-square values. Table 9 summarizes the daily input data used to run the HEA model for all lakes in this study.
Table 9 Daily input data used for HEA in this study

<table>
<thead>
<tr>
<th>No</th>
<th>Parameters</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Water temperature</td>
<td>°C</td>
</tr>
<tr>
<td>2</td>
<td>pH</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Secchi depth</td>
<td>m</td>
</tr>
<tr>
<td>4</td>
<td>Dissolved Oxygen</td>
<td>mg/l</td>
</tr>
<tr>
<td>5</td>
<td>NH₄</td>
<td>mg/l</td>
</tr>
<tr>
<td>6</td>
<td>NO₃-N</td>
<td>mg/l</td>
</tr>
<tr>
<td>7</td>
<td>PO₄-P</td>
<td>mg/l</td>
</tr>
<tr>
<td>8</td>
<td>Solar radiation</td>
<td>J/cm².d</td>
</tr>
<tr>
<td>9</td>
<td>Chl-a</td>
<td>µg/l</td>
</tr>
<tr>
<td>10</td>
<td>Biovolume</td>
<td>µm³</td>
</tr>
<tr>
<td>11</td>
<td>Biochemical oxygen demand</td>
<td>mg/l</td>
</tr>
<tr>
<td>12</td>
<td>Chemical oxygen demand</td>
<td>mg/l</td>
</tr>
<tr>
<td>13</td>
<td>Turbidity</td>
<td>NTU</td>
</tr>
<tr>
<td>14</td>
<td>Conductivity</td>
<td>µS/cm</td>
</tr>
<tr>
<td>15</td>
<td>Rainfall</td>
<td>mm</td>
</tr>
</tbody>
</table>

The software HEA has been applied by means of the bootstrap training mode for 100 runs randomly picking 80% of the data for training and 20% of the data for testing. In bootstrap, one computes an object and creates an artificial list by randomly drawing elements from that list. In due course, some elements will be picked more than once. A new object is computed and the exercises are repeated numerous times and seek the best model by looking at the distribution of these objects. The absent of distinct seasonality on all the three study sites allows the use of bootstrap method to predict the dynamic of algal biovolume and chlorophyll-a.

3.6.1 Data Requirements for the HEA

The principle framework of the rule discovery of HEA is shown in Figure 13. HEA utilises programming to evolve the structure of the rule set and optimised the random parameters on the rule set by using a general genetic algorithm. The attraction with rules is that they can be easily be interpreted by a user and can be treated independently of the system in which they were created (Whigham & Fogel, 2003). HEA allows discovering rule sets which predict unseen data and represent causal relationships between physical and chemical variables and algal population dynamics (Cao, et al., 2006b).

The hybrid EA was applied for short-term forecasting of chlorophyll-a in Lake Putrajaya (chlorophyll-a), Lake Kenyir (biovolume) and Lake Penang (biovolume and chlorophyll-a) respectively. Criteria used to determine the best performing predictive models by HEA in this research are listed as follows;

- The root mean square error (RMSE)
- The square of correlation coefficient ($r^2$) of a linear regression between predicted and measured data
- Visual comparison between predicted and measured data
3.6.2 Data Treatment for the HEA

The time-series ecological data from: Lake Putrajaya (biweekly, six-year data), Lake Kenyir (monthly, one-year data) and Lake Penang (monthly, one-year data) were used for this research. These time series data were interpolated to produce daily values for each data using Microsoft Excel software for modelling by HEA using bootstrap method. Seven days’ time-lag between the input and output time-series data was imposed to the data for forecasting purposes. Pre-processing of data to clean up and eradicate missing values was done following the treatment as explained in section 3.1.1.

Another time series data set from Lake Kenyir and Lake Penang were merged together to assess HEA’s capability in generating generic rules for merged data set from two different tropical lakes with similar trophic status and mixing conditions.

Data training were randomly assigned for 80 per cent of the data while the remaining 20 per cent were used for testing the rule sets. All HEA experiments were performed on a supercomputer (IBM eServer 1350 Linux) with an appropriate peak speed by means of the C++ programming language. The parameters setting for both structure optimisation (genetic programming) and parameter optimisation (genetic algorithm) is as tabulated in Table 10. N represents the population size of the rule model describing how many rule models will be produced in the initial population and the following iteration. Pop size is the parameter population (vector) and max gen is the stopping criteria utilised to end the optimisation process.
Table 10 Parameter settings of HEA for rule model discovery (Cao et al., 2006b)

<table>
<thead>
<tr>
<th>Parameter setting</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure optimisation (GP)</td>
<td>N=200</td>
</tr>
<tr>
<td>Parameter optimisation (GA)</td>
<td>Pop size =100</td>
</tr>
<tr>
<td></td>
<td>Max gen=200</td>
</tr>
</tbody>
</table>

Biovolume calculation

One of the objectives for this research was to use HEA to develop generic predictive rules for algal growth in tropical lakes with different morphometry and trophic states. This was conducted by merging time series data from lakes Penang and Kenyir and setting a common output for both dataset. Algal biovolume was chosen as the output. Therefore, the conversion of Chl-a data to biovolume was performed on Chl-a from Lake Penang. Phytoplankton biomass was estimated by biovolume, multiplying the density of each species by its respective cell volume. This was done by the following steps:

1. Total cell number for the top four most significant algal species was obtained from raw data of lakes Penang and Kenyir
2. Individual cell biovolume for respective algal species was obtained from literature
3. The cell number was multiplied to the cell biovolume from literature in order to find specific algal biovolume value.

Data of individual cell biovolume for alga was obtained from Makarewicz (1993), Barros et al., (2006), Rodrigo et al., (2009) and HELCOM PEG Biovolume Reporting website (http://www.ices.dk/env/repfor/index.asp).

3.6.3 Sensitivity Analysis

Sensitivity analysis was conducted based on rule based model to estimate output sensitivity to input variables in order to evaluate the applicability of the output models. In order to analyse sensitivity for variables in developed models, one variable varied while the others were fixed in average value. This step is repeated for each of the variable present in the rule. Sensitivities are distributed into two parts due to partition of the “IF” condition in the rule discovery. The following are steps taken during the sensitivity analyses for the best performing predictive rule sets:

- all data will be divided according to the THEN and ELSE branches of the IF criteria
- the values of mean, standard deviation (STDEV), mean-STDEV and mean+STDEV will be calculated for each variable considered in the equations of the THEN or ELSE-branch
- the sensitivities of algal concentrations to specific variables will be calculated by fixing other variables to the mean values in the given THEN and ELSE equations.
3.6.4. Generic Predictive Rules for Particular Lake Category

This research also aims to develop rule-based agents that are predictive for algal biovolume by accurately forecast the timing and magnitude of algal biovolume level and generic for tropical lake ecosystem category. This was done by developing the predictive rules-based agents for algal populations as follows;

1. rule discovery by means of HEA and followed by,
2. rule generalisation by means of merged time series data of lakes belonging to the same category

The HEA was developed for potential use in forecasting the algal biovolumes level in both Lake Penang and Lake Kenyir environments. Biovolumes was used as an output to standardise the results from this experiment. Therefore, the framework was developed to use only input variables that are available in both lakes. The maximum number of input parameters and greatest duration of data possible were used for this experiment. Interpolated daily values were used for these modelling experiments while the training and testing of the rule-based models developed throughout this experiment were 80% and 20% respectively. In order to design the model for 7-days ahead forecasting, a time lag of 7 days was imposed between the input and output data.

3.7. DYRESM Model

DYRESM was used in this study to investigate the dynamics of temperature profile in Lake Putrajaya ecosystems. From hydrodynamic point of view, provided that the lake is neither extremely long and narrow, nor extremely broad and shallow, a one-dimensional hydrodynamic approach is satisfactory. A one-year sampling was carried out from April 2008 at station PLg2 which is the deepest point (see Fig. 2) in Lake Putrajaya during which all input variables required by DYRESM was measured. Table 11 summarizes the input data to run the DYRESM.

For the purpose of using DYRESM, Lake Putrajaya is assumed to be one-dimensional. Daily meteorological input data (incident short wave radiation, long wave radiation, air temperature, wind speed, vapour pressure and rainfall), daily inflow and outflow volumes, salinities and temperatures were used for DYRESM model. This model was used to simulate temperature stratification and from the temperature stratification, dissolved oxygen and phosphorus dynamics were developed.
Table 11 List of input data required for running the DYRESM model

<table>
<thead>
<tr>
<th>No</th>
<th>Parameters</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Configuration file</td>
<td>Set up data to run DYRESM</td>
</tr>
<tr>
<td>2</td>
<td>Morphometry</td>
<td>Morphometric characteristic of the Lake Putrajaya</td>
</tr>
<tr>
<td>3</td>
<td>Initial profile</td>
<td>Initial vertical profile of water temperature and salinity</td>
</tr>
<tr>
<td>4</td>
<td>Meteorology</td>
<td>Meteorological data</td>
</tr>
<tr>
<td>5</td>
<td>Inflow</td>
<td>Daily average inflow data (volume, temperature and salinity)</td>
</tr>
<tr>
<td>6</td>
<td>Outflow</td>
<td>Daily values of withdrawal data</td>
</tr>
<tr>
<td>7</td>
<td>Parameters</td>
<td>Parameters of the simulation</td>
</tr>
</tbody>
</table>

3.8. Fitness evaluation

Results of best rule-based model in every single run are tested each for its validity and generality by calculating the predicted values against the measured data. The best performing rule-based model was chosen based on RMSE, $r^2$, rule of appropriateness, simplicity and visual assessment of measured versus predicted data.

The RMSE is define as the testing error for validation of the result of different rules and is given by the following:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2}$$  \hspace{5cm} (20)

where $m$ is the number of testing data points, $y_i$ and $\hat{y}_i$ are the $i$th observed value and the $i$th predicted value of Chl-a or biovolumes respectively.

Visual assessment by experts, r-squared and RMSE strategies are used for examining result from SALMO-OO, SALMO-PLUS and HEA in this research.
CHAPTER 4

4.0 RESULTS

This chapter is divided into two sections: the first section discusses results from applications of the data driven HEA model to tropical lakes (Lake Putrajaya, Lake Kenyir and Lake Penang). Results from the development of forecasting models for algal biovolume from merged limnological time series data from Lake Penang and Lake Kenyir by means of HEA are also explained. The second section presents results of model structure and parameter optimisation based on the process based model SALMO-PLUS utilising particle swarm optimisation for multi objective parameter optimisation.

4.1 Experimental Results from HEA and the Generic Rules for Tropical Lakes

The software HEA was applied to develop models for the three tropical lakes Lake Putrajaya, Lake Penang and Lake Kenyir (for details of the lakes see chapter 3) that allow 7-days-ahead forecasts of algal community dynamics reflected by chlorophyll-a and algal biovolume data within these lakes. The three lakes matched with following lake categories: tropical-polymictic mesotrophic (Lake Putrajaya) and tropical-stratified mesotrophic (Lake Penang and Lake Kenyir). Tropical-polymictic mesotrophic lakes are typically not deeper than 8m and therefore predominantly mixed. The average depth of Lake Putrajaya is 6.6m. Tropical-stratified mesotrophic lakes are deep lakes that are stably stratified. The average depth of Lake Penang is 17.9m and of Lake Kenyir is 37m respectively.

Three specific models have been developed for all three lakes and one generic model has been developed for tropical-stratified mesotrophic lakes based on merged data of the lakes Kenyir and Penang.

The software HEA has been applied by means of the bootstrap training mode for 100 runs where 20% of the data were randomly picked for testing and 80% of the data were used for training.

The bootstrap training mode was chosen instead of the split-sample training mode due to its capacity to more comprehensively explore the information content of small data sets as typical for the tropical lakes studied by this research, and to apply cross-validation. The use of one year datasets for the lakes Kenyir and Penang was justified because of the fact that tropical-stratified lakes show no distinctive seasonality in contrast to temperate-dimictic and Mediterranean-monomictic lakes.

The bootstrap training mode was also chosen because according to Adèr et al. (2008) resampling techniques such as bootstrapping is a very powerful technique that in most cases provides an impression of the distributional properties of the data set relative to the test statistic of choice and aim at enhancing estimation quality of the parameter estimates. As such, the bootstrap training mode chosen for this study is a
way to overcome the relatively limited number of available data and create more data in order to better assess the statistical properties of the data.

The following criteria were used to assess the models’ validity and determine the best performing predictive models by HEA:

- The root mean square error (RMSE) as a quantitative measure of fitting between the measured data and the model outputs for each variable
- The square of correlation coefficient ($r^2$) of a linear regression between predicted and measured data
- Visual comparison between predicted and measured data
- Plausibility test of input sensitivity analyses.

Sensitivity analysis is a procedure to study the behaviour of a model when global parameters are systematically varied, with an aim to check whether operations on the data influence the results (Adèr et al. 2008). In this study the sensitivity analysis was applied to the relationships between input variables and the forecasted outputs of the models that also revealed the input selection frequency over 100 bootstrap training runs.

All models were represented as “IF-THEN-ELSE” rule sets. Certain rules obtained by HEA may resulted in distinct value or number as a whole or part of the output component. This value may also represent a part or whole of the rule of HEA model that best performed with related data from study site used in the experiment. In an example given below, value of 30.962 is part of the rules obtained by HEA during the experiment. This number is then tested against the formula “exp(Secchi depth)” by replacing the notation with appropriate Secchi depth values on respective date in order to differentiate which Secchi depth measurement comes under THEN and/or ELSE-branch.

An example of rule obtained in one of the experiment:

```
"IF exp(Secchi depth) < 30.962
THEN Biovolume = ((pH*SD)*((pH*SD)*88.238))
ELSE Biovolume = (((147.948/SD)/PO_4-P) + exp(pH))*14.098")
```
4.1.1 Predictive Chl-a Model for Lake Putrajaya

Experiment 1

Six years of daily interpolated water quality data were used to develop a 7-days-ahead forecasting model for Chl-a of Lake Putrajaya. The one hundred runs of bootstrap training resulted in 100 models which were automatically ranked by HEA according to their RMSE and $r^2$.

Figure 14 illustrates the selection frequencies of the water quality variables considered as inputs for the forecasting of Chl-a in Lake Putrajaya during the 100 runs of bootstrap training. It clearly shows that pH, Secchi depth (SD) and dissolved oxygen (DO) appear to be key input variables for forecasting Chl-a whereas solar radiation (SR) seems to be the least important input variable.

Figure 14 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>NO$<em>3$-N$</em>{inflow}$</td>
<td>Nitrate inflow</td>
</tr>
<tr>
<td>pH</td>
<td>pH</td>
<td>PO$<em>4$-P$</em>{inflow}$</td>
<td>Phosphate inflow</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>Turb</td>
<td>Turbidity</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
<td>Cond</td>
<td>Conductivity</td>
</tr>
<tr>
<td>NO$_3$-N</td>
<td>Nitrate</td>
<td>SR</td>
<td>Solar radiation</td>
</tr>
<tr>
<td>PO$_4$-P</td>
<td>Phosphate</td>
<td>Rainfall</td>
<td>Rainfall</td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the following model has been identified as the best performing model:

IF pH is equal or less than 7.9,

THEN the chlorophyll a concentration is calculated by the equation;

\[ \text{Chl-a} = \text{pH-ln}((19.01-(\text{PO}_4\text{-P}\times138.424))) \]

Otherwise, by the ELSE equation:
Chl-a = pH+((exp(SD)-pH)-NO3-N_{inflow})

**Figure 15** Structure of the Chl-a model for Lake Putrajaya a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE-branch (note: values in the legend box are the range of input data)

The model achieved a root mean square error (RMSE) of 2.35 and an $r^2$ of 0.33

The IF-condition of the model stipulates the threshold value pH=7.9 for distinguishing between low and high Chl-a concentrations to be seen in the input sensitivity plot for the THEN- and the ELSE-branches of the rule in Fig. 15a and Fig. 15b respectively. If pH exceeds 7.9, higher Chl-a concentrations are forecasted by the ELSE-branch (Fig. 15b) compared to lower Chl-a concentrations forecasted for pH $\leq$ 7.9 by the THEN-branch (Fig. 15a). The input sensitivities for both branches of the rule indicate a positive relationship between Chl-a and pH. The highest change rate of Chl-a was observed for an increase of pH from 7.1 to 7.7 reflected by the THEN-branch. However, only a small increase in Chl-a concentrations is shown by the ELSE-branch within the range 7.8 $\leq$ pH $\leq$8.4.

In terms of nutrient dynamics, the THEN-branch shows that highest PO4-P concentrations coincide with higher Chl-a concentrations most likely reflecting sub-optimal underwater light conditions for phytoplankton growth during the monsoon season. By contrast the ELSE-branch seems to reflect optimal conditions for
phytoplankton growth outside the monsoon season where highest Secchi depths of up
to 2.4m and lowest NO$_3$-N$_{inflow}$ concentrations coincide with highest Chl-a
concentrations (see Fig. 15b).

Figure 16 shows the validation of the Chl-a model for Lake Putrajaya for the years 2003 to 2008. The model matches well peak events of Chl-a from 2005 to 2008 by predicting timing and magnitudes of such events. However the model fails to properly predict Chl-a dynamics between 2003 and 2004 even though matching the average Chl-a concentrations. During these two years 2003 and 2004, the Lake Putrajaya was still in an early development stage and gradually flooded with extreme hydrodynamic conditions resulting in highly fluctuating phytoplankton dynamics as can be seen in Fig. 16. The $r^2=0.33$ that has been achieved for the 7-days-ahead forecasting of chlorophyll-a in Lake Putrajaya can be considered as a reasonable result in spite of the fact that the data used for modelling captured the first 5 years of this man-made lake and therefore reflected unusual physical-chemical conditions in this lake. Furthermore the shallowness of the lake in combination with intense management impacts added to the highly stochastic nature of this lake during these first 5 years.

Input variable selection by HEA reveals pH and Secchi depth as the most important input variables for forecasting chlorophyll-a in Lake Putrajaya. This shows that HEA result is complimentary to SALMO-OO basic logistic equation that consider temperature, nutrient, light and mortality influences (section 2.3.2), where Secchi depth is related to light and pH does affect the form and availability of nutrient elements in water.
**Experiment 2**

Another experiment (experiment 2) was performed on four years of daily interpolated water quality data to develop a 7-days-ahead forecasting model for Chl-a of Lake Putrajaya. Data from the years 2003 and 2004 which represent early period of lake inundation was excluded from this experiment. All input variables used in experiment 1 were retained in this experiment. The one hundred runs of bootstrap training resulted in 100 models which were automatically ranked by HEA according to their RMSE and $r^2$.

Figure 17 illustrates the selection frequencies of the water quality variables considered as inputs for the forecasting of Chl-a in Lake Putrajaya during the 100 runs of bootstrap training. It clearly shows that pH, PO$_4$-P and Secchi depth (SD) appear to be key input variables for forecasting Chl-a whereas rainfall, solar radiation and NO$_3$-N seems to be the least important input variable.

![Figure 17 Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya (experiment 2)](image)

**Definitions for input symbols:**

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>NO$<em>3$-N$</em>{inflow}$</td>
<td>Nitrate inflow</td>
</tr>
<tr>
<td>pH</td>
<td>pH</td>
<td>PO$<em>4$-P$</em>{inflow}$</td>
<td>Phosphate inflow</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>Turbidity</td>
<td>Turbidity</td>
</tr>
<tr>
<td>NO$_3$-N</td>
<td>Nitrate</td>
<td>Cond</td>
<td>Conductivity</td>
</tr>
<tr>
<td>PO$_4$-P</td>
<td>Phosphate</td>
<td>Rainfall</td>
<td>Rainfall</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
<td>SR</td>
<td>Solar radiation</td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the following model has been identified as the best performing model:

IF(NO$_3$-N$_{inflow}$*43.926)*21.694 is less than 209.839

THEN the chlorophyll a concentration is calculated by the equation;
Chl-a = exp(SD) - (exp(ln(|DO|))*Ninflow)

Otherwise, by the ELSE equation:
Chl-a = (pH+(pH+(20.323-WT))) - (PO4-P*(((-39.669)/SD))

IF
NO3-Ninflow *43.926)*21.694 < 209.839

THEN
Chl-a = exp(SD) - (exp(ln(|DO|))*Ninflow)

ELSE
Chl-a = (pH+(pH+(20.323-WT))) - (PO4-P*(((-39.669)/SD))

Figure 18 Structure of the Chl-a model for Lake Putrajaya (experiment 2) a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE-branch (note: values in the legend box are the range of input data)

The model achieved a root mean square error (RMSE) of 2.28 and an $r^2$ of 0.41

The sensitivity analyses for the THEN- and the ELSE- branch are plotted in Fig.18. The IF-condition of the model determines a threshold value for NO3-Ninflow. Accordingly, the IF-condition of the model stipulates the threshold value NO3-Ninflow=0.22mg/l for distinguishing between high and low Chl-a concentrations in
Lake Putrajaya. As can be seen in the input sensitivity plot Fig 18, the THEN-branch of the rule in Fig. 18a applies to slightly higher Chl-a than the ELSE branch (Fig. 18b).

The sensitivity analysis for the THEN- and ELSE-branch indicate that the sensitivity of Chl-a is always high to Secchi depth, but the changing trends are opposite for the THEN- and ELSE-branches. When NO3-Ninflow value is less than 0.22 mg/l, Secchi depth is positively related to an increasing algal biomass (Fig. 18a). An increased Secchi depth provides more photosynthetic active light to the water column, promoting the algal growth. The high algal growth is fast consuming NO3-N nutrient and causes decrease in DO level. On the opposite, when NO3-Ninflow value is more than 0.22 mg/l, high level of Chl-a coincide with low Secchi depth, possibly caused by shading from increased algal biomass in the water column (Fig. 18b). The rising temperature level is negatively correlated with algal biomass indicating high water temperature of 31.1°C is limiting the algal growth. On the contrary, Figure 18b shows strong positive relationship between PO4-P and pH towards Chl-a concentrations.

In terms of nutrient dynamics, the THEN-branch reflects optimal phytoplankton growth with highest Secchi depths of up to 2.2m and lowest NO3-Ninflow concentrations coincide with highest Chl-a concentrations (see Fig. 18a). Increase in PO4-P concentrations coincide with higher Chl-a concentrations most likely reflecting sub-optimal for phytoplankton growth during the warmer temperature (dry) season in the lake (Fig. 18b).

Figure 19 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Putrajaya for the years 2005 to 2008 (experiment 2)

Figure 19 shows the validation of the Chl-a model for Lake Putrajaya for the years 2005 to 2008. The timing of the predicted Chl-a compares well with the measured data for most of the simulation period although the magnitude predicted is slightly under-estimated. Generally, the increasing trend of Chl-a concentrations
towards the year 2008 was successfully simulated by the model although the magnitude of prediction is slightly lower. This result can be considered promising taking into consideration the stochastic nature of the lake conditions as reflected by the data even though the first two years of the input data was excluded. The root mean square error (RMSE) of 2.28 and an \( r^2 \) of 0.41 achieved for this experiment are slightly better than results achieved from experiment 1 of Lake Putrajaya (RMSE=2.35 and an \( r^2=0.33 \)). In fact, the \( r^2=0.41 \) achieved for the 7-days-ahead forecasting of Chlorophyll-a in Lake Putrajaya during experiment 2 can be considered an improvement from experiment 1 and realistic result despite the highly fluctuating phytoplankton dynamics (Fig. 19). As discussed earlier, the limited data from this shallow man-made lake coupled with intense human interference further complicating the nature of this lake.
4.1.2 Predictive Algal Biovolume Model for Lake Kenyir

One year of daily interpolated water quality data of Lake Kenyir was used to develop a 7-days-ahead forecasting model for algal biovolume. One hundred runs of bootstrap training of HEA utilising randomly 80% of the lake data for training and 20% for testing resulted in 100 models.

Figure 20 illustrates the selection frequency of specific input variables during the 100 runs of HEA by the bootstrap training mode for Lake Kenyir. Highest frequencies were observed for PO\textsubscript{4}-P, pH and Secchi depth appearing as key input variables for forecasting algal biovolume whilst lowest frequencies were observed for rainfall and solar radiation (SR). The information about input selection frequencies as shown in Figure 20 can also be utilised for optimising the CPU time for developing models by HEA for cases with large datasets and high numbers of input variables. This is important particularly when the number of input variables available increases significantly.

![Figure 20](image)

**Figure 20** Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Kenyir

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>NO\textsubscript{3}-N\textsubscript{inflow}</td>
<td>Nitrate inflow</td>
</tr>
<tr>
<td>pH</td>
<td>pH</td>
<td>PO\textsubscript{4}-P\textsubscript{inflow}</td>
<td>Phosphate inflow</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>Cond</td>
<td>Conductivity</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
<td>Rainfall</td>
<td>Rainfall</td>
</tr>
<tr>
<td>NO\textsubscript{3}-N</td>
<td>Nitrate</td>
<td>SR</td>
<td>Solar radiation</td>
</tr>
<tr>
<td>PO\textsubscript{4}-P</td>
<td>Phosphate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As result of the 100 bootstrap runs by HEA the following best-performing model has been identified:

IF SD is greater than 2.94m,

THEN the algal biovolume level is calculated by the equation:

\[
BV = ((\exp(pH) + ((WT/PO_4-P)+35.230)*(((111.255)/\ln(NO_3-N)) + \ln(NO_3-N)))
\]

Otherwise, by the ELSE equation:

\[
BV = ((\exp(pH) + ((WT/PO_4-P)-99.014))\times WT)
\]

The model achieved a root mean square error (RMSE) of 42 and an \( r^2 \) of 0.86.

The rule suggests the Secchi depth=2.94m as threshold for distinguishing between high and low algal biovolume in Lake Kenyir. As can be seen in Fig. 21a the THEN-branch applies to slightly higher algal biovolume than the ELSE-branch.
The sensitivity analysis for the THEN-branch (Fig. 21a) indicates a negative relationship between algal biovolume and $\text{PO}_4$-$\text{P}$ concentrations by showing highest biovolumes at lowest $\text{PO}_4$-$\text{P}$ concentrations and vice versa. By contrast biovolumes are rising with increasing $\text{NO}_3$-$\text{N}$ concentrations. The observed relationships between biovolume and the nutrients suggest that Lake Kenyir experiences nitrogen limitation confirmed by the average $\text{NO}_3$-$\text{N}$/PO$_4$-$\text{P}$ of 0.58. A N:$\text{P}$ ratio smaller than 10 based on $\text{NO}_3$-$\text{N}$ and $\text{PO}_4$-$\text{P}$ values indicates a nitrogen limitation (Kappers, 1980).

Algal biovolume experiences little sensitivity to changes in pH and water temperature (WT) where increases in both input variables showed only slight increases of algal biovolumes.

The sensitivity analysis for the ELSE-branch (Fig. 21b) reveals a similar relationship between algal biovolume and $\text{PO}_4$-$\text{P}$ concentrations, pH and water temperature as observed for the THEN-branch at a somewhat higher biovolume level.

The Figure 22 shows the validation of the algal biovolume model for Lake Kenyir for the year 1992. The timing and magnitudes of the predicted algal biovolumes compare reasonably well with the measured data for the whole year by matching accurately the first peak in February and slightly under-estimating the second peak in September. The good performance of this model is also manifested by the $r^2 = 0.86$ and RMSE=42.

Even though this model has been developed and validated from 12 months of data only it can be assumed that it captured the moderate seasonality of tropical-stratified lakes being characterised by marginally lower algal biovolumes during monsoon seasons from October till December and April to May. During the rainy season, the Secchi depth decreases due to the highly turbid water entering the lake causing temporary light limitation to algal photosynthesis. The onset of monsoon season is also the period when water from the nutrient-enriched hypolimnion partially mixes with the nutrient poor epilimnion due to hydrodynamic turbulences (Yusoff et al., 1998). Yusoff and Ambak (1999) reported low chlorophyll-a in the lake due to shortage of nutrients especially nitrogen. This is concurring with the results of this study where nitrate increase results in increasing level of algal biovolume.

The explanation is depicted in Figure 23 that demonstrates the seasonal distribution of the data set. Figure 23 shows most data points that are below 2.941 meter are distributed in very late October until the middle of December. The peak of monsoon seasons that brought heavy rains from late October to the middle of December in the Lake Kenyir area are illustrated in Figure 24. High discharge rates decrease retention times, washing out cells and therefore, reduce the opportunity for large algal population to develop in the lake.
Figure 22 Validation of the 7-days-ahead forecasting model for algal biovolume (µm$^3$) in Lake Kenyir for 1992

Figure 23 showed most algal data points that are at Secchi depths below 2.9 meter are distributed in very late October until the middle of December. This is the peak of monsoon seasons in the Lake Kenyir area as illustrated in Figure 24.

Figure 23 Data division and seasonal distribution by the HEA model for algal biovolume (µm$^3$) in Lake Kenyir for 1992

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT/temperature</td>
<td>Water temperature</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
</tbody>
</table>
Figure 24 Seasonal data division by the HEA model for algal biovolume (µm³) showing monsoon seasons in Lake Kenyir for 1992

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT/temperature</td>
<td>Water temperature</td>
</tr>
<tr>
<td>Rainfall</td>
<td>Rainfall</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
</tbody>
</table>
4.1.3 Predictive Chl-a Model for Lake Penang

Twelve months of daily interpolated water quality data of Lake Penang between 2005 and 2006 were used to develop a 7-days-ahead forecasting model for chlorophyll-a. One hundred runs of HEA produced 100 models based on the bootstrap training mode utilising randomly 80% of the data for training and 20% for testing.

Figure 25 highlights Secchi depth, NH₄, pH and COD as key input variables for Chl-a in Lake Penang by having the highest selection frequencies.

As a result of the 100 bootstrap runs by HEA the best-performing model has been identified as follows:
IF \((\exp((SD*SD))\) is equal or less than 41.203;

THEN the chlorophyll a concentration is calculated by the equation;

\[
Chl-a = \frac{(\ln(\ln(NH_4)))*\ln(pH*83.295))}{WT}
\]

Otherwise, by the ELSE equation:

\[
Chl-a = \frac{(\ln(pH)*42.277)/\ln((\exp(COD)-(-353.408))))}{WT}
\]

The model achieved a root mean square error (RMSE) of 0.9337 and an \(r^2\) of 0.89.

The value of 41.203 is part of the rules obtained from HEA model that best performed with data from Lake Penang in this experiment. The corresponding SD value for the model was then obtained by trial and error approach, taking into account the range of SD value from Lake Penang.

\[
\text{IF} \quad \exp((SD*SD)) \leq 41.203 \\
\quad \text{THEN} \quad Chl-a=\frac{(\ln(\ln(NH_4)))*\ln(pH*83.295))}{WT} \\
\quad \text{ELSE} \quad Chl-a=\frac{(\ln(pH)*42.277)/\ln((\exp(COD)-(-353.408))))}{WT}
\]

\[\text{Figure 26 Structure of the Chlorophyll-a model for Lake Penang a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE-branch}\]

The IF condition of the rule is based on the Secchi depth (SD) and distinguishes between medium to high Chl-a levels, represented by the THEN-branch of the model (Fig. 26a) and medium Chl-a levels represented by the ELSE-branch of the model (Fig. 26b).
Validation of the 7-days-ahead forecasting model for Chl-a in Lake Penang for 2005 and 2006 is shown in Figure 27. The Figure 28 illustrates the seasonal correspondence of the IF criteria \( \exp((SD*SD) \) in relation with Chl-a showing that the THEN-branch is activated only from early August till early October 2005 and from early April till late June 2006. These are the periods when Secchi depth was below 1.8m. Figure 29 shows seasonal data division by the model with 7-day-delay input data in Lake Penang for 2005 and 2006 in relation to Secchi depth and measured chlorophyll-a.

The THEN-branch of the model includes water temperature (WT), pH and NH_4 (see Fig. 26a) and considers a strong negative relationship of Chl-a with NH_4 but almost neutral to slight negative relationships with WT and pH. The depicted relationship between Chl-a and NH_4 that Chl-a reaches highest concentrations when NH_4 is low and vice versa may hint to the fact that NH_4 is rapidly consumed by phytoplankton as soon as there are optimum algal growth conditions.

The sensitivity analyses for the ELSE-branch of the model as plotted in Fig 26b shows a strong negative relationship between Chl-a and chemical oxygen demand (COD) with lowest COD concentrations at high Chl-a levels and vice versa.

COD is positively correlated with BOD (Mohanty et al., 2008) and COD values are always higher than BOD. The source of COD in Lake Penang is from water of the river Muda which occasionally is pumped into the lake as storage water. Nakasone (2003) reported COD runoff concentrations of around 10–20 mg/L (average 12.9 mg/L) during the irrigation period in a small watershed covering 205 hectare, whose main industry is agriculture with paddy fields making up 23% of the land use of the watershed. In this research, the average COD concentration in Lake Penang is 22.16 mg/l (range from 10.8 to 35.1 mg/l). This high values occur when Lake Penang receives water that is pumped from the River Muda Paddy Irrigation Scheme that covers an area of 9035 hectares (Mon & Chang, 2008).

COD is a measure of the oxygen required for chemical oxidation of organic matter and oxidisable inorganic substances (Kawabe & Kawabe, 1997; Haroon et al., 2010). COD measures the total oxygen demand in the presence of a powerful oxidising agent such as ozone or a mixture of sulphuric acid and potassium dichromate; it is a measure of the total amount of oxygen required to stabilise the waste. The source of oxidising agent especially potassium is believed to be originated from potassium fertilizer applications which is common practise in the paddy field irrigated by River Muda.

Oxidisable pollutants in the lake triggered the depletion of dissolved oxygen in the lake. Therefore, high COD can poison the natural population of unadapted algal, thereby leading to the lower level of Chl-a. This is more apparent in lakes where water currents are generally not suffice to diffuse large amounts of organic substances.

Seretaki (as cited in Wei et al., 2001) found COD to play an important role in the formation of the algal community. Wei et al.(2001) also found that the growth of Microcystis would be inhibited by the increment of COD under the limitation of total nitrogen and increase in COD concentration will inhibit Synedra (diatoms) growth.
This is comparable with the current finding where Lake Penang was observed to be N-limited with an N/P ratio=8.8 (based on NO₃-N and PO₄-P).

Fig 26b also shows a slight positive relationship of Chl-a with pH by observing slight alkaline conditions at low Chl-a concentrations and medium alkaline conditions at increasing Chl-a concentrations. The relationship with pH indicates that with increasing algal biomass and therefore photosynthetic activity in the lake, the concentration of dissolved CO₂ is exhausting.

The comparison between the observed and 7-days-ahead forecasted Chl-a concentrations for 12 months between 2005 and 2006 in Lake Penang in Fig. 27 attests the model’s good forecasting validity as has also been reflected by the r² of 0.89. The observed Chl-a peaks in early September 2005, early February and early May 2006 are reasonably well forecasted by the model both in terms of the timing of fast algal growth as well as of the magnitude of Chl-a.

Figure 27 Validation of the 7-days-ahead forecasting model for Chl-a in Lake Penang for 2005 and 2006
Figure 28 Data division for the rule set in Lake Penang for 2005 and 2006 showing the input data are divided into two parts depending on the Secchi depth value.

Figure 29 Data division seasonal distribution by the model with 7-day-delay input data in Lake Penang for 2005 and 2006
4.1.4 Predictive Algal Biovolume Model for Lake Penang

Twelve months of daily interpolated water quality data of Lake Penang between August 2005 and July 2006 were used to develop a 7-days-ahead forecasting model for algal biovolume (BV). One hundred runs of HEA produced 100 models based on the bootstrap training mode utilising 80% and 20% of the data for training and testing respectively.

Figure 30 highlights COD (Chemical Oxygen Demand), TSS (Total Suspended Solids) and Secchi depth (SD) as key input variables for algal biovolume in Lake Penang by having the highest selection frequencies. Solar radiation (SR) seems to be the least important input variable for Lake Penang.

**Figure 30** Frequency of input variable selection by HEA during 100 bootstrap runs for algal biovolume prediction for Lake Penang

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH</td>
<td>NO₂⁻-N</td>
<td>Nitrate</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>PO₄³⁻-P</td>
<td>Phosphate</td>
</tr>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>NH₄</td>
<td>Ammonium</td>
</tr>
<tr>
<td>DO</td>
<td>Dissolved oxygen</td>
<td>TSS</td>
<td>Total suspended solids</td>
</tr>
<tr>
<td>BOD</td>
<td>Biochemical oxygen demand</td>
<td>SR</td>
<td>Solar radiation</td>
</tr>
<tr>
<td>COD</td>
<td>Chemical oxygen demand</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the best-performing model has been identified as the following:
IF COD is equal or less than 26.3 mg/l;

THEN the algal biovolume is calculated by the equation:

\[ BV = (((WT+((-76.690)/DO))+pH)*(WT+((pH*83.322)/(BOD+NO3-N)))) \]

Otherwise, by the ELSE equation:

\[ BV = ((\exp(pH)+\exp(\exp(NH4)))+(TSS)) \]

The model achieved a root mean square error (RMSE) of 41 and an \( r^2 \) of 0.95

The IF-condition of the model stipulates the threshold value COD=26.31 mg/l for distinguishing between high and low algal biovolume concentrations to be seen in the input sensitivity plot for the THEN- and the ELSE-branches of the rule in Fig. 31a and Fig 31b respectively. If COD exceeds 26.31 mg/l, lower algal biovolume are forecasted by the ELSE-branch (Fig 31b) compared to higher algal biovolume forecasted for COD \( \leq \) 26.31 mg/l by the THEN-branch (Fig 31a). The input range (%)

**Figure 31** Structure of the algal biovolume model for Lake Penang a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE-branch
sensitivities for both branches of the rule indicate a positive relationship between algal biovolume and pH with the highest change rate of algal biovolume was observed for an increase of pH from 7.4 to 8.0 reflected by the ELSE-branch. In contrast, only a minor increase in algal biovolume was observed in the THEN-branch within the pH range of 7.6 to 8.8.

The THEN-branch shows a weak positive relationship between algal biovolume and NO$_3$-N and water temperature respectively. This could be explained by algal consuming NO$_3$-N at an optimal pH and water temperature ranges. Increased in DO concentration from 4.9 mg/l to 7.6 mg/l was observed with an increasing algal biovolume, denoting higher photosynthesis rate is producing more oxygen in the lake. However, a strong negative relationship between algal biovolume and BOD was observed in the THEN-branch. This shows higher number algal biovolume is consuming more oxygen thus the high BOD value at higher algal biovolume while lower algal biovolume is consuming less DO therefore, the lower BOD observed.

Apart from strong positive relationship between algal biovolume and pH, the sensitivity analysis for the ELSE-branch reveals a strong positive relationship between algal biovolume and TSS and slightly weaker positive relationship to NH$_4$. As mentioned earlier, the ELSE-branch cater for lower concentration of algal biovolume, thus at this stage, the algal is rapidly growing and consuming NH$_4$. This consequently contributes to the increasing TSS due to algal shading effect.

The comparison between the observed and 7-days ahead forecasted algal biovolume for 12 months from August 2005 until July 2006 is shown in Fig. 32. Again, the model good forecasting validity was demonstrated by the $r^2 = 0.95$. The observed algal biovolume peak in January 2006 is realistically well forecasted by the model both in terms of the timing and magnitude.

![Figure 32 Validation of the 7-days-ahead forecasting model for algal biovolume in Lake Penang for the years 2005 to 2006](image-url)
4.2 Towards a Generic Forecasting Model for Algal Biovolume in Mesotrophic Tropical-stratified Lakes

This section documents results of the application of HEA for the development of a forecasting model for algal biovolume from merged limnological time series data from Lake Penang and Lake Kenyir. As indicated before, the two lakes Kenyir and Penang studied during this research belong to the category of mesotrophic tropical-stratified lakes. After having demonstrated in sections 4.1.2 and 4.1.4 that predictive models for algal biomass can be developed for both lakes separately by means of hybrid evolutionary algorithms (HEA) it would be desirable to have a generic algal biomass model available that is applicable and comparable for mesotrophic tropical-stratified lakes. Ideally, the rule-based models for algal population dynamics would be generic to some extent and applicable to more than one water body of the same category. Recknagel et al. (2008b, 2008c) demonstrated the possibility that generic data driven and process-based models can be developed for water bodies that belong to the same ecosystem category by utilising merged data sets. Welk et al. (2008) followed successfully this pathway by developing generic models for cyanobacteria populations in three different lake categories.

In this study two experiments have been conducted:

(1) 24 months of daily interpolated data from lakes Kenyir and Penang were merged and used to develop a 7-days-ahead forecasting model for algal biovolume for lakes Kenyir and Penang respectively. Appropriate attention was given to the correct positioning for these 24 months of merged data in the Microsoft Excel spread sheet in order to reflect continuity in the calendar month and to preserve the seasonality aspect in each dataset. To begin the experiment, Lake Penang data from August 2005 until December 2005 was placed before Lake Kenyir data (January to December 1992) and Lake Penang data from January 2006 until July 2006 was placed after Lake Kenyir data in sequence.

(2) 12 months of daily interpolated data from Lake Kenyir and Lake Penang were used to develop a 7-days-ahead forecasting model for algal biovolume for each lake separately. The model for Lake Kenyir was then being applied to Lake Penang and vice versa.

Algal biovolume was chosen as an output for both experiments. Phytoplankton cells numbers (cells/ml) data from both Lake Penang and Lake Kenyir were converted to biovolumes following the instructions by Strathmann (1967). The reason for that was that only Chl-a data were available from Lake Penang and only algal biovolume data were available from Lake Kenyir but algal abundance data were available from both lakes.

The individual biovolumes for the dominant four algal populations in each lake were multiplied with their abundances and summed up to get the algal community biovolume per unit volume. Data on species specific biovolumes were adopted from literature (see Ahlgren, 1970; Makarewicz, 1993; Barros et al., 2006; Rodrigo et al., 2009) and website (http://www.helcom.fi/helcom/en_GB/biovolumes/). In cases where the algal specific
biovolumes were not available, the closest genus biovolume was chosen as an alternative. The total algal biovolume for a lake is the summation of each species’ biovolume from respective lake.

4.2.1 Experiment 1

Figure 33 illustrates the selection frequencies of specific input variables during the 100 runs of HEA. It clearly shows that Secchi depth (SD) and pH appear to be key input variables for forecasting algal biovolume using merged data of Lake Kenyir and Lake Penang.

![Figure 33](image)

**Figure 33** Frequency of input variable selection by HEA during 100 bootstrap runs for merged data of lakes Penang and Kenyir

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH</td>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>NO3-N</td>
<td>Nitrate</td>
</tr>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>PO4-P</td>
<td>Phosphate</td>
</tr>
</tbody>
</table>

One hundred runs of HEA produced 100 models based on the bootstrap training mode. As result of the 100 bootstrap runs by HEA the best-performing model has been identified as the following:

IF exp(Secchi depth) < 30.962

THEN Biovolume = ((pH*SD)*((pH*SD)*88.238))

ELSE Biovolume = ((((147.948/SD)/PO4-P) + exp(pH))*14.098)

The model achieved a root mean square error (RMSE) of 83 and an $r^2$ of 0.76.
The value of “30.962” is part of the rule obtained from HEA model that best performed with data from these lakes. HEA may give result in terms of rules with formulation such as above, together with associated input variables. The rules are applied using Microsoft Excel spreadsheet.

\[
\text{IF} \quad \exp(\text{SD}) < 30.962 \\
\text{THEN} \\
\text{ELSE}
\]

\[
\text{Biovolume} = \left(\frac{147.948}{\text{SD}}/\text{PO}_4\text{P} + \exp(\text{pH})\right) \times 14.098
\]

\[
\text{Biovolume} = ((\text{pH} \times \text{SD}) \times ((\text{pH} \times \text{SD}) \times 88.238))
\]

**Figure 34** Structure of the model for algal biovolume from merged Lake Kenyir and Lake Penang input data a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE branch

The IF-condition of the model specifies the threshold value for Secchi depth at 3.43 meter. If it exceeds the value of 3.43, slightly higher algal biovolumes are forecasted by the ELSE-branch (Fig. 34b) compared to somewhat lower algal biovolumes forecasted for by the THEN-branch (Fig. 34a).

The input sensitivities plot for the THEN- and the ELSE- branches of the rule indicate a positive relationship between algal biovolume and pH. Fig. 34a and Fig. 34b both indicate that higher pH values coincide with faster algal growth resulting in higher algal biovolumes. Algal biovolume increase was observed across the pH range from pH 7.1 to pH 8.5 obviously reflecting the CO₂ consumption by algal photosynthesis.

The input sensitivities plotted for THEN-branch of the rule also indicate a strong positive relationship between algal biovolume and Secchi depth by showing highest algal biovolume at higher Secchi depth value (Fig. 34a) obviously reflecting some light limitation by turbid water most likely occurring during the monsoon season.
The sensitivity analysis for the ELSE-branch reveals a slightly negative relationship between algal biovolume and Secchi depth by showing that higher algal abundances cause self-shading effects. A negative relationship was also observed for the ELSE-branch (Fig. 34b) between biovolume and PO₄-P concentrations by showing highest biovolumes at lowest PO₄-P concentrations. This indicated the consumption of PO₄-P nutrient resulting in higher biovolumes of phytoplankton.

Figures 35 and 36 provide the validation results of the model for 7 days ahead forecasting of algal biovolume level in lakes Kenyir and Penang for the period of August 2005-July 2006 and January 1992–December 1992 respectively. Generally the model predicts the timing of peak events very well for both lakes. The model matches well the magnitude of a single major peak event of algal biovolume in Lake Penang in January 2006. Even though the small peaks observed in August 2005 and June 2006 in Lake Penang were not accurately predicted in terms of timing and peak, the magnitude of the prediction is well forecasted.

However, the model slightly overestimates and underestimates magnitude of months with relatively high peak event such as February and October 1992 in Lake Kenyir respectively. Despite these inaccuracies, the overall r² value of the linear regression between measured and predicted data of all 24 months for Lake Penang and Lake Kenyir amounts to 0.68 and 0.67 respectively. The validation results are considered good as they have achieved reasonable r² values despite the highly stochastic nature caused by periodic water withdrawals and seasonal induced up-welling in Lake Penang and Lake Kenyir respectively.

![Figure 35](image)

**Figure 35** Validation of the 7-days-ahead forecasting model for algal biovolumes in Lake Kenyir for 1992 developed by experiment 1
The resulting rule-based models proved to be both predictive and explanatory. It demonstrated that the interpretation of the model can be brought into the context of empirical and causal knowledge on algal biovolume dynamics under specific water quality conditions in tropical lake ecosystems. The model for algal biovolume in both Lake Penang and Lake Kenyir proved to be reasonably valid for the lake datasets and months tested in each case study.
4.2.2 Experiment 2

Twelve months of daily interpolated data from each lake were used separately from each other to generate rule-based models for algal biovolumes for each lake. Input data for pH, Secchi depth, water temperature, dissolved oxygen, NO$_3$-N and PO$_4$-P and output data for algal biovolume were used for this experiment. The ‘Kenyir model’ was then tested for Lake Penang and the ‘Penang model’ was tested for Lake Kenyir.

‘Lake Kenyir model’

Twelve months of daily interpolated data from Lake Kenyir were used for this experiment and the resulted models were then applied to Lake Penang. Result from application of the rule-based model to Lake Kenyir is shown first.

Figure 37 highlights pH, PO$_4$-P and Secchi depth (SD) as key input variables for algal biovolume in Lake Kenyir by having the highest selection frequencies. Water temperature (WT) seems to be the least important input variable for Lake Kenyir.

![Figure 37](image)

**Figure 37** Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Kenyir

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH</td>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>NO$_3$-N</td>
<td>Nitrate</td>
</tr>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>PO$_4$-P</td>
<td>Phosphate</td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the following model has been identified as the best performing model:

IF (exp(SD)>37.539)
THEN BV=(((exp(pH)/(PO$_4$-P/NO$_3$-N))+406.114)*59.630)
ELSE

\[ BV = (((\exp(pH) - (SD/NO_3-N)) + (NO_3-N/(pH/PO_4-P))) \times 41.898) \]

The model achieved a root mean square error (RMSE) of 41 and an \( r^2 \) of 0.87.

\[
\text{IF} \quad \exp(SD) > 37.539
\]

THEN

\[
\text{Biovolume} = \left( \left( \frac{\exp(pH) - (SD/NO_3-N)}{NO_3-N/(pH/PO_4-P)} \right) + 406.114 \right) \times 59.630
\]

ELSE

\[
\text{Biovolume} = (((\exp(pH)/(PO_4-P/NO_3-N)) + 406.114) \times 59.630)
\]

---

**Figure 38** Structure of the model of algal biovolume for Lake Kenyir

a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE branch (note: values in the legend box are the range of input data)

The IF-condition of the model stipulates the Secchi depth of 3.6 meter as threshold value for distinguishing between low and high algal. If Secchi depth exceeds 3.6 meter, lower algal biovolume are forecasted by the THEN-branch (Fig. 38a) compared to higher algal biovolume forecasted for Secchi depth lower than 3.6 meter by the ELSE-branch (Fig. 38b).

The input sensitivities for both branches of the rule indicate a positive relationship between algal biovolume and pH. A positive relationship was also observed between algal biovolume and PO\(_4\)-P for both branches with relationship in ELSE-branch showed stronger positive relationship compared to THEN-branch. The highest change rate of algal biovolume was observed for an increase of Secchi depth from 2.8 to 3.4 meter, PO\(_4\)-P concentrations from 0.03 to 0.06 mg/l and NO\(_3\)-N concentrations from 0.01 to 0.03 mg/l as reflected by the ELSE-branch. However, a weak positive relationship of algal biovolume and pH level was observed with only a
small increase in algal biovolume is shown by both branches for an increase of pH from 7.1 to 7.4 for THEN-branch and from 6.8 to 7.4 for ELSE-branch respectively.

The THEN-branch shows that highest PO₄-P concentrations coincide with higher algal biovolume most likely reflecting sub-optimal underwater light conditions for phytoplankton growth during the rainy season as was observed in the experiment with Lake Putrajaya at section 4.3.1. NO₃-N seems to be the limiting nutrient where the rapid consumption of NO₃-N has restricted the algal growth in Lake Kenyir at higher Secchi depth according to THEN-branch. By contrast the ELSE-branch seems to reflect optimal conditions for phytoplankton growth outside the monsoon season where highest Secchi depths of up to 3.4 meter, highest NO₃-N and PO₄-P concentrations coincide with highest algal biovolume (see Fig. 38b).

![Figure 39 Validation of the 7-days-ahead forecasting model for algal biovolume concentration in Lake Kenyir for the year 1992](image.png)

The Figure 39 shows the validation of the algal biovolume model for Lake Kenyir for the year 1992. The model matches well three peak events of algal biovolume in 1992 by reasonably predicting timing and magnitudes of such events. However the model fails to accurately predict algal biovolume magnitude in March 1992 peak and mistimed the peak in August 1992 even though matching perfectly the small peak of algal biovolume in May 1992. An $r^2$ value of 0.87 has been achieved for the 7-days-ahead forecast of algal biovolume in Lake Kenyir that can be considered as a good result. However, it must be noted that this prediction was based on one-year set of data and must be treated with caution.
‘Lake Penang model’

Twelve months of daily interpolated data from Lake Penang were used for this experiment and the resulted models were then applied to Lake Kenyir. Result from application of the rule-based model to Lake Penang is shown first.

Figure 40 highlights water temperature (WT), Secchi depth (SD) and pH as key input variables for algal biovolume in Lake Penang by having the highest selection frequencies. PO₄-P, dissolved oxygen (DO) and NO₃-N seem to be the least important input variable for Lake Penang.

**Figure 40** Frequency of input variable selection by HEA during 100 bootstrap runs for Lake Penang 2005/2006

Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH</td>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>SD</td>
<td>Secchi depth</td>
<td>NO₃-N</td>
<td>Nitrate</td>
</tr>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>PO₄-P</td>
<td>Phosphate</td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the following model has been identified as the best performing model:

IF WT ≥ 31.3°C
THEN BV = exp(exp(exp(ln(|SD|))))
ELSE
BV = exp((SD - (228.288/(SD - WT))))

The model achieved a root mean square error (RMSE) of 67 and an $r^2$ of 0.85
The IF-condition of the model stipulates the threshold value WT=31.3°C for distinguishing between low and high algal biovolume. If water temperature is equal to or exceeds 31.3°C, higher algal biovolume concentrations are forecasted by the ELSE-branch (Fig. 41b) compared to lower algal biovolume concentrations forecasted for water temperature lower than 31.3°C by the THEN-branch (Fig. 41a).

The input sensitivities for both branches of the rule indicate a positive relationship between algal biovolume concentrations and Secchi depth (SD).

The highest change rate of algal biovolume was observed for an increase of Secchi depth from 2.0 to 2.2 meter reflected by the THEN-branch with almost 5 times double the algal biovolume concentrations. The increasing concentration of algal biovolume was also observed in the ELSE-branch within the range 1.6- 2.4m albeit with a lesser degree of growth.

Figure 41 Structure of the model of algal biovolume for Lake Penang a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE branch (note: values in the legend box are the range of input data)
The ELSE-branch shows that highest Secchi depth coincide with higher algal biovolume concentrations most likely reflecting optimal underwater light conditions for phytoplankton growth. By contrast the THEN-branch seems to reflect optimal conditions for phytoplankton growth with the Secchi depths range from 2.0 to 2.2m (see Fig. 41a).

The input sensitivities for ELSE-branch of the rule also indicate a negative relationship between water temperature and algal biovolume concentrations where an increase of water temperature of up to 31.1°C is shown to decrease the algal biovolume concentrations. If this condition is met, then the algal growth is determined by the THEN-branch where Secchi depth plays a single role in algal growth functions according to the model.

![Figure 42 Validation of the 7-days-ahead forecasting model for algal biovolume in Lake Penang for 2005 and 2006](image)

The Figure 42 shows the validation of the algal biovolume concentration model for Lake Penang for August 2005 until July 2006. The model matches well a single major peak event of algal biovolume by well predicting timing and magnitudes of such event during January 2006. Both the early and end of year forecasted peaks were within a reasonable magnitude even though the model fails to properly matches the early peak in the beginning (August 2005) and the lake peak towards the end of Jun 2006.

After successful validation of rule-based model on Lake Kenyir and Lake Penang respectively, it was necessary to test the generality of the Lake Kenyir model by applying it to data from Lake Penang and vice versa (a water body that belongs to the same lake ecosystem category but was not used during the training process). The purpose of testing the rule-based model on different dataset from where it has been trained was to investigate whether the model could be generic for lakes that were not involved in the training process but belong to the same lake ecosystem category as those used for model training. However, this experiment that attempted to utilise rule-based model generated from Lake Kenyir onto Lake Penang and vice versa, in order to obtain generic rules was not successful. Several reasons can be proposed for the model shortcoming. Lake Penang has a range of biovolume from 2,685 to 59,325 cm$^3$/m$^3$ while Lake Kenyir has a range from 25,683 to 63,315cm$^3$/m$^3$. Therefore, the
model from Lake Kenyir has had no exposure to or experience with the conditions and pattern coinciding with any event below 25,683 cm$^3$/m$^3$ which helps to explain why the model from Lake Kenyir is not successfully applied in Lake Penang. The ranges of Secchi depth for both Lake Kenyir and Lake Penang are 2.7–5.6 m and 1.5–2.8 m respectively. The model had different exposure to Secchi depth value in respective lakes and very minimum overlap (around 0.1 m) exists in terms of Secchi depth value for both lakes. Ranges for water temperatures in both lakes also shown minimal overlap with temperature in Lake Kenyir range from 24.2 to 30.9°C while the temperature range in Lake Penang is 29.4–32.0°C. The similar pattern of minimal overlap was observed in pH data where the range in Lake Kenyir is 6.6 to 7.5 while in Lake Penang, the range is from 7.2 to 9.4.

Obviously, training the rule-based model with data that encompassed the entire range of conditions experienced in both lakes would provide improved results and this has been proven when the data from both lakes was merged as in the experiment of section 4.2.1. Therefore, the first way suggested earlier where merged data from Lake Kenyir and Lake Penang was pooled together with due consideration given on the seasonal pattern is a more suitable way to develop models generic for lake of same ecosystem categories.
4.3 Generic Forecasting Model for Algal Biovolume in Mesotrophic Tropical-stratified Lakes using Electronically Measured Data

Pursuant to the successful application of HEA for the development of a forecasting model for algal biovolume from merged limnological time series data from Lake Penang and Lake Kenyir, the next step would be to have an electronically measured input data for developing model for mesotrophic tropical-stratified lakes. Section 4.2 has demonstrated that predictive models for algal biomass can be developed for the both lakes separately by means of hybrid evolutionary algorithms (HEA). The following section document results from HEA experiment using electronically measured data only to develop a forecasting model.

Algal biovolume was chosen as an output for this experiment for the same reason as mentioned earlier in section 4.2. Conversion of phytoplankton numbers (cells/ml) to biovolume for both Lake Penang and Lake Kenyir data follows the instructions by Strathmann (1967) as mentioned in section 4.2. The data treatment steps detailed in section 4.2 was also adopted for this experiment. Electronically measured data from lakes Kenyir and Penang were merged and used to develop a 7-days-ahead forecasting model for algal biovolume with appropriate positioning and sequencing for the 24 months of merged data in the Microsoft Excel spread sheet to reflect continuity and seasonality in the calendar month.

4.3.1 Experiment with Merged Electronically Measured Data for Lakes Kenyir and Penang

Figure 43 illustrates the selection frequencies of specific input variables during the 100 runs of HEA. It clearly shows that pH and water temperature appear to be key input for forecasting algal biovolume among the other common electronically measured input data in merged data of Lake Kenyir and Lake Penang.

![Figure 43](image-url)  
**Figure 43** Frequency of input variable selection by HEA during 100 bootstrap runs for merged electronically measured data of lakes Penang and Kenyir
Definitions for input symbols:

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH</td>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One hundred runs of HEA produced 100 models based on the bootstrap training mode. As result of the 100 bootstrap runs by HEA the best-performing model has been identified as the following:

IF (WT ≤ 29.4)

THEN Biovolume = ((59.475 - WT) * ((WT * 2.519) + exp(pH)))

ELSE Biovolume = ((33.287 - WT) * (((WT * 225.912) + 139.048) + exp(pH)))

The model achieved a root mean square error (RMSE) of 10251 and an $r^2$ of 0.65.

**Figure 44** Structure of the model for algal biovolume from merged Lake Kenyir and Lake Penang electronically measured input data a) input sensitivity of the THEN-branch, b) input sensitivity of the ELSE branch
The IF-condition of the model specifies the threshold value for water temperature at 29.4°C. If the water temperature exceeds 29.4 °C, higher algal biovolumes are forecasted by the THEN-branch (Fig. 44a) compared to lower algal biovolumes forecasted for by the ELSE-branch (Fig. 44b).

Both the THEN- and the ELSE- branches of input sensitivities plot for the rule indicate a positive relationship between algal biovolume and pH. Higher growth in algal biovolumes was observed in the THEN- branch of the rule (Fig. 44a) compared to the ELSE-branch (Fig. 44b). Range of pH from 6.9 to 7.3 (THEN-branch) suits faster algal growth resulting in higher algal biovolumes compared to pH range from 7.4 to 8.4 (ELSE-branch). Algal biovolume increase was observed across the pH range from pH 6.9 to pH 8.5 obviously reflecting the CO₂ consumption by algal photosynthesis.

Both the THEN- and the ELSE- branches of input sensitivities plot for the rule indicate a negative relationship between algal biovolume and water temperature. The input sensitivities plot for THEN-branch of the rule indicate a weaker negative relationship between algal biovolume and water temperature by showing highest algal biovolume at lower water temperature (Fig. 44a) reflecting temperature limitation most likely occurring during the dry season.

The sensitivity analysis for the ELSE-branch reveals a stronger negative relationship between algal biovolume and water temperature by showing that lower algal abundances caused by warmer water temperature. The resulted lower algal biovolumes indicates the temperature limitation of algal growth in these lakes.

Figures 45 and 46 provide the validation results of the model for 7 days ahead forecasting of algal biovolume level in lakes Kenyir and Penang using only electronically measured input data for the period of August 2005-July 2006 and January 1992–December 1992 respectively. Generally the model predicts the timing of peak events reasonably well for both lakes.

Experiment using electronically measured input data only shows the model matches well the magnitude of major peak event of algal biovolume in Lake Kenyir in September 1992 albeit with slightly mistiming. However, similar to results from section 4.2, the model developed using electronically measured input data overestimates and underestimates magnitude of months with relatively high peak event such as February and October 1992 in Lake Kenyir respectively. Even though the early of the year small peak in Lake Kenyir was not satisfactorily captured by the model in terms of timing and peak, the subsequent downward trend was reasonably forecasted.

The overall $r^2$ value of the linear regression between measured and predicted data of all 24 months for both Lake Penang and Lake Kenyir are 0.65 respectively. Again, the validation results are considered good as they have achieved reasonable $r^2$ values despite utilising only the electronically measured input data for prediction in highly stochastic nature of both lakes. These $r^2$ value of the linear regression are comparable with $r^2$ value of 0.76 obtained by experiment in section 4.2 (utilising electronic and non-electronically measured input data).
The model matches well the timing of a single major peak event of algal biovolume in Lake Penang in January 2006. However, the magnitude of the prediction was slightly lower than the measured data. Even though the year end small peak forecasting in Lake Penang was not accurately predicted in terms of timing and peak, the magnitude of the prediction is reasonably forecasted.
Apart from generating rule-based models that are both predictive and explanatory, this result has shown that HEA was proven to be capable of utilising available electronically measured input data to generate models with similar capability. Despite the fact that electronically measured input data available from these two lakes are limited to pH, water temperature and dissolved oxygen only, the model obtained for algal biovolume in both Lake Penang and Lake Kenyir proved to be reasonably valid for the lake datasets and months tested in each case study. Models obtained by HEA from electronically measure input data are comparable in terms of causal knowledge with regard to interpretation of algal biovolume dynamics under tropical lake ecosystems.
4.3.2 Experiment with Merged Electronically Measured Data for Lake Putrajaya

Four years of daily interpolated electronically measured water quality data were used to develop a 7-days-ahead forecasting model for Chl-a of Lake Putrajaya. The one hundred runs of bootstrap training resulted in 100 models which were automatically ranked by HEA according to their RMSE and $r^2$.

Figure 47 illustrates the selection frequencies of the electronically measured water quality variables considered as inputs for forecasting Chl-a in Lake Putrajaya during the 100 runs of bootstrap training. It clearly shows that pH, water temperature and conductivity appear to be key electronically measured input variables for forecasting Chl-a whereas rainfall seems to be the least important input variable.

![Frequency of electronically measured input variable selection by HEA](image)

**Figure 47** Frequency of electronically measured input variable selection by HEA during 100 bootstrap runs for Lake Putrajaya

<table>
<thead>
<tr>
<th>Input symbols</th>
<th>definitions</th>
<th>Input symbols</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Water temperature</td>
<td>DO</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td>pH</td>
<td>pH</td>
<td>Turb</td>
<td>Turbidity</td>
</tr>
<tr>
<td>Rainfall</td>
<td>Rainfall</td>
<td>Cond</td>
<td>Conductivity</td>
</tr>
</tbody>
</table>

As result of the 100 bootstrap runs by HEA the following model has been identified as the best performing model:

IF pH is less than 8.3,

THEN the chlorophyll a concentration is calculated by the equation;

$$\text{Chl-a} = \text{DO-ln}(((29.716-\text{WT})/(\text{DO/Cond})))$$
Otherwise, by the ELSE equation:
\[
Chl-a = 10.876 - \ln(((DO/60.065)*4.582))
\]

\[
\text{IF}
\]
\[
\text{pH}<8.3
\]
\[
\text{THEN}
\]
\[
Chl-a = DO - \ln(((29.716-WT)/(DO/Cond)))
\]
\[
\text{ELSE}
\]
\[
Chl-a = 10.876 - \ln(((DO/60.065)*4.582))
\]

**Figure 48** Structure of the Chl-a model using electronically measured input data for Lake Putrajaya a) input sensitivity according to the THEN-branch, b) input sensitivity according to the ELSE-branch (note: values in the legend box are the range of input data)

The model achieved a root mean square error (RMSE) of 2.4 and an \( r^2 \) of 0.29.

The IF-condition of the model defines the threshold value pH=8.3 for distinguishing between low and high Chl-a concentrations to be seen in the input sensitivity plot for the THEN- and the ELSE-branches of the rule in Fig. 48a and Fig. 48b respectively. If pH exceeds 8.3, higher Chl-a concentrations are forecasted by the ELSE-branch (Fig. 48b) compared to lower Chl-a concentrations forecasted for pH < 8.3 by the THEN-branch (Fig. 48a). The input sensitivities of the rule indicate a positive relationship between Chl-a and DO for THEN-branch (Fig. 48a) while a negative relationship was observed between Chl-a and DO for ELSE-branch (Fig. 48b). For lower Chl-a concentrations, an increase of Chl-a was observed in the THEN-branch input sensitivity plot for an increase of DO from 6.9 to 7.9 mg/l. A reversal pattern for DO was observed at higher Chl-a concentrations in which high algal growth is consuming more DO leading to a decrease in DO level with an increase of Chl-a concentrations as shown by the ELSE-branch (Fig. 48b).
The THEN-branch shows a negative relationship between Chl-a concentrations and water temperature where highest water temperature coincides with lower Chl-a concentrations reflecting limitation of water temperature conditions for phytoplankton growth. The input sensitivities of the rule for THEN-branch (Fig. 48a) also indicate a weak negative relationship between Chl-a and conductivity. The Chl-a and conductivity relationship for THEN-branch reflects conditions for phytoplankton growth (see Fig. 48a). Conductivity is a measure of the ability of water to conduct an electric current. It is determined primarily by the number of ionic particles present. The THEN-branch (Fig. 48a) gives conductivity range of 73-103 µS/cm for Lake Putrajaya which is low in comparison to the Interim National Water Quality Standards for Malaysia (INWQS) of 1000 µS/cm. Low values of conductivity are common for soft waters such as Lake Putrajaya. As conductivity ranges also indicate cultural sources of ions such as runoff from urbanized areas, variations in watershed geology can result in natural fluctuations in conductivity as well. The low conductivity range inferred that the concentration of free metal ions is in short supply in the lake, therefore low level of metal pollution. Therefore, higher algal growth was observed at low level of conductivity as shown in Fig. 48a.

**Figure 49** Validation of the 7-days-ahead forecasting model for Chl-a in Lake Putrajaya for the years 2005 to 2008 using electronically measured input data

Figure 49 shows the validation of the Chl-a model using only electronically measured data for Lake Putrajaya for the years 2005 to 2008. The model matches reasonably peak events of Chl-a from 2005 to 2008 by predicting timing and magnitudes of such events. However the model over and under predicted the Chl-a dynamics in Lake Putrajaya for Jan-Feb 2005 and July 2005 respectively. Experiment using only electronically measured input data for Lake Putrajaya also shows highly fluctuating phytoplankton dynamics as can be seen in Fig. 49 due to extreme hydrodynamic conditions of the lake. The $r^2=0.29$ that has been achieved for the 7-
days-ahead forecasting of chlorophyll-a in Lake Putrajaya using electronically measured input data is comparable with the results from experiment using full input data as discussed in section 4.2. The result from this experiment is considered as a reasonable result despite the fact that only electronically measured input data were used for modelling but still capable of revealing basic pattern of algal growth in the lake. As mentioned in section 4.2, this newly built shallow lake coupled with frequent human interventions added to the highly stochastic nature of this lake.

In this research, rainfall data is considered available in real time via online network on top of its common standing as one of the electronically measurable parameters. This is based on the fact that Malaysia has already implemented comprehensive online hydrological data monitoring complete with real time rainfall information. The daily updated rainfall database is available for free to the general public as well.

The results from section 4.3 shows potential as an early warning model and will help identify tropical lakes experiencing deteriorating health conditions. This would cater to the management need for eutrophication forecasting and quick decision making in order to prepare for the recurrent algal bloom in the region.
4.4 Model Identification for Different Lake Categories by SALMO PLUS

One of the objectives of this research is to identify the optimal model structure for lakes with different physical, chemical and biological conditions. Data sets of Lakes Penang and Kenyir (Malaysia), Saidenbach (Germany), Roodeplaat (South Africa) and South Para (South Australia) covering a wide range of environmental conditions (see Table 12) have been chosen for this research to determine the most suitable model structures by means of the lake simulation system SALMO-PLUS.

SALMO-PLUS has been designed as a tool for assembling the optimum performing process-based model structure for a particular lake or lake category from a library of alternative process models implemented by object-oriented programming in JAVA (Martin & Recknagel, 2010). The design of SALMO-PLUS is a promising attempt to overcome structural rigidity of conventional process-based lake ecosystem models (Recknagel et al., 2008b).

<table>
<thead>
<tr>
<th>Lake</th>
<th>Category</th>
<th>Climate</th>
<th>Maximum depth (meter)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penang</td>
<td>Mesotrophic tropical-stratified</td>
<td>Tropical</td>
<td>44</td>
</tr>
<tr>
<td>Kenyir</td>
<td></td>
<td></td>
<td>145</td>
</tr>
<tr>
<td>Saidenbach</td>
<td>Dimictic mesotrophic</td>
<td>Temperate</td>
<td>45</td>
</tr>
<tr>
<td>Roodeplaat</td>
<td>Warm-monomictic eutrophic</td>
<td>Mediterranean</td>
<td>45.1</td>
</tr>
<tr>
<td>South Para</td>
<td>Warm-monomictic mesotrophic</td>
<td>Mediterranean</td>
<td>38</td>
</tr>
</tbody>
</table>

The library of the lake simulation system SALMO-PLUS includes alternative process models for algal growth and algal grazing as well as zooplankton growth and zooplankton mortality developed for four specific lake models by different authors (see Table 13).
Table 13 Sources of process models included in the library of SALMO-PLUS

<table>
<thead>
<tr>
<th>Source</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
</tr>
<tr>
<td>Hongping and Jianyi (2002)</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
</tr>
</tbody>
</table>

In order to identify the optimum model structure for a particular lake by testing all possible combinations of process-models in Tab. 13, SALMO-PLUS utilises particle swarm optimization (PSO) (Kennedy & Eberhart, 1995; Kennedy & Eberhart, 2001) obtained from an open source library and integrated into SALMO-PLUS.

Based on PSO, SALMO-PLUS consecutively performs: (1) multi-objective optimisation of the model structure, and (2) multi-objective optimisation of parameters of the optimum model structure where the minimisation of error between measured and calculated state variables, is used as a goal function:

$$\sum_{i=1}^{r} (Y_i(\text{obs}) - Y_i(\text{calc})) \rightarrow \text{Min}$$ (21)

In addition to the automatic optimisation steps (1) and (2), fine tuning of the model validity for a specific lake application was conducted by interactive adjustment of parameters according to specific lake conditions such as specific stock ratios between planktivorous and piscivorous fish by means of the parameters MOMIN and MOT determining zooplankton mortality. This step requires solid understanding of the process equations and is usually conducted by model authors, and uses both the goal function (1), and visual comparison of the measured and calculated state trajectories as criteria. Common quantitative expressions for assessing the goal function (1) are the squared correlation coefficient of a linear regression between predicted and measured data and the root mean square error (RMSE) that measures the differences between values predicted by the model and the measured values.
4.5 Mesotrophic tropical-stratified Lakes Kenyir and Penang

The data pre-processing for lakes Kenyir and Penang included following steps:

1. Determination of timing and extent of thermal stratification by means of monthly measured temperature profiles of the lakes
2. Calculation of decadal volumes, surface areas and mean mixing depths for the total lakes as well as epilimnion and hypolimnion section of the water bodies by means of measured water level and hypsographic curves.
3. Determination of decadal real mixing depths of the epilimnion of the lakes in 10 day steps
4. Determination of decadal photosynthetic active solar radiation entering the lakes’ surfaces
5. Determination of decadal mean water temperatures of the lakes’ epi- and hypolimnions
6. Determination of decadal inflow and outflow volumes, factor for strong rain events, as well as inflow phosphate, nitrate and detritus concentrations of the lakes’ epi- and hypolimnions
7. Daily interpolation from decadal values of all input variables
8. Determination of initial values for the state variable and lake specific parameter values such as LTMAX (maximum light transmission at 5m depth)

The following parameters that are related to the process models for algal growth and grazing as well as for zooplankton growth and mortality (see Table 14) were selected for multi-objective optimisation by SALMO-PLUS: KP, YX, KI, YZP, YZN, TOPTX, PHOTXMIN, POTXMAX, MOMIN, MOT and TOPTZ. The unit and definitions for each selected parameters are also given in Table 14.
<table>
<thead>
<tr>
<th>Parameters name and unit</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP (mgP/m³)</td>
<td>Half-saturation constant of the relationship between phosphorus and rate of photosynthesis of the phytoplankton at minimum phytoplankton biomass</td>
</tr>
<tr>
<td>YX (g wet weight/mgP)</td>
<td>Phosphorus-related yield coefficient of phytoplankton</td>
</tr>
<tr>
<td>KI (J/cm².d)</td>
<td>Half-saturation constant of the dependence of the photosynthesis rate of phytoplankton on light</td>
</tr>
<tr>
<td>YZP (g wet weight/mgP)</td>
<td>Phosphorus-related yield coefficient of zooplankton</td>
</tr>
<tr>
<td>YZN (g wet weight/gN)</td>
<td>Nitrogen-related yield coefficient of zooplankton</td>
</tr>
<tr>
<td>TOPTX (°C)</td>
<td>Optimal temperature for phytoplankton growth</td>
</tr>
<tr>
<td>PHOTXMIN (d⁻¹)</td>
<td>Value of PHOTX; PHOTXH near 0°C and optimal light and nutrient conditions</td>
</tr>
<tr>
<td>PHOTXMAX (d⁻¹)</td>
<td>Maximum value of PHOTX; PHOTXH at optimal conditions</td>
</tr>
<tr>
<td>MOMIN (d⁻¹)</td>
<td>Rate of zooplankton mortality near 0°C at zooplankton biomass Z, ZH much higher than KMO</td>
</tr>
<tr>
<td>MOT (°C⁻¹. d⁻¹)</td>
<td>Slope of the function MORTZ(T) and MORTZH(TH)</td>
</tr>
<tr>
<td>TOPTZ (°C)</td>
<td>Optimal temperature for feeding activity of the zooplankton</td>
</tr>
</tbody>
</table>
4.5.1 Simulation Results for Lake Kenyir by Means of the Simulation System SALMO-PLUS

Model Structure Optimisation

After the goal function (1) has been applied to the state variables PO₄-P and NO₃-N for which measured data of Lake Kenyir were available, the following optimal model structure has been identified by SALMO-PLUS:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
<td></td>
</tr>
</tbody>
</table>

Model Structure and Parameter Optimisation

Table 15 summarises the expert-based pre-set and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Lake Kenyir.

<table>
<thead>
<tr>
<th>State variables</th>
<th>Expert-based pre-set state variable values</th>
<th>Interactively fine-tuned state variable values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO₄-P Epilimnion</td>
<td>70 μg/l</td>
<td>60 μg/l</td>
</tr>
<tr>
<td>PO₄-P Hypolimnion</td>
<td>110 μg/l</td>
<td>120 μg/l</td>
</tr>
<tr>
<td>NO₃-N Epilimnion</td>
<td>10.0 mg/l</td>
<td>15.0 mg/l</td>
</tr>
<tr>
<td>NO₃-N Hypolimnion</td>
<td>10.0 mg/l</td>
<td>12.0 mg/l</td>
</tr>
<tr>
<td>Diatoms Epilimnion and Hypolimnion</td>
<td>2.0 and 0.2 cm³/m³</td>
<td>2.0 and 0.2 cm³/m³</td>
</tr>
<tr>
<td>Green Algae Epilimnion and Hypolimnion</td>
<td>5.0 and 0.2 cm³/m³</td>
<td>6.0 and 0.2 cm³/m³</td>
</tr>
<tr>
<td>Blue-green Algae Epilimnion and Hypolimnion</td>
<td>1.0 and 0.2 cm³/m³</td>
<td>2.0 and 0.2 cm³/m³</td>
</tr>
<tr>
<td>Zooplankton Epilimnion and Hypolimnion</td>
<td>2.0 and 0.3 cm³/m³</td>
<td>3.0 and 0.3 cm³/m³</td>
</tr>
</tbody>
</table>
Table 16 summarises the expert-based pre-set, optimised and interactively fine-tuned parameter values achieved by SALMO-PLUS for Lake Kenyir.

**Table 16** Expert-based pre-set, optimised and interactively fine-tuned parameter values achieved by SALMO-PLUS for Lake Kenyir; X1=diatoms, X2=green algae and X3=blue-green algae

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Expert-based pre-set parameter values</th>
<th>Optimised parameter values by SALMO-PLUS</th>
<th>Interactively fine-tuned parameter values</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP</td>
<td>X1=18</td>
<td>X1=14.1</td>
<td>X1=14.1</td>
</tr>
<tr>
<td></td>
<td>X2=16</td>
<td>X2=7.5</td>
<td>X2=7.5</td>
</tr>
<tr>
<td></td>
<td>X3=27</td>
<td>X3=38.4</td>
<td>X3=38.4</td>
</tr>
<tr>
<td>YX</td>
<td>X1=0.8</td>
<td>X1=1.2</td>
<td>X1=1.2</td>
</tr>
<tr>
<td></td>
<td>X2=0.41</td>
<td>X2=0.8</td>
<td>X2=0.8</td>
</tr>
<tr>
<td></td>
<td>X3=1</td>
<td>X3=0.5</td>
<td>X3=0.5</td>
</tr>
<tr>
<td>KI</td>
<td>X1=29</td>
<td>X1=37.45</td>
<td>X1=37.45</td>
</tr>
<tr>
<td></td>
<td>X2=29</td>
<td>X2=28.32</td>
<td>X2=28.32</td>
</tr>
<tr>
<td></td>
<td>X3=28</td>
<td>X3=28.32</td>
<td>X3=28.32</td>
</tr>
<tr>
<td>YZP</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>TOPTX</td>
<td>X1=21</td>
<td>X1=28.35</td>
<td>X1=28.35</td>
</tr>
<tr>
<td></td>
<td>X2=23</td>
<td>X2=26.03</td>
<td>X2=26.03</td>
</tr>
<tr>
<td></td>
<td>X3=30</td>
<td>X3=15</td>
<td>X3=15</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>X1=0.17</td>
<td>X1=0.1</td>
<td>X1=0.1</td>
</tr>
<tr>
<td></td>
<td>X2=0.35</td>
<td>X2=0.53</td>
<td>X2=0.53</td>
</tr>
<tr>
<td></td>
<td>X3=0</td>
<td>X3=0</td>
<td>X3=0</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>X1=2.37</td>
<td>X1=2</td>
<td>X1=2</td>
</tr>
<tr>
<td></td>
<td>X2=3.3</td>
<td>X2=3.5</td>
<td>X2=3.5</td>
</tr>
<tr>
<td></td>
<td>X3=2.37</td>
<td>X3=1.19</td>
<td>X3=1.19</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.0225</td>
<td>0.0075</td>
</tr>
<tr>
<td>MOT</td>
<td>0.006</td>
<td>0.0027</td>
<td>0.01218</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>20</td>
<td>27.35</td>
<td>27.35</td>
</tr>
</tbody>
</table>

The corresponding simulation results are represented in Fig. 50a and Fig. 50b.
<table>
<thead>
<tr>
<th>Lake Kenyir (Malaysia) 1992</th>
<th>PO₄-P Concentration (μg/l)</th>
<th>NO₃-N Concentration (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALMO-PLUS reference model structure (A1-A2-A3-A4) with expert-based pre-set parameters</td>
<td><img src="image1.png" alt="Graph" /> ( r^2 = 0.001 )</td>
<td><img src="image2.png" alt="Graph" /> ( r^2 = 0.02 )</td>
</tr>
<tr>
<td>SALMO-PLUS optimised model structure (A1-A2-A3-A4) with optimised and fine-tuned parameters</td>
<td><img src="image3.png" alt="Graph" /> ( r^2 = 0.003 )</td>
<td><img src="image4.png" alt="Graph" /> ( r^2 = 0.02 )</td>
</tr>
</tbody>
</table>

**Figure 50a** Simulation results of SALMO-PLUS for PO₄-P and NO₃-N concentrations in Lake Kenyir (1992). X-axis in days; ––– = simulated epilimnion; ---- = simulated hypolimnion; o = measured epilimnion
<table>
<thead>
<tr>
<th>Lake Kenyir (Malaysia) 1992</th>
<th><strong>Total Phytoplankton (cm³/m³)</strong></th>
<th><strong>Algal functional groups (cm³/m³)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>SALMO-PLUS reference model structure (A1-A2-A3-A4) with expert-based pre-set parameters</td>
<td><img src="image1" alt="Graph of Total Phytoplankton" /></td>
<td><img src="image2" alt="Graph of Algal functional groups" /></td>
</tr>
<tr>
<td>SALMO-PLUS optimised model structure (A1-A2-A3-A4) with optimised and fine-tuned parameters</td>
<td><img src="image3" alt="Graph of Total Phytoplankton" /></td>
<td><img src="image4" alt="Graph of Algal functional groups" /></td>
</tr>
</tbody>
</table>

**Figure 51b** Simulation results of SALMO-PLUS for Total phytoplankton and Algal functional groups biomass in Lake Kenyir (1992). X-axis in days; — = simulated epilimnion; ---- = simulated hypolimnion; For algal functional groups simulation, — = Diatoms; --- = Green Algae; x = Blue-green Algae.
The Fig. 50a indicates that the simulation results for PO₄-P concentrations in the epilimnion of Lake Kenyir improved after the parameters were optimised and matched well the magnitude and trends of the observed data. This was also reflected by a slightly higher $r^2$ value (0.003) and RMSE=71.5. By contrast there were no improvements in the simulation results for NO₃-N concentrations by parameter optimisation and the seasonal trends of the observed data were only poorly matched. One reason for the ineffective parameter optimisation for NO₃-N may be the fact that parameters subjected to optimisation were related to phyto- and zooplankton growth and zooplankton mortality.

The Fig. 50b summarises the simulation results for the biomass of Total phytoplankton and algal functional groups. A significant improvement of the results in terms of seasonality of algal biomass has been achieved by means of the parameter optimisation. The total phytoplankton trajectory in Fig. 50b (bottom left) reaches a first peak in the post-monsoon months where the lake receives nutrient rich water from the entering rivers and by surface runoff. This peak is largely produced by green algae (see Fig. 50b, bottom right) that are known to be opportunistic. However the fast growth of Blue-green algae contributes to this seasonal algal biomass peak as well. The second significant peak of algal biomass has been simulated before the onset of monsoon season in September and October where the lake experiences high wind events creating a complex pattern of horizontal circulation and strong seiches, which frequently re-suspends nutrients from the hypolimnion into the epilimnion of the lake. Resulting higher nutrient concentrations in the epilimnion as also simulated in Fig. 50a stimulate growth of phytoplankton leading to the biomass peak on September and October (Fig. 50b). Interestingly the simulation results suggest that this second peak is largely produced by diatoms (Fig. 50b, bottom right) which are favoured by turbulences by keeping them longer in the euphotic zone. Similar observations on seasonal phytoplankton growth in Lake Kenyir have been discussed by Yusoff et al., (1998).

Algal functional groups predicted by the SALMO-PLUS model are realistic for the mesotrophic conditions exhibited by Lake Kenyir regarding dynamic interactions between the dominant groups of diatoms, green algae and lesser important group of blue-green algae. The prediction on phytoplankton functional groups dynamics from SALMO-PLUS model is in agreement with findings from Yusoff et al.,(1998) where green algae were reported to be the most abundant group in terms of species and densities. The occurrence of blue-green algae and diatoms in the lake was also observed. The blue-green algae were predicted to peak during January–February period of the year on the same pattern with green algae where the effect of monsoon flushing is at its highest albeit with lower magnitude. However, the growth of blue green algae is suppressed by growth of green algae and diatoms in the following months throughout the whole year of simulation.

The SALMO-PLUS model predicts that green algae peak from January and maintain a steady growth until the middle of the year when diatoms growth supersedes green algae especially during September to October period. This growth peaks coincide with the onset and the ending of monsoon season in Lake Kenyir and could be explained by the prevailing condition of monsoonal season onset where the lake responds vigorously to high wind events caused by monsoon winds. The wind affects lakes by creating a complex circulation pattern that promotes rapid
reproduction of diatoms by maintaining them in suspension throughout the lake area thereby exposing them to nutrients. This is explainable by looking at study on tropical lake (Lake Lanao) by Lewis (1978) where a strong tendency for diatoms to thrive was observed when turbulence is maximal. This enables diatoms to out-compete blue-green algae and green algae for nutrients before eventually die-off due to monsoon seasons at the end of the year. Phytoplankton population decrease during the monsoon season mainly due to the dilution and reduced solar radiation (Yusoff & Lock, 1995).

In general, the particle swarm optimisation method and multi objective parameter optimisation have improved the SALMO-PLUS model validity for Lake Kenyir compared to the simulation using reference SALMO-OO model structure and expert-based pre-set parameter values. This method has proven to be able to identify the most suitable model structure and optimum parameters for Lake Kenyir.
4.5.2 Simulation Results for Lake Penang by Means of the Simulation System SALMO-PLUS

Model Structure Optimisation

The following optimal model structure has been identified by SALMO-PLUS after the goal function (1) of particle swarm optimisation method was applied to the state variables PO₄-P and Chl-a for which measured data of Lake Penang were available:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
<td></td>
</tr>
</tbody>
</table>

Model Structure and Parameter Optimisation

Multi objective parameter optimisation using particle swarm optimisation method was then performed on selected parameters for Lake Penang. Table 17 summarises the expert-based pre-set and interactively fine-tuned values for state variables achieved by SALMO-PLUS for Lake Penang.

**Table 17** Expert-based pre-set and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Lake Penang

<table>
<thead>
<tr>
<th>State variables</th>
<th>Expert-based pre-set state variable values</th>
<th>Interactively fine-tuned state variable values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO₄-P Epilimnion</td>
<td>70 μg/l</td>
<td>20 μg/l</td>
</tr>
<tr>
<td>PO₄-P Hypolimnion</td>
<td>110 μg/l</td>
<td>60 μg/l</td>
</tr>
<tr>
<td>NO₃-N Epilimnion</td>
<td>10.0 mg/l</td>
<td>2.0 mg/l</td>
</tr>
<tr>
<td>NO₃-N Hypolimnion</td>
<td>10.0 mg/l</td>
<td>4.0 mg/l</td>
</tr>
<tr>
<td>Diatoms Epilimnion and Hypolimnion</td>
<td>2.0 and 0.1 cm³/m³</td>
<td>0.2 and 0.1 cm³/m³</td>
</tr>
<tr>
<td>Green Algae Epilimnion and Hypolimnion</td>
<td>5.0 and 0.1 cm³/m³</td>
<td>3.0 and 0.1 cm³/m³</td>
</tr>
<tr>
<td>Blue-green Algae Epilimnion and Hypolimnion</td>
<td>1.0 and 0.2 cm³/m³</td>
<td>0.2 and 0.2 cm³/m³</td>
</tr>
<tr>
<td>Zooplankton Epilimnion and Hypolimnion</td>
<td>2.0 and 0.3 cm³/m³</td>
<td>15.0 and 0.3 cm³/m³</td>
</tr>
</tbody>
</table>
Table 18 summarises the expert-based pre-set, optimised and interactively fine-tuned parameter values achieved by SALMO-PLUS for Lake Penang.

**Table 18** Expert-based pre-set, optimised and interactively fine-tuned parameters values achieved by SALMO-PLUS for Lake Penang; (X1=diatoms, X2=green algae, X3= blue-green)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Expert-based pre-set parameter values</th>
<th>Optimised parameter values by SALMO-PLUS</th>
<th>Interactively fine-tuned parameter values</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP</td>
<td>X1=18</td>
<td>X1=26.1</td>
<td>X1=26.1</td>
</tr>
<tr>
<td></td>
<td>X2=16</td>
<td>X2=17.28</td>
<td>X2=17.28</td>
</tr>
<tr>
<td></td>
<td>X3=27</td>
<td>X3=26.5</td>
<td>X3=26.5</td>
</tr>
<tr>
<td>YX</td>
<td>X1=0.8</td>
<td>X1=1.08</td>
<td>X1=1.08</td>
</tr>
<tr>
<td></td>
<td>X2=0.41</td>
<td>X2=0.44</td>
<td>X2=0.44</td>
</tr>
<tr>
<td></td>
<td>X3=1</td>
<td>X3=0.54</td>
<td>X3=0.54</td>
</tr>
<tr>
<td>KI</td>
<td>X1=29</td>
<td>X1=43.5</td>
<td>X1=43.5</td>
</tr>
<tr>
<td></td>
<td>X2=29</td>
<td>X2=14.5</td>
<td>X2=14.5</td>
</tr>
<tr>
<td></td>
<td>X3=28</td>
<td>X3=32.93</td>
<td>X3=32.93</td>
</tr>
<tr>
<td>YZP</td>
<td>0.8</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>118.14</td>
<td>118.14</td>
</tr>
<tr>
<td>TOPTX</td>
<td>X1=21</td>
<td>X1=14.73</td>
<td>X1=14.73</td>
</tr>
<tr>
<td></td>
<td>X2=23</td>
<td>X2=8.8</td>
<td>X2=8.8</td>
</tr>
<tr>
<td></td>
<td>X3=30</td>
<td>X3=45</td>
<td>X3=45</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>X1=0.17</td>
<td>X1=0.2</td>
<td>X1=0.2</td>
</tr>
<tr>
<td></td>
<td>X2=0.35</td>
<td>X2=0.525</td>
<td>X2=0.525</td>
</tr>
<tr>
<td></td>
<td>X3=0</td>
<td>X3=0</td>
<td>X3=0</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>X1=2.37</td>
<td>X1=3.3</td>
<td>X1=3.3</td>
</tr>
<tr>
<td></td>
<td>X2=3.3</td>
<td>X2=4.9</td>
<td>X2=4.9</td>
</tr>
<tr>
<td></td>
<td>X3=2.37</td>
<td>X3=1.3</td>
<td>X3=1.3</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>MOT</td>
<td>0.006</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>20</td>
<td>29.9</td>
<td>29.9</td>
</tr>
</tbody>
</table>

Assuming less planktivorous and piscivorous fish effects on zooplankton abundance were observed at Lake Penang in comparison with Lake Kenyir, the zooplankton pre-set value was set =15.0 cm³/m³. All the parameter values derived from multi objective parameter optimisation were maintained for simulation of SALMO-PLUS. For MOMIN, MOT and TOPTZ, further fine-tuning was made to increase the zooplankton abundance in Lake Penang. This was in turn affecting the algal functional groups dynamic.

**Data treatment**

The simulation results using SALMO-PLUS for Lake Penang were achieved after a minor data treatment where the Secchi depth was reduced to 50% less than the original value with an assumption that this lake has a reduced transparency due to higher detritus level. The corresponding simulation results are represented in Fig. 51a and Fig. 51b.
The Fig 51a indicates the simulation results from the model SALMO-PLUS on PO$_4$-P and Chlorophyll-a in Lake Penang.

<table>
<thead>
<tr>
<th>Lake Penang (Malaysia) 2005</th>
<th>PO$_4$-P Concentration (µg/l)</th>
<th>Chlorophyll-a Concentration (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SALMO-PLUS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reference model structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A1-A2-A3-A4) with expert-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>based pre-set parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="PO$_4$-P Simulation" /></td>
<td><img src="image" alt="Chlorophyll-a Simulation" /></td>
</tr>
<tr>
<td></td>
<td>$r^2=0.001$</td>
<td>$r^2=0.004$</td>
</tr>
<tr>
<td><strong>SALMO-PLUS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>optimised model structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A1-A2-A3-A4) with</td>
<td></td>
<td></td>
</tr>
<tr>
<td>optimised and fine-tuned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="PO$_4$-P Simulation" /></td>
<td><img src="image" alt="Chlorophyll-a Simulation" /></td>
</tr>
<tr>
<td></td>
<td>$r^2=0.004$</td>
<td>$r^2=0.05$</td>
</tr>
</tbody>
</table>

**Figure 52a** Simulation results of SALMO-PLUS for PO$_4$-P and Chlorophyll-a concentrations in Lake Penang (2005). X-axis in days; ——— = simulated epilimnion; ---- = simulated hypolimnion; o = measured epilimnion
**PO₄-P simulation**

The results from Lake Penang simulation by SALMO-PLUS (Fig. 51a) using optimised model structure and fine-tuning of optimised parameter produced a good result with regards to prediction of PO₄-P state variable. In contrast, little improvement was observed in terms of statistical result for PO₄-P prediction between reference model structure with expert-based pre-set parameters (Fig. 51a, top left) and optimised model structure with optimised parameters (Fig. 51a, bottom left). The statistical value improvement was poor as reflected by the $r^2$ value of 0.004 and RMSE value of 29 from the optimised model structure and fine-tuning of optimised parameters simulation. This is compared to $r^2=0.001$ and RMSE=28 obtained for the simulation of SALMO-PLUS using reference model structure with expert-based pre-set parameters values (Fig. 51a, top left). However, the magnitude for PO₄-P nutrient dynamic was well matched by SALMO-PLUS simulation in both cases.

The SALMO-PLUS simulation predicted the timing of PO₄-P peak in the middle of the year a bit too early but produced great prediction improvement visually and quantitatively towards the year end where the PO₄-P measured data was adequately simulated.

**Chlorophyll–a simulation**

Simulation of chlorophyll-a dynamics in Lake Penang follows closely the measured phytoplankton data. Results from SALMO-PLUS simulation give the magnitude of chlorophyll-a prediction within the range of 3.0 to 62.0 mg/l. Chlorophyll-a peaks was simulated during months of August, October, February and June marking a constant algal growth pattern for approximately every three months in Lake Penang. Chlorophyll-a is calculated by the following function:

\[ \text{Chlorophyll-a} = 2.5 \times X \]

where $X = X_1 + X_2 + X_3$; $X_1$=diatoms; $X_2$=green algal and $X_3$=blue-green algae

Results from SALMO-PLUS simulation using optimised model structure and fine-tuned optimised parameter showed great improvement in term of chlorophyll-a prediction when compared to results from reference model structure with expert-based pre-set parameters values. This was reflected by the $r^2$ value of 0.05 and RMSE=15 resulted from optimised model structure with fine-tuned optimised parameter simulation (Fig. 51a, bottom right) compared to $r^2=0.004$ and RMSE=19 from reference model structure with expert-based pre-set parameters (Fig. 51a, top right). The timing and magnitude of chlorophyll-a simulation fits considerably well with the measured data for Lake Penang. Apart from the highest peak in August 2005, the SALMO-PLUS describes the first peak of measured data in October 2005 very well. The model predicted another peak in the month of June 2006 but with a slightly higher magnitude compared to the measured data. Even though the SALMO-PLUS simulation does not adequately predict the two extreme low data points during the month December 2005, it manages to capture reasonably well the fluctuations in the following months until the end of simulation.
The simulation results of SALMO-PLUS for total phytoplankton and algal functional groups biomass in Lake Penang are represented in Fig. 51b.

<table>
<thead>
<tr>
<th>Lake Penang (Malaysia) 2005</th>
<th>Total Phytoplankton (cm$^3$/m$^3$)</th>
<th>Algal functional groups (cm$^3$/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SALMO-PLUS reference model structure (A1-A2-A3-A4) with expert-based pre-set parameters</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>SALMO-PLUS optimised model structure (A1-A2-A3-A4) with optimised and fine-tuned parameters</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Figure 53** Simulation results of SALMO-PLUS for Total phytoplankton and Algal functional groups biomass in Lake Penang (2005). X-axis in days; ----- = simulated epilimnion; ---- = simulated hypolimnion; For algal functional groups simulation, --- = Diatoms; - - - = Green Algae; x = Blue-green Algae.
**Total phytoplankton simulation**

The prediction of phytoplankton biomass in Lake Penang using optimised model structure with fine-tuned optimised parameters (Fig. 51b, bottom left) yielded a reasonable growth pattern and magnitude with the maximum level of phytoplankton biomass predicted to be at 25.0 cm$^3$/m$^3$ and phytoplankton growth to peak at approximately every three months during the simulation year. The dynamics of phytoplankton simulated by SALMO-PLUS in this study is comparable with findings from Makhlough (2008). The maximum phytoplankton biomass values predicted is comparable with value from Lake Kenyir (26.6 cm$^3$/m$^3$) from this study denoting similarity in trophic level between these two lakes. Simulation using optimised model structure with fine-tuned optimised parameters (Fig. 51b, bottom left) revealed lowest level of phytoplankton biomass in Lake Penang was predicted to be in May 2006. This is explainable by looking at seasonality where period of February to May is a dry season for the area (Taylor et al., 1992). As a storage dam, the water from Lake Penang is being pumped out by demand to cater for diminishing water supply at the River Dua WTP downstream during dry season. Thus, the pumped out water contributed to the lower level of phytoplankton in the Lake Penang during May 2006.

**Algal functional groups simulation**

Prediction of algal functional groups in Lake Penang by SALMO-PLUS is realistic of the mesotrophic conditions of the lake with the dynamic interaction between diatoms and green algae were in synchronised and dominated by green algae. This is in agreement with result from study by Makhlough (2008) where the dominant division was observed to be the Chlorophyta. The magnitude of green algae prediction is slightly higher in simulation from optimised structure with fine-tuned optimised parameters (Fig. 51b, bottom right) when compared to results from simulation with reference model structure with expert-based pre-set parameters (Fig. 51b, top right). Results from simulation with reference model structure and expert-based pre-set parameters (Fig. 51b, top right) gives a slightly lower abundance of diatoms and blue-green algae for the first half of simulation year compared to results from simulation with optimised model structure and fine-tuned optimised parameters (Fig. 51b, bottom right).

The blue-green algae was simulated having very low abundance by SALMO-PLUS. Makhlough (2008) also reported very low Cyanophyta from January to July 2006. The results from algal functional dynamic confirms to the expectation that Lake Penang is a typical tropical water body with no seasonal pattern observed.
4.6 Dimictic Mesotrophic Saidenbach Reservoir

4.6.1 Simulation Results by Means of the Simulation System SALMO-PLUS for Saidenbach Reservoir, Germany

Saidenbach Reservoir (Saxony, Germany) is of great importance for drinking water supply and has been investigated continuously and intensively since 1975 (Horn, 2003). The total volume of this reservoir is 22.4 Mm$^3$ with maximum depth of 45m and average depth of 15.3m (Horn, 2003). Saidenbach Reservoir has been classified as mesotrophic with green algae and large diatoms dominating its phytoplankton community (Cetin, 2007). SALMO-PLUS simulation was performed on measured data for phytoplankton biomass, phytoplankton and nutrient (phosphate) in 1975 that were supplied by Benndorf and Recknagel (1982) using particle swarm optimisation. Saidenbach reservoir was used for testing because it was used to test the original SALMO model, its data are of high quality and include zooplankton biomass data, which can be difficult to obtain (Cetin, 2007).

Model Structure Optimisation

After the goal function (1) has been applied to the state variables PO$_4$-P, total phytoplankton and zooplankton for which measured data of Saidenbach Reservoir were available, SALMO-PLUS has identified the following as the best model structure (in bold):

<table>
<thead>
<tr>
<th>Process models</th>
<th>Authors</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>-</td>
<td>-</td>
<td>A3</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td>-</td>
<td>D2</td>
<td>-</td>
<td>D4</td>
<td></td>
</tr>
</tbody>
</table>

The SALMO-PLUS simulation revealed the optimum model structure for algal growth and algal grazing in this research for Saidenbach Reservoir that are similar with the best performing model structure reported by Cetin (2007) and Recknagel et al. (2008b). Both Cetin (2007) and Recknagel et al. (2008b) selected Park et al. (1974) and Arhonditsis and Brett (2005) model structure as the best performing model for algal growth and grazing in Saidenbach Reservoir respectively.
Model Structure and Parameter Optimisation

Table 19 summarises the expert-based pre-set and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Saidenbach Reservoir and Table 20 summarises the expert-based pre-set and optimised parameter values achieved by SALMO-PLUS for Saidenbach Reservoir.

### Table 19
Expert-based pre-set and interactively fine-tuned initial values for state variables of Saidenbach Reservoir for SALMO-PLUS simulation

<table>
<thead>
<tr>
<th>State variables</th>
<th>Expert-based pre-set state variable values</th>
<th>Interactively fine-tuned state variable values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$PO_{4}$-P</td>
<td>8.0 $\mu$g/l</td>
<td>8.0 $\mu$g/l</td>
</tr>
<tr>
<td>$NO_3$-N</td>
<td>18.0 mg/l</td>
<td>18.0 mg/l</td>
</tr>
<tr>
<td>Diatoms</td>
<td>0.06 cm$^3$/m$^3$</td>
<td>0.06 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Green Algae</td>
<td>0.06 cm$^3$/m$^3$</td>
<td>0.06 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Blue-green Algae</td>
<td>0.04 cm$^3$/m$^3$</td>
<td>0.04 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Zooplankton</td>
<td>0.3 cm$^3$/m$^3$</td>
<td>0.3 cm$^3$/m$^3$</td>
</tr>
</tbody>
</table>

### Table 20
Expert-based pre-set and fine-tuned optimised parameter values achieved by SALMO-PLUS for Saidenbach Reservoir (X1=diatoms, X2=green algae, X3=blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Expert-based pre-set parameter values</th>
<th>Optimised parameter values by SALMO-PLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>STW</td>
</tr>
<tr>
<td>KP</td>
<td>X1=21.6</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>X2=19.2</td>
<td>24.6</td>
</tr>
<tr>
<td></td>
<td>X3=21.6</td>
<td>20.5</td>
</tr>
<tr>
<td>YX</td>
<td>X1=0.96</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>X2=0.328</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>X3=1</td>
<td>1.5</td>
</tr>
<tr>
<td>KI</td>
<td>X1=29</td>
<td>28.8</td>
</tr>
<tr>
<td></td>
<td>X2=23.2</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>X3=28</td>
<td>31.1</td>
</tr>
<tr>
<td>YZP</td>
<td>0.96</td>
<td>1.31</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>108.8</td>
</tr>
<tr>
<td>TOPTX</td>
<td>X1=16.8</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td>X2=27.6</td>
<td>28.8</td>
</tr>
<tr>
<td></td>
<td>X3=25.0</td>
<td>16.5</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>X1=0.17</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>X2=0.35</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>X3=0</td>
<td>0</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>X1=2.37</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>X2=3.96</td>
<td>4.32</td>
</tr>
<tr>
<td></td>
<td>X3=2.37</td>
<td>1.87</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.018</td>
<td>0.012</td>
</tr>
<tr>
<td>MOT</td>
<td>0.0048</td>
<td>0.006</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>16</td>
<td>23.1</td>
</tr>
</tbody>
</table>
The corresponding simulation results are represented in Fig. 52a and Fig. 52b. The simulation results of SALMO-PLUS for PO₄-P concentration and total phytoplankton in Saidenbach Reservoir are represented in Fig. 52a.

<table>
<thead>
<tr>
<th>Saidenbach Reservoir (Germany) 1975</th>
<th><strong>PO₄-P Concentration (μg/l)</strong></th>
<th><strong>Total phytoplankton (cm³/m³)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>SALMO-PLUS Reference structure (A1-A2-A3-A4) with expert-based pre-set parameters</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>SALMO-PLUS Optimised (B1-D2-A3-D4) structure with expert-based pre-set parameters</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>SALMO-PLUS Optimised structure (B1-D2-A3-D4) with optimised parameters</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Figure 54** Simulation results of SALMO-PLUS for PO₄-P and Total phytoplankton in Saidenbach Reservoir (1975), X-axis in days; — = simulated epilimnion, --- = simulated hypolimnion, = measured data.
**PO4-P simulation**

The simulation results for Saidenbach reservoir by SALMO-PLUS using optimised model structure with optimised parameter produced a good result with reasonable magnitude and timing of PO4-P state variable dynamics (Figure 52a, bottom left). The statistical value improvement was also good as reflected by improvement from $r^2$ value of 0.002 and RMSE value of 5.2 from the reference model structure with expert-based pre-set parameter simulation ((Figure 52a, top left) to $r^2$ value of 0.43 and RMSE=2.3 from the optimised model structure with optimised parameters simulation (Figure 52a, bottom left). The $r^2$ value of 0.43 with RMSE=2.3 from this study are also better than $r^2$ value of 0.18 and RMSE=3.63 reported in study by Cetin (2007).

The PO4-P simulation by SALMO-PLUS using optimised model structure with optimised parameters was over-predicted for the first quarter of the simulation but the subsequent prediction was well matched with measured data until the end of simulation. The end point of PO4-P measured data was also reasonably simulated by SALMO-PLUS. In general, the SALMO-PLUS simulation using optimised model structure with optimised parameters has managed to predict PO4-P state variable dynamics in Saidenbach Reservoir very well in terms of magnitude and peak of timing (Figure 52a, bottom left) compared to both simulations using reference model structure with expert-based pre-set parameters (Figure 52a, top left) and using optimised model structure with expert-based pre-set parameters (Figure 52a, middle left).

**Total phytoplankton simulation**

The SALMO-PLUS prediction of phytoplankton biomass in Saidenbach Reservoir using optimised model structure with optimised parameters (Figure 52a, bottom right) gave good results compared to SALMO-PLUS simulation using optimised model structure with expert-based pre-set parameter (Figure 52a, middle right) and reference model structure with expert-based pre-set parameters (Figure 52a, top right). The improvement on statistical value was also observed with an $r^2=0.302$ and RMSE=2.7 for SALMO-PLUS simulation from optimised model structure with optimised parameters. This was better compared to $r^2=0.05$ and RMSE=5.3 obtained from simulation using reference model structure with expert-based pre-set parameters (Figure 52a, top right) and $r^2=0.11$ and RMSE=5.8 obtained from simulation using optimised model structure with expert-based pre-set parameters (Figure 52a, middle right).

The maximum level of total phytoplankton biomass simulated using optimised model structure with optimised parameters was 7.2 cm$^3$/m$^3$. This was slightly over predicted, however, the SALMO-PLUS simulation using optimised model structure with optimised parameters managed to capture peak during autumn. In this simulation experiment, slight over prediction is better than under prediction since the magnitude did not differ much. Generally, the SALMO-PLUS simulation using optimised model structure with optimised parameters is a better compromise compared to the other simulation result due to realistic magnitude and phytoplankton growth timing.
The simulation results of SALMO-PLUS for zooplankton biomass in Saidenbach Reservoir are represented in Fig. 52b.

<table>
<thead>
<tr>
<th>Saidenbach Reservoir (Germany)</th>
<th>Zooplankton biomass (cm$^3$/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td></td>
</tr>
</tbody>
</table>

**SALMO-PLUS Reference model structure (A1-A2-A3-A4) with expert-based pre-set parameters**

![Graph](attachment:graph1.png) $r^2=0.05$

**SALMO-PLUS Optimised model structure (B1-D2-A3-D4) with expert-based pre-set parameters**

![Graph](attachment:graph2.png) $r^2=0.06$

**SALMO-PLUS Optimised model structure (B1-D2-A3-D4) with optimised parameters**

![Graph](attachment:graph3.png) $r^2=0.31$

**Figure 55** Simulation results of SALMO-PLUS for zooplankton biomass in Saidenbach Reservoir (1975); *, X-axis in days; --- = simulated epilimnion, --- = simulated hypolimnion, = measured data
**Zooplankton biomass simulation**

The simulation of zooplankton biomass dynamics in Saidenbach Reservoir by SALMO-PLUS is realistic for the mesotrophic conditions of the reservoir with the dynamic interaction between zooplankton (Fig. 52b) and total phytoplankton (Fig. 52a) were observed during summer. Fluctuation of zooplankton was high during summer period where growth of green algae and large diatoms was also simulated to peak (Fig. 52b, bottom). The SALMO-PLUS simulation using optimised model structure with optimised parameters (Fig. 52b, bottom) resulted in better statistical value with $r^2=0.31$ and RMSE=2.2 compared to $r^2=0.05$ and RMSE=2.6 than simulation using reference model structure with expert-based pre-set parameters (Fig. 52b, top) and $r^2=0.05$ and RMSE=2.8 from simulation using optimised model structure with expert-based pre-set parameters (Fig. 52b, middle). However, even though SALMO-PLUS simulation using optimised model structure with optimised parameters manages to simulate zooplankton dynamic reasonably well, it still cannot adequately predict any of the higher measured data observed at the beginning of summer in Saidenbach reservoir.
4.7 Warm-monomictic Hypertrophic Lake

4.7.1 Simulation Results by Means of the Simulation SALMO-PLUS for Roodeplaat Dam, South Africa

Roodeplaat Dam is located 25 km north east of Pretoria, South Africa and has exhibited eutrophication problems since the 1980s (Cetin, 2007). This recreational lake with a surrounding area of 668 km² is categorised as eutrophic to hypertrophic and exhibit warm monomictic conditions where stratification occurs once during summer. Roodeplaat Dam is severely eutrophied and has been placed amongst the most eutrophied impoundments in South Africa (Steyn et al., 1976).

Model Structure Optimisation

After the goal function (1) of particle swarm optimisation method has been applied to the state variables PO₄-P, NO₃-N, Total phytoplankton and Chl-a for which measured data of Roodeplaat Dam were available, the following optimal model structure has been identified by SALMO-PLUS:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td></td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
</tbody>
</table>

The best performing model structure for algal growth and algal grazing yielded in this research by SALMO-PLUS for Roodeplaat Dam are different from the best performing model structure reported by Cetin (2007) and Recknagel et al. (2008b). Both Cetin (2007) and Recknagel et al. (2008b) selected Arhonditsis and Brett (2005) model structure as the best performing model for algal growth and grazing for Roodeplaat Dam.

Model Structure and Parameter Optimisation

Multi objective parameter optimisation using particle swarm optimisation method was then conducted on selected parameter for Roodeplaat Dam. Table21 summarises the reference and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for Roodeplaat Dam and Table 22 summarises the initial and optimised parameter values achieved by SALMO-PLUS for Roodeplaat Dam.
Table 21 Reference and interactively fine-tuned initial values for state variables of Roodeplaat Dam for SALMO-PLUS simulation

<table>
<thead>
<tr>
<th>State variables</th>
<th>Reference Initial State Variable Values</th>
<th>Initial State Variable Values interactively fine-tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO₄-P</td>
<td>130 μg/l</td>
<td>130 μg/l</td>
</tr>
<tr>
<td>NO₃-N</td>
<td>1.03 mg/l</td>
<td>1.03 mg/l</td>
</tr>
<tr>
<td>Diatoms</td>
<td>6.5 cm³/m³</td>
<td>6.5 cm³/m³</td>
</tr>
<tr>
<td>Green Algae</td>
<td>4.5 cm³/m³</td>
<td>4.5 cm³/m³</td>
</tr>
<tr>
<td>Blue-green Algae</td>
<td>10.0 cm³/m³</td>
<td>10.0 cm³/m³</td>
</tr>
<tr>
<td>Zooplankton</td>
<td>5.0 cm³/m³</td>
<td>5.0 cm³/m³</td>
</tr>
</tbody>
</table>

Table 22 Initial values for parameters and optimised parameter values achieved by SALMO-PLUS for Roodeplaat Dam (X1=diatoms, X2=green algae, X3= blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Values</th>
<th>Parameter Values optimised by SALMO-PLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>STW</td>
</tr>
<tr>
<td>KP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=14.8</td>
<td>20.2</td>
<td>7.4</td>
</tr>
<tr>
<td>X2=12.8</td>
<td>19.2</td>
<td>6.4</td>
</tr>
<tr>
<td>X3=21.6</td>
<td>10.8</td>
<td>10.8</td>
</tr>
<tr>
<td>YX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=0.64</td>
<td>0.32</td>
<td>0.96</td>
</tr>
<tr>
<td>X2=0.49</td>
<td>0.73</td>
<td>0.24</td>
</tr>
<tr>
<td>X3=0.8</td>
<td>1.14</td>
<td>1.20</td>
</tr>
<tr>
<td>KI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=29</td>
<td>14.5</td>
<td>43.5</td>
</tr>
<tr>
<td>X2=29</td>
<td>43.5</td>
<td>43.5</td>
</tr>
<tr>
<td>X3=28</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>YZP</td>
<td>0.68</td>
<td>0.34</td>
</tr>
<tr>
<td>YZN</td>
<td>130</td>
<td>103.2</td>
</tr>
<tr>
<td>TOPTX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=25.2</td>
<td>24.6</td>
<td>12.6</td>
</tr>
<tr>
<td>X2=25.6</td>
<td>13.7</td>
<td>19.9</td>
</tr>
<tr>
<td>X3=24.0</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=0.17</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>X2=0.35</td>
<td>0.52</td>
<td>0.17</td>
</tr>
<tr>
<td>X3=0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1=2.37</td>
<td>1.18</td>
<td>1.18</td>
</tr>
<tr>
<td>X2=3.3</td>
<td>2.98</td>
<td>4.94</td>
</tr>
<tr>
<td>X3=2.37</td>
<td>2.97</td>
<td>3.55</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.0075</td>
</tr>
<tr>
<td>MOT</td>
<td>0.0048</td>
<td>0.0024</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>16</td>
<td>8.00</td>
</tr>
</tbody>
</table>
The corresponding simulation results are represented in Fig. 53a and Fig. 53b. The simulation results of SALMO-PLUS for PO$_4$-P and NO$_3$-N in Roodeplaat Dam are represented in Fig. 53a.

<table>
<thead>
<tr>
<th>Roodeplaat Dam (South Africa) 2003</th>
<th>PO$_4$-P Concentration (µg/l)</th>
<th>NO$_3$-N Concentration (mg/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SALMO-PLUS Reference model structure (A1-A2-A3-A4) with default parameters</strong></td>
<td><img src="image1" alt="Graph PO$_4$-P" />  $r^2=0.31$</td>
<td><img src="image2" alt="Graph NO$_3$-N" />  $r^2=0.28$</td>
</tr>
<tr>
<td><strong>SALMO-PLUS Optimised model structure (A1-A2-A3-A4) with default parameters</strong></td>
<td><img src="image3" alt="Graph PO$_4$-P" />  $r^2=0.29$</td>
<td><img src="image4" alt="Graph NO$_3$-N" />  $r^2=0.35$</td>
</tr>
<tr>
<td><strong>SALMO-PLUS Optimised model structure (A1-A2-A3-A4) with optimised parameters</strong></td>
<td><img src="image5" alt="Graph PO$_4$-P" />  $r^2=0.31$</td>
<td><img src="image6" alt="Graph NO$_3$-N" />  $r^2=0.33$</td>
</tr>
</tbody>
</table>

*Figure 56a* Simulation results of SALMO-PLUS for PO$_4$-P and NO$_3$-N in Roodeplaat Dam (2003), X-axis in days; — = simulated epilimnion, -- = simulated hypolimnion, = measured data
**PO₄-P simulation**

The Fig. 53a indicates that the simulation results for PO₄-P concentrations in Roodeplaat Dam by SALMO-PLUS using optimised model structure with optimised parameter produced a fair result with regards to magnitude and timing of PO₄-P dynamics. However, there was no significant improvement shown by the SALMO-PLUS simulation using optimised model structure with optimised parameters (Fig. 53a, bottom left) compared to simulation using optimised model structure with default parameters (Fig. 53a, middle left) and simulation using reference model structure with default parameters (Fig. 53a, top left). The $r^2$ value obtained from simulation using optimised model structure with optimised parameters was 0.31 and RMSE=148. This is slightly lower than the statistical results from simulation reported by Cetin (2007) with $r^2=0.26$ and RMSE=129. The magnitude and timing of PO₄-P concentrations simulated by SALMO-PLUS are comparable with result from Cetin (2007). However, SALMO-PLUS shows over prediction towards the end of simulation and did not adequately predict the measured data.

**NO₃-N simulation**

The SALMO-PLUS simulation for NO₃-N concentrations in Roodeplaat Dam using optimised model structure with optimised parameter produced little improvement with regards to statistical values. An $r^2=0.33$ and RMSE=0.9 was obtained by simulation using optimised model structure with optimised parameter (Fig. 53a, bottom right) compared to $r^2=0.35$ and RMSE=0.91 obtained from simulation using optimised model structure with default parameters (Fig. 53a, middle right) and $r^2=0.28$ and RMSE=1.0 obtained from simulation using reference model structure with default parameters (Fig. 53a, top right). Generally, the magnitude and timing of NO₃-N concentrations was reasonably predicted during the first half of simulation only. The SALMO-PLUS simulation using optimised model structure with optimised parameter (Fig. 53a, bottom right) over predicted the NO₃-N concentrations towards the year end of the simulation. However, the magnitude for the over prediction is less compared to the simulation using optimised model structure with default parameters and simulation using reference model structure with default parameters. Simulations results of NO₃-N from other authors were not available for comparison purposes.

**Total phytoplankton simulation**

The Fig. 53b indicates that the simulation of total phytoplankton biomass by SALMO-PLUS in Roodeplaat Dam using optimised model structure with optimised parameters (Fig. 53b, bottom left) gave better results compared to SALMO-PLUS simulation using optimised model structure with default parameter (Fig. 53b, middle left) and simulation using reference model structure with default parameters (Fig. 53b, top left). Even though no significant statistical improvement was observed in terms of $r^2$ and RMSE values, the simulation using optimised model structure with optimised parameters revealed better growth curve with reasonable magnitude. However, the simulation for total phytoplankton was over predicted during summer peak in the middle of the year. The $r^2=0.05$ and RMSE=13.0 obtained from SALMO-PLUS simulation using optimised model structure with optimised parameters does not differ much compared to $r^2=0.03$ and RMSE=15.6 and $r^2=0.06$ and RMSE=14.4 obtained from SALMO-PLUS simulation using reference model structure with default parameter and simulation using reference model structure with default parameters.
The simulation results of SALMO-PLUS for Total phytoplankton and Chlorophyll-a concentrations in Roodeplaat Dam are represented in Fig. 53b.

<table>
<thead>
<tr>
<th>Roodeplaat Dam (South Africa) 2003</th>
<th><strong>Total Phytoplankton (cm$^3$/m$^3$)</strong></th>
<th><strong>Chlorophyll-a concentration (mg/l)</strong></th>
</tr>
</thead>
</table>
| **SALMO-PLUS**
Reference model structure (A1-A2-A3-A4) with default parameters | ![Graph](image1.png) $r^2=0.06$ | ![Graph](image2.png) $r^2=0.057$ |
| **SALMO-PLUS**
Optimised model structure (A1-A2-A3-A4) with default parameters | ![Graph](image3.png) $r^2=0.039$ | ![Graph](image4.png) $r^2=0.039$ |
| **SALMO-PLUS**
Optimised model structure (A1-A2-A3-A4) with optimised parameters | ![Graph](image5.png) $r^2=0.05$ | ![Graph](image6.png) $r^2=0.049$ |

**Figure 57** Simulation results of SALMO-PLUS for Total phytoplankton and Chlorophyll-a concentrations in Roodeplaat Dam (2003) X-axis in days —— = simulated epilimnion; ---- = simulated hypolimnion; o = measured epilimnion
Results from Recknagel et al. (2008b) on phytoplankton simulation for Roodeplaat Dam showed better outcome compared to this study. Recknagel et al. (2008b) obtained an $r^2=0.13$ and RMSE=10.64 with better magnitude and timing of phytoplankton growth. The best results for phytoplankton growth simulation in Roodeplaat Dam was obtained by Cetin (2007) with an $r^2=0.27$ and RMSE=9.48 where phytoplankton predictions is visually and quantitatively very good and describe the measured data very well.

Chlorophyll-a simulation

The Fig. 53b indicates the simulation of chlorophyll-a in Roodeplaat Dam using optimised model structure with optimised parameters (Fig. 53b bottom right) gives satisfactory predictions ($r^2=0.049$, RMSE=33.8) for most of the months except for the over prediction during summer chlorophyll-a peak. Generally, prediction of chlorophyll-a dynamics in Roodeplaat Dam follows closely the phytoplankton dynamic fluctuations as expected. The SALMO-PLUS simulation using optimised model structure with optimised parameters does not differ much in term of statistical values compared to the simulation using reference model structure with default parameter (Fig. 53b, top right) ($r^2=0.05$, RMSE=36.0) and simulation using optimised model structure with default parameters (Fig. 53b, middle left) ($r^2=0.039$, RMSE=39.1), however, the SALMO-PLUS simulation using optimised model structure with optimised parameters adequately predicted the magnitude of chlorophyll-a measured data from Roodeplaat Dam. There are no results on chlorophyll-a available from other authors for comparison purposes.

Structure and parameters optimisations are performed on a consistent basis during the experiment. In nearly all experiment, specifying an optimisation percentage/level was found to improve results performance. However, higher levels of optimisation bound to increase compilation time even though the results may not be significantly different from one experiment to another. For most cases, level of less than 50% optimisation is a good balance between optimisations time and performance gain. Nevertheless, structure and parameters optimisation performed on data from Roodeplaat Dam data does not improve the results significantly. This could be due to the coding not providing the compiler with opportunities for further performance improvements. The next step would then be to analyse and restructure the program at the source code level to achieve better performance.
4.8 Warm Monomictic and Mesotrophic (South Para Reservoir, South Australia)

4.8.1 Simulation results by Means of the Simulation SALMO-PLUS for South Para Reservoir, South Australia

South Para Reservoir is part of water storages of the South Para Catchment System, located 60 km north-east of Adelaide, South Australia. South Para reservoir exhibits strong thermal and chemical stratification throughout summer (Schoofs, 1996). Data for South Para Reservoir for this experiment using SALMO-PLUS starts from early July 2008 until end of June 2008. Stratification begins from middle of November after the spring turn-over which happens over September for this southern hemisphere lake.

Model Structure Optimisation

After the goal function (1) of particle swarm optimisation method has been applied to the state variables PO4-P, total phytoplankton and zooplankton for which measured data of South Para Reservoir were available, the following optimal model structure has been identified by SALMO-PLUS (in bold):

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>-</td>
<td>B2</td>
<td>-</td>
<td>-</td>
<td>B4</td>
</tr>
<tr>
<td>Hongping and Jianyi (2002)</td>
<td>-</td>
<td>-</td>
<td>C3</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Model Structure and Parameter Optimisation

Multi objective parameter optimisation using particle swarm optimisation method was then conducted on selected parameter for South Para Reservoir. Table 23 summarises the reference and interactively fine-tuned initial values for state variables achieved by SALMO-PLUS for South Para Reservoir.
Table 23 Reference and interactively fine-tuned initial values for state variables of South Para Reservoir for SALMO-PLUS simulation

<table>
<thead>
<tr>
<th>State variables</th>
<th>Reference Initial State Variable Values</th>
<th>Initial State Variable Values interactively fine-tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO₄-P</td>
<td>10.0 μg/l</td>
<td>5.0 μg/l</td>
</tr>
<tr>
<td>NO₃-N</td>
<td>3 mg/l</td>
<td>0.039 mg/l</td>
</tr>
<tr>
<td>Diatoms</td>
<td>6.5 cm³/m³</td>
<td>0.3 cm³/m³</td>
</tr>
<tr>
<td>Green Algae</td>
<td>5.0 cm³/m³</td>
<td>12.9 cm³/m³</td>
</tr>
<tr>
<td>Blue-green Algae</td>
<td>5.0 cm³/m³</td>
<td>0.01 cm³/m³</td>
</tr>
<tr>
<td>Zooplankton</td>
<td>5.0 cm³/m³</td>
<td>1.5 cm³/m³</td>
</tr>
</tbody>
</table>

Tab. 24 summarises the initial and optimised parameter values achieved by SALMO-PLUS for South Para Reservoir. The corresponding simulation results are represented in Fig. 54a and Fig. 54b.

**PO₄-P simulation**

The results from South Para Reservoir simulation by SALMO-PLUS using optimised model structure and optimised parameters (Fig. 54a, bottom left) produced a poor result with regards to prediction of PO₄-P state variable. Even though the $r^2=0.0015$ obtained was better than $r^2=0.000003$ from simulation by reference model structure and default parameter (Fig. 54a, top left), the PO₄-P fluctuations was poorly simulated.

**O₂ simulation**

In contrast, little improvement was observed in terms of statistical result for O₂ prediction between reference model structure with default parameters (Fig. 54a, top right) and optimised model structure with optimised parameters (Fig. 54a, bottom right). The statistical value improvement was small as reflected by the $r^2$ value of 0.873 and RMSE value of 2.1 from the optimised model structure and optimised parameters simulation. This is compared to $r^2=0.842$ and RMSE=1.8 obtained for the simulation of SALMO-PLUS using reference model structure with default parameters (Fig. 54a, top right). The magnitude for O₂ dynamic was well matched by SALMO-PLUS simulation in both cases.
Table 24 Initial values for parameters and optimised parameter values achieved by SALMO-PLUS for South Para Reservoir (X1=diatoms, X2=green algae, X3=blue-green; STW=winter, VZF=spring, STS=summer, VZH=autumn)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Parameter Values</th>
<th>Parameter Values optimised by SALMO-PLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>STW</td>
</tr>
<tr>
<td>KP</td>
<td>X1=18</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>X2=12.8</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>X3=21.6</td>
<td>32.4</td>
</tr>
<tr>
<td>YX</td>
<td>X1=0.8</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>X2=0.328</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>X3=1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>KI</td>
<td>X1=29</td>
<td>43.5</td>
</tr>
<tr>
<td></td>
<td>X2=29</td>
<td>43.5</td>
</tr>
<tr>
<td></td>
<td>X3=28</td>
<td>14</td>
</tr>
<tr>
<td>YZP</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>165</td>
</tr>
<tr>
<td>TOPTX</td>
<td>X1=21</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>X2=27.6</td>
<td>41.4</td>
</tr>
<tr>
<td></td>
<td>X3=24</td>
<td>12</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>X1=0.17</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>X2=0.35</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>X3=0</td>
<td>0</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>X1=2.37</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>X2=3.3</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td>X3=2.37</td>
<td>3.55</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.022</td>
</tr>
<tr>
<td>MOT</td>
<td>0.006</td>
<td>0.0073</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>South Para Reservoir (South Australia) 2008</td>
<td><strong>PO₄-P Concentration (µg/l)</strong></td>
<td><strong>O₂ Concentration (mg/l)</strong></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>SALMO- PLUS</strong> Reference model structure (A1-A2-A3-A4) with default parameters</td>
<td><img src="image1.png" alt="Graph" /> ( r^2=0.00003 )</td>
<td><img src="image2.png" alt="Graph" /> ( r^2=0.842 )</td>
</tr>
<tr>
<td><strong>SALMO- PLUS</strong> Optimised model structure (A1-B2-C3-B4) with default parameters</td>
<td><img src="image3.png" alt="Graph" /> ( r^2=0.0015 )</td>
<td><img src="image4.png" alt="Graph" /> ( r^2=0.873 )</td>
</tr>
<tr>
<td><strong>SALMO- PLUS</strong> Optimised model structure (A1-B2-C3-B4) with optimised parameters</td>
<td><img src="image5.png" alt="Graph" /> ( r^2=0.0015 )</td>
<td><img src="image6.png" alt="Graph" /> ( r^2=0.873 )</td>
</tr>
</tbody>
</table>

**Figure 58a** Simulation results of SALMO-PLUS for PO₄-P and O₂ concentrations in South Para Reservoir (2008) X-axis in days. —— = simulated epilimnion; ---- = simulated hypolimnion; o = measured
**Total phytoplankton simulation**

The Fig. 54b indicates that the simulation results for the biomass of total phytoplankton in South Para Reservoir improved after the model structure and parameters were optimised (Fig 54b, bottom left). The simulation follows closely the measured total phytoplankton data. The $r^2=0.607$ obtained from simulation with optimised model structure and optimised parameter was better than $r^2=0.38$ obtained from simulation using reference model structure with default parameter (Fig. 54b, top left). This $r^2$ was similar to $r^2$ from simulation with optimised structure and default parameter (Fig. 54b, middle left) suggesting that the results could be further improved by manual fine tuning. The SALMO-PLUS simulation for all the three exercise (Fig. 54b, left) did not adequately capture the peak during summer season in South Para Reservoir.

**Zooplankton simulation**

The Fig. 54b (right) summarises the simulation results for the zooplankton biomass using optimised model structure and optimised parameter.

A little improvement of the results in term of $r^2$ for zooplankton simulation was achieved when results from reference model structure with default parameters (Fig. 54b, top right) was compared to results from optimised model structure with optimised parameters (Fig. 54b, bottom right). This was reflected by the $r^2$ value of 0.413 and RMSE=0.76 from optimised model structure with optimised parameter simulation (Fig. 54a, bottom right) compared to $r^2=0.318$ and RMSE=0.71 from reference model structure with default parameters (Fig. 54a, top right).

Even though the SALMO-PLUS simulation from optimised model and parameter does produce better $r^2$, the timing and magnitude of zooplankton simulation shown does not fits well with the measured data for South Para Reservoir. These results have shown that PSO method needs to be coupled with manual fine-tuning by experts in order to achieve better results.
Figure 59 Simulation results of SALMO-PLUS for Total phytoplankton and zooplankton in South Para Reservoir (2008). X-axis in days ——— = simulated epilimnion; ---- = simulated hypolimnion; o = measured.
4.9 Extension of the SALMO-OO Simulation Library by Additional Process Model

In response to the needs to investigate the complex ecosystem dynamics that these lakes possess, this research has proposed to utilise the SALMO-OO model. The incorporation of new process model from Law et al. (2009) is intended to add another alternative of process model to the current library of SALMO-OO. It is postulated that this new process model can improve prediction especially on the tropical lake ecosystem processes regarding algal succession and interaction between zooplankton and different algal groups.

The new process model from Law et al. (2009) containing algal growth, algal grazing, zooplankton growth and zooplankton mortality functions are now available in the SALMO-OO library and are ready for selection as shown in Figure 55. The present five different model structures available in the SALMO-OO library each for algal growth, algal grazing, zooplankton growth and zooplankton mortality and the new process model from Law et al. (2009) (given symbol E) are listed in Table 25.

**Table 25** List of present process model structure in SALMO-OO showing addition of Law et al. (2009) (symbol as model E)

<table>
<thead>
<tr>
<th>Source</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
</tr>
<tr>
<td>Hongping and Jianyi (2002)</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
</tr>
<tr>
<td><strong>Law et al. (2009)</strong></td>
<td><strong>E1</strong></td>
<td><strong>E2</strong></td>
<td><strong>E3</strong></td>
<td><strong>E4</strong></td>
</tr>
</tbody>
</table>

A new category describing the trophic and mixing status of Lake Kenyir and Lake Penang known as ‘mesotrophic tropical-stratified’ has also been created on the selection page of the model SALMO-OO (see section 3.4, Fig. 10). The latest algal and zooplankton process models in SALMO-OO with the inclusion of Law et al. (2009) are listed in Table 26. As shown in Table 26, the SALMO-OO simulation library now contains five alternative process models available for selection. This research has also contributed to a centralised documentation for library of 1) algal photosynthesis, respiration and grazing and 2) zooplankton growth and mortality in SALMO-OO model.

The input sign and parameters notation from the Law et al. (2009) process model which have similar meanings and definitions with SALMO-OO model were standardised and put into uniform naming conforming to the SALMO-OO standard and convention. No new parameters which are not found in SALMO-OO were observed in the new process model from Law et al. (2009). Parameters from Law et al. (2009) with same meaning but hold
different values when compared to SALMO-OO parameters are listed in Table 27 with their original values retained during this experiment.

Figure 60 Snapshot from SALMO-OO model selection page showing the new process model from Law et al. (2009), (symbol as model number 2) and the new category for mesotrophic tropical-stratified lake.

Clearly, the calibration and application for a single category of lake is not sufficient to establish the applicability of this new model having new value of water quality related parameters. It is necessary to use the new process model for other lakes with different category as well and check the consistency of the parameter values used for the calibration. Therefore, another datasets representing two separate water bodies which are Saidenbach Reservoir and Roodeplaat Dam with different trophic conditions were used to further validate and test the feasibility of the new process model extension to describe trophic status of respective study lakes by looking at the prediction of the algal growth, grazing and zooplankton growth and mortality respectively. Brief description of these study lakes were given in the previous section respectively.
<table>
<thead>
<tr>
<th>Table 26 SALMO-OO algal and zooplankton processes formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benndorf and Recknagel (1982)</strong></td>
</tr>
<tr>
<td><strong>Park et al. (1974)</strong></td>
</tr>
<tr>
<td><strong>Hongping and Jianyi (2002)</strong></td>
</tr>
<tr>
<td><strong>Arhonditsis and Brett (2005)</strong></td>
</tr>
<tr>
<td><strong>Law et al. (2009)</strong></td>
</tr>
</tbody>
</table>

**Algal growth**

AGRO\[i\] = \(PHO[i]-RA[i]\)

AGRO\[j\] = \(PHO[j]-RA[j]\)

AGRO\[i\] = \(PHO[i]-RA[i]\)

AGRO\[i\] = \(PHO[i]-RA[i]\)

AGRO\[i\] = \(PHO[i]-RA[i]\)

**Algal photosynthesis**

PHO\[i\] = ((PHOMAX\[i\] - PHOMIN\[i\]) * (IREDZ / (KZI + IREDZ)) * PHONP\[i\];

PHO\[i\] = PHOMAX - PHOT\[i\]

PHO\[i\] = PHOMAX - PHOT\[i\]

PHO\[i\] = PHOMAX * \((2.718*FP/(EPS*ZMIX)) \times \exp(-I/(IS*FP))\) 

PHO\[i\] = PHOMAX * \((2.718*FP/(EPS*ZMIX)) \times \exp(-I/(IS*FP))\) 

PHO\[i\] = PHOMAX * \((2.718*FP/(EPS*ZMIX)) \times \exp(-I/(IS*FP))\) 

PHO\[i\] = PHOMAX * \((2.718*FP/(EPS*ZMIX)) \times \exp(-I/(IS*FP))\) 

PHO\[i\] = PHOMAX * \((2.718*FP/(EPS*ZMIX)) \times \exp(-I/(IS*FP))\) 

**Algal respiration**

RA\[i\] = (RATOPT - RATMIN) / TOPTA, *T + RATMIN + 0.3 * PHO\[i\]

RA\[i\] = RATOPT, *PHOT\[i\], PHOMAX

RA\[i\] = RATMIN, *exp(0.038 * T)

RA\[i\] = RATOPT, *exp(0.07 * (T - TOPTA))

RA\[i\] = RATOPT, *exp(-ktz(T - Tref))\)
<table>
<thead>
<tr>
<th>Algal grazing</th>
<th>AGRA[i] = (GMAX - GMIN)<em>(exp(- R</em>abs(ln(T/TOPTZ)))+GMIN)*((A[i]*PFAi)/Z)/(KAG/KZ+(A[i]*PFAi)/KZ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZGRO = (AZMAX - AZMIN)/GMAX*G_T,E,H * Σ (G[i] * T_E,H * UAZ, RZ_T,E,H)</td>
<td></td>
</tr>
<tr>
<td>ZGRO = (ΣAGRA<em>Fe) / (RZMIN+0.2</em>((Σ(PF,*Ai)) / (KAG+Σ(PF,*Ai))) * GT)</td>
<td></td>
</tr>
<tr>
<td>ZGRO = (AZMAX<em>Σ(PAF,<em>Ai)-FMIN)/(KAG+Σ(PAF,<em>Ai)-FMIN)+RZOPT</em>exp(1.066</em>T)+0.04</em>AGRA)</td>
<td></td>
</tr>
<tr>
<td>ZMO = (MOMIN+MOT*T_T,E,H) * Z_T,E,H / (KMO+Z_T,E,H)</td>
<td></td>
</tr>
<tr>
<td>ZMO = MO</td>
<td></td>
</tr>
<tr>
<td>ZMO = (MORTZ*Z'') / (KMO''+Z'')</td>
<td></td>
</tr>
</tbody>
</table>

| Zooplankton growth | AGRA[i] = GMAX*((PFA,*A[i])/(KAG+(PFA,*A[i])))*GT; |
| ZGRO = (ΣAGRA*Fe) / (RZMIN+0.2*((Σ(PF,*Ai)) / (KAG+Σ(PF,*Ai))) * GT) |
| ZGRO = (ΣAGRA*Fe) / (RZMIN+0.2*((Σ(PF,*Ai)) / (KAG+Σ(PF,*Ai))) * GT) |
| ZGRO = (ΣAGRA*Fe) / (RZMIN+0.2*((Σ(PF,*Ai)) / (KAG+Σ(PF,*Ai))) * GT) |

| Zooplankton mortality | ZGRO = (GMAX*Fefᵢ)((Ai*P/Cphyto)² + ω * DETᵢ²)*exp(-kₜz(Tᵢ-Tᵢ_ref)²) |
| ZGRO = (GMAX*Fefᵢ)((Ai*P/Cphyto)² + ω * DETᵢ²)*exp(-kₜz(Tᵢ-Tᵢ_ref)²) |
| ZGRO = (GMAX*Fefᵢ)((Ai*P/Cphyto)² + ω * DETᵢ²)*exp(-kₜz(Tᵢ-Tᵢ_ref)²) |
| ZMO = MORTZ*((Z') / (KMO''+Z'')) |

162
Table 27 List of parameter definitions and unit specific to the new process model with their original values retained for this experiment

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition and units</th>
<th>BG</th>
<th>G</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>KB</td>
<td>Background light extinction coefficient, m^{-1}</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>RO</td>
<td>Constant phytoplankton respiration rate, d^{-1}</td>
<td>0.08</td>
<td>0.08</td>
<td>0.1</td>
</tr>
<tr>
<td>KL</td>
<td>Light extinction coefficient due to Chl-a, m^{-2}mg^{-1}</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Gmax</td>
<td>Maximum phytoplankton growth rate, d^{-1}</td>
<td>1.2</td>
<td>1.8</td>
<td>2.2</td>
</tr>
<tr>
<td>KPi</td>
<td>Half saturation constant for PO_4-P uptake by algae, mg/m^3</td>
<td>18</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>KT</td>
<td>Effect of temperature on phytoplankton processes, °C^{-2}</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>KTZ</td>
<td>Effect of temperature on zooplankton processes, °C^{-2}</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>P/Cphyto</td>
<td>Phosphorus to carbon ratio for phytoplankton, mgPmgC^{-1}</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The approach adopted for testing new process model is as follows;

1) Best model structure for each study lake was first obtained by means of structure optimisation and multi objective parameter optimisation as described in Section 4.1.

2) The six water bodies’ dataset was subjected to gradual change in algal growth, algal grazing, zooplankton growth and zooplankton mortality process model respectively from the present best structure until the entire process model contains the new process model of Law et al. (2009). Simulations runs were performed for each substitution of the new process model (i.e one simulation run per each substitution). An example of the matrix used to perform this test on Saidenbach Reservoir is shown in Table 28.

Running this new process model does not cause an additional time for SALMO-OO computer simulation and one can finish a single run within few minutes.
The following criteria were used to assess whether this new model library has improved SALMO-OO model simulation:

1. Qualitative consideration was used where appearance of forecast plots against the measured data of phytoplankton, zooplankton and phosphate and nitrate state variable (where data was available) were compared for similarity in trend and magnitude. The timing and duration of peak both in predicted and measured data are also compared for assessment. Seasonality of phytoplankton functional groups needed to be realistic of the trophic state of the water body modelled.

2. Quantitative analysis of simulation results for phytoplankton, zooplankton (where data was available) and phosphate state variables.
   - The square of correlation coefficient ($r^2$) of a linear regression between predicted and measured data
   - The root mean square error (RMSE)

3. Based on the qualitative and quantitative results, each simulation run was ranked based on the above criteria. Only the best result is shown and discussed in the following section
4.9.1 Mesotrophic Tropical-stratified (Lake Kenyir)

After the goal function (1) has been applied to the state variables PO$_4$-P and NO$_3$-N for which measured data of Lake Kenyir were available, the following optimal model structure has been identified:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
<td></td>
</tr>
</tbody>
</table>

The above optimal model structure was gradually replaced with the Law et al. (2009) process model proposed for this research following the matrix as suggested in Table 28. Experiment with the new process models for Lake Kenyir resulted in the following best structure combination for further assessment:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>-</td>
<td>E4</td>
</tr>
<tr>
<td>Law et al. (2009)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>E4</td>
<td></td>
</tr>
</tbody>
</table>

The combinations of Law et al. (2009) zooplankton mortality function and Benndorf and Recknagel (1982) for other process functions as shown above were selected as the possible best structure for Lake Kenyir and later were subjected for further analysis. The simulation results with the new process models by SALMO-OO for PO$_4$-P and NO$_3$-N, in Lake Kenyir are represented in Fig. 56. Only state variables with measured data (PO$_4$-P and NO$_3$-N concentrations), total phytoplankton biomass and algal functional groups are discussed in the following paragraphs.

The SALMO-OO model simulation results for Lake Kenyir (Fig. 56) based on the above structure combination produce a reasonable fit of the measured data for NO$_3$-N concentration ($r^2$=0.04 and RMSE=6.41) (see Fig 56, top right). The simulation for PO$_4$-P state variable produce output trajectories with a reasonable fit to the measured data ($r^2$=0.04 and RMSE=203) (Fig 56, top left). The NO$_3$-N concentration was simulation within the reasonable magnitude compared to measured data even though the NO$_3$-N peak in Lake Kenyir during month of March-April was poorly matched. However, the simulation of PO$_4$-P was over predicted starting from April until November of the simulation year.
Figure 61 Lake Kenyir (1992) simulation results from the SALMO-OO optimised model structure and combination with zooplankton mortality process model from Law et al. (2009) for PO₄-P and NO₃-N concentrations. X-axis in days; ── = simulated epilimnion; ---- = simulated hypolimnion, o = measured data. For algal functional groups: : = Diatoms; - - - = Green Algae; x = Blue-green Algae
Simulation of algal functional group dynamic shows a dynamic dominant between diatoms (X1), green algae (X2) and blue-green algae (X3) (see Figure 56, bottom right). The dominance of diatoms and green algae was observed during January to June followed by periods of blue-green algae dominance from July until November. On comparison with the simulation result on algal functional group for Lake Kenyir from section 4.5.1, blue-green algae were simulated to be prevalence in contrast with the prediction of lesser dominant influence of blue-green algae by SALMO-PLUS. The green algae dominance was similar with prediction in section 4.5.1. The simulation of diatoms dominance from January until June from this experiment with new process model was in contrast with the dominance from July until November (see section 4.5.1, Fig 50b, bottom right).

The simulation results for Lake Kenyir by SALMO-OO with new process model for zooplankton mortality describe the total phytoplankton biomass reasonably well. The model does predict the highest abundance of algae in early February with another major peak simulated both during end of March to early April period and during August. The lowest abundance of total algae was simulated during the end of February. Overall, this combination of Law et al. (2009) zooplankton mortality function and Benndorf and Recknagel (1982) for other process functions does not improved simulation of state variables in Lake Kenyir.
4.9.2 Mesotrophic Tropical-stratified (Lake Penang)

After the goal function (1) has been applied to the state variables input data for which measured data of Lake Penang were available, the following optimal model structure has been identified:

<table>
<thead>
<tr>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
</tbody>
</table>

The above optimal model structure was gradually replaced with the Law et al. (2009) process model proposed for this research following the matrix as suggested in Table 28. Experiment with the new process models for Lake Penang resulted in the following best structure for further assessment:

<table>
<thead>
<tr>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982))</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>A4</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>-</td>
<td>-</td>
<td>B3</td>
<td>-</td>
</tr>
<tr>
<td>Law et al. (2009)</td>
<td>-</td>
<td>E2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td>D1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The combinations of Law et al. (2009) model for algal grazing function, Arhonditsis and Brett (2005) model for algal growth, Park et al. (1974) model for zooplankton growth and Benndorf and Recknagel (1982) model for zooplankton mortality as shown above were selected as the possible best structure for Lake Penang and were subjected for further analysis. Only state variables with measured data (PO4-P and chlorophyll-a), total phytoplankton biomass and algal functional groups are discussed in the following paragraphs. The simulation results in combination with the new process models by SALMO-OO for PO4-P and Chlorophyll-a in Lake Penang are represented in Fig. 57.

The simulation results by SALMO-OO using the process model of Law et al. (2009) algal grazing function for Lake Penang describe the PO4-P very well (Fig. 4.9, top left) with the successful prediction in timing and magnitude. The simulation with Law et al. (2009) algal grazing function gives reasonable prediction with the simulation results by SALMO-OO using the process model of Law et al. (2009) algal grazing function for Lake Penang describe the PO4-P very well (Fig. 4.9, top left) with the successful prediction in timing and magnitude. The simulation with Law et al. (2009) algal grazing function gives reasonable prediction with \( r^2 = 0.40 \) and RMSE = 11.7 (Fig 57, top left). The major PO4-P peaks in September and April were
also well described by the simulation trajectory. The simulation results by SALMO-OO for Chlorophyll-a at Lake Penang in combination with Law et al. (2009) algal grazing function gives a good $r^2$ value of 0.35 and RMSE=49.3 (see Fig 57, top right). However, visual observation showed the magnitude of simulation was over predicted even though the timing for peaks in Chlorophyll-a fluctuations in Lake Penang was fairly described.

Prediction of total phytoplankton biomass by this process model combination described the peak seasons of algal on Lake Penang during months of August to October (Fig. 57, bottom left). This peak was contributed by blue-green algae dominance as described later in algal functional groups simulation (Fig. 4.9, bottom right). Simulation of algal functional groups dynamic in Lake Penang shows prevailing blue green algae domination (see Fig. 4.9, bottom right) with very minor occurrence of diatoms and green algae. This is in contrast with the simulation of algal functional groups in Lake Penang by reference SALMO-OO model structure from section 4.5.2. The simulation results from section 4.5.2 were describing a mesotrophic condition of Lake Penang while the simulation from this section were more suitable in describing a eutrophic lake, as can be seen from the higher level of chlorophyll-a simulated (see Fig. 57, top right). Thus it can be said that based on these results, the SALMO-OO model combination with the new algal grazing process model from Law et al. (2009) does not show any improvement to the simulation of phytoplankton dynamic in Lake Penang.
<table>
<thead>
<tr>
<th>Lake Penang</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Malaysia)</td>
</tr>
<tr>
<td>2005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>PO₄-P Concentration (µg/l)</strong></th>
<th><strong>Chlorophyll-a Concentration (mg/l)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="PO₄-P graph" /></td>
<td><img src="image" alt="Chlorophyll-a graph" /></td>
</tr>
<tr>
<td>$r^2=0.40$</td>
<td>$r^2=0.35$</td>
</tr>
<tr>
<td>RMSE=11.7</td>
<td>RMSE=49.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Total Phytoplankton (cm³/m³)</strong></th>
<th><strong>Algal functional groups (cm³/m³)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Total phytoplankton graph" /></td>
<td><img src="image" alt="Algal functional groups graph" /></td>
</tr>
</tbody>
</table>

**Figure 62** Simulation results of SALMO-OO for PO₄-P, Chlorophyll-a, Total phytoplankton and algal functional groups in Lake Penang (2005). X-axis in days. —— = simulated epilimnion; ---- = simulated hypolimnion; For algal functional groups simulation, —— = Diatoms; - - - = Green Algae; x = Blue-green Algae
4.9.3 Dimictic and Mesotrophic (Saidenbach Reservoir, Germany)

After the goal function (1) has been applied to the state variables for which measured data of Saidenbach Reservoir were available, the following optimal model structure has been identified (in bold):

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>A3</td>
<td>-</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td></td>
<td>-</td>
<td>D2</td>
<td>-</td>
<td>D4</td>
</tr>
</tbody>
</table>

The above optimal model structure for Saidenbach Reservoir was gradually replaced with the Law et al. (2009) process model proposed for this research following the matrix as suggested in Table 28. Experiment with the new process models for Saidenbach Reservoir resulted in the following best structure for further assessment:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td></td>
<td>-</td>
<td>D2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Law et al. (2009)</td>
<td></td>
<td></td>
<td></td>
<td>E3</td>
<td>E4</td>
</tr>
</tbody>
</table>

The combinations of Law et al. (2009) model for zooplankton growth and mortality function, Arhonditsis and Brett (2005) model for algal grazing and Park et al. (1974) model for algal growth as shown above were selected as the possible best structure for Saidenbach Reservoir and were subjected for further analysis. The simulation results with the new process models by SALMO-OO for PO₄-P, total phytoplankton biomass, zooplankton and algal functional groups concentrations in Saidenbach Reservoir are represented in Fig. 58.

The SALMO-OO simulation results of PO₄-P state variable for Saidenbach Reservoir from the above model combination gives satisfactory predictions ($r^2=0.16$, RMSE=11.5) for most of the year (Fig. 58, top left), except for the over prediction of the second half of the year which is not described by the model simulation. Even though the statistical values from this simulation of PO₄-P were not better than simulation results from section 4.3.1, it was comparable with results from Cetin (2007) study. Cetin (2007) obtained $r^2=0.03$ and RMSE=5.78 for PO₄-P simulation in Saidenbach Reservoir from combinations of Park et al. (1974) and Arhonditsis and

The simulation results from the above model combination give a good result for the prediction of total phytoplankton biomass in Saidenbach Reservoir with $r^2=0.4$ and RMSE=2.98 (Fig. 58, top right). The simulation of timing and magnitude of total phytoplankton biomass during the spring peak also fits more closely to the measured data of Saidenbach Reservoir. The simulation results from the above model combination also predicted the decrease in phytoplankton biomass during the end of the simulation reasonably well. However, the simulation of summer peak for phytoplankton biomass of Saidenbach Reservoir is over predicted by this model combination. This simulation result is comparable with the result obtained by SALMO-PLUS using particle swarm optimisation ($r^2=0.30$ and RMSE=2.3) (see section 4.3.1) and with the result from study by Cetin (2007) ($r^2=0.23$ and RMSE=1.7).

The simulation results for Saidenbach Reservoir using the above model combinations describe the zooplankton biomass moderately with $r^2$ value of 0.24 and RMSE=2.6 (Fig. 58, bottom left). However, the three extreme summer data point during and the timing of summer peak was inadequately predicted by this simulation. The SALMO-OO simulation results adopting Law et al. (2009) zooplankton process model functions was comparable with simulation result from SALMO-PLUS (section 4.3.1). The SALMO-PLUS by means of swarm optimisation obtained $r^2=0.31$ and RMSE=2.3. However, visual observation and statistically wise, the zooplankton biomass result from study by Cetin (2007) with $r^2=0.30$ and RMSE=3.09 is better compared to results from this experiment. Even though no improvement has been observed by the application of the new zooplankton growth and grazing models to the SALMO-OO simulation library, the zooplankton result has shown that this new process is comparable with the present similar model functions in the library.

The simulation results for Saidenbach Reservoir describe the algal functional group dynamics in consistent with mesotrophic conditions lake of Saidenbach Reservoir with a balanced abundance of each functional groups contributing to the total biomass present (Fig. 58, bottom right). A clear dominance of green algae during spring and early summer was observed, with very low occurrence of blue-green algae throughout the year. The major peak of green algae was simulated in summer while a minor peak in spring was followed by the succession of diatoms in early spring and late summer. However, a situation where no high occurrence of blue-green algae observed during any time of the year is more common for an oligotrophic lake rather than a mesotrophic lake. Even for mesotrophic conditions, there should be some blue-green algae biomass available during summer (Cetin, 2007) and this is not given by this combination. It should be noted that this experiment with Law et al. (2009) zooplankton growth and grazing models was conducted without parameter optimisation and therefore stand a chance for further improvement. These model combinations have the potential for improving results of phosphate simulation even though the zooplankton process model prediction has not been improved. In terms of the model combination best suited to improve SALMO-OO simulation for Saidenbach Reservoir, the Law et al. (2009) zooplankton process model is not among the best option.
<table>
<thead>
<tr>
<th>Saidenbach Reservoir (Germany) 1975</th>
<th>PO$_4$-P Concentration(µg/l)</th>
<th>Total Phytoplankton (cm$^3$/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1.png" alt="Graph" /> $r^2=0.16$ RMSE=11.5</td>
<td><img src="image2.png" alt="Graph" /> $r^2=0.4$ RMSE=3.0</td>
</tr>
<tr>
<td>Zooplankton (cm$^3$/m$^3$)</td>
<td><img src="image3.png" alt="Graph" /> $r^2=0.24$ RMSE=2.6</td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>Algal functional groups (cm$^3$/m$^3$)</td>
<td><img src="image5.png" alt="Graph" /></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 63** Simulation results of SALMO-OO for PO$_4$- total phytoplankton biomass, zooplankton and algal functional groups in Saidenbach Reservoir (1975). X-axis in days; –– = simulated epilimnion; o = measured data; For algal functional groups simulation, —— =Diatoms; - - - = Green Algae; x = Blue-green Algae.
4.9.4 Warm Monomictic and Hypertrophic (Roodeplaat Dam, South Africa)

After the goal function (1) of particle swarm optimisation method has been applied to the state variables for which measured data of Roodeplaat Dam were available, the following optimal model structure has been identified:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td></td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The optimised structure for Roodeplaat Dam was found to be similar with the reference SALMO-OO model structure. The above optimal model structure for Roodeplaat Dam was gradually replaced with the Law et al. (2009) model proposed in this research following the matrix as suggested in Table 28. Experiment with the Law et al. (2009) process models resulted in the following best structure for further assessment:

<table>
<thead>
<tr>
<th>Author</th>
<th>Process models</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td></td>
<td>A1</td>
<td>A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Law et al. (2009)</td>
<td></td>
<td></td>
<td></td>
<td>E3</td>
<td>E4</td>
</tr>
</tbody>
</table>

The combinations of Law et al. (2009) model for zooplankton growth and mortality functions with Benndorf and Recknagel (1982) model for algal growth and grazing function as shown above were selected as the possible best structure for Roodeplaat Dam and were subjected for further analysis. Only state variables with measured data (PO₄-P, NO₃-N and Chlorophyll-a), total phytoplankton biomass and algal functional groups are discussed in the following paragraphs. Fig. 59a and Fig. 59b illustrates the simulation results for the above combination on Roodeplaat Dam.

The SALMO-OO simulation results of PO₄-P state variable for Roodeplaat Dam from the above model combination gives satisfactory predictions with $r^2=0.74$ and RMSE=146 (Fig. 59a, top left). The dynamic and magnitude of PO₄-P concentrations trend was well simulated although the magnitude of the summer peak was over predicted. The result for PO₄-P concentrations from this experiment is comparable with the results reported by Cetin (2007) ($r^2=0.28$ and RMSE=101) and results from section 4.7.1 of this research ($r^2=0.31$ and RMSE=128).
The simulation of NO$_3$-N concentration for Roodeplaat Dam yield a reasonable result with $r^2=0.25$ and RMSE 1.08, although the simulation during the second half of the year was over predicted in terms of magnitude (Fig. 59a, top right). The $r^2$ obtained from the above model combinations is comparable to $r^2=0.35$ and RMSE=0.6 obtained from the SALMO-OO PSO experiment (section 4.7.1).

The SALMO-OO simulation from the above model combination for total phytoplankton biomass in Roodeplaat Dam yields moderate $r^2$ value (0.27) and RMSE=38.8 (Fig. 59a, bottom left). However, the magnitude of spring phytoplankton peak is over estimated even though the summer highest measured data was well captured by the simulation. In contrast, results from Cetin (2007) for total phytoplankton biomass in Roodeplaat Dam showed better visual closeness between measured and predicted ($r^2=0.27$ and RMSE=9.48) even though the statistical results is comparable with the results from this experiment.

The simulation of chlorophyll-a concentration yields a reasonable result with $r^2=0.25$ and RMSE 97 (Fig. 59a, bottom right) and follows closely the dynamic of total phytoplankton biomass as expected although the simulation during the second half of the year was over predicted.

The SALMO-OO simulation from the model combination for diatoms concentrations in Roodeplaat Dam yields moderate $r^2$ value (0.27) and RMSE=2.0 (Fig. 59b, top left). The simulation did not sufficiently capture the highest peak of diatoms during winter and fail to simulate the diatoms dynamic during summer. Green algae simulation was also poorly reflected by the model combinations with an over prediction of green algae magnitude during summer (Fig. 59b, top right) even though a sensible value of $r^2=0.64$ and RMSE=22.3 was obtained. Blue-green algae simulation was moderately simulated with $r^2=0.10$ and RMSE=19.7 obtained from the SALMO-OO model and proposed model combinations (Fig. 59b, bottom left). However, the autumn magnitude of blue-green algae peak was over predicted by the model simulation.

The algal functional groups abundances simulated by the model combination proposed in this experiment reflect a mesotrophic to eutrophic system where a high abundance of green algae and blue green algae throughout the year was apparent (Fig. 59b, bottom right). Except for the model overestimation of green algae abundances during spring and early summer and underestimation of diatoms in summer, this model combination gave predictions of realistic hypertrophic conditions with complete dominance of blue-green algae during most of the year and very little abundances of green algae and diatoms. The model combinations of Law et al. (2009) for zooplankton growth and mortality functions with Benndorf and Recknagel (1982) for algal growth and grazing function does not improved the interaction between zooplankton and phytoplankton for Roodeplaat Dam.
Figure 64a Simulation results of SALMO-OO for PO$_4$-P and NO$_3$-N concentrations in Roodeplaat Dam (1992). X-axis in days, —— = simulated epilimnion; ---- = simulated hypolimnion, o = measured data.
Roodeplaat Dam (South Africa) 2003

<table>
<thead>
<tr>
<th>Roodeplaat Dam (South Africa) 2003</th>
<th>Diatoms (cm$^3$/m$^3$)</th>
<th>Green algae (cm$^3$/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Diatoms Graph" /></td>
<td>$r^2=0.27$ RMSE=2.0</td>
<td><img src="image2.png" alt="Green algae Graph" /> $r^2=0.64$ RMSE=22.3</td>
</tr>
<tr>
<td><img src="image3.png" alt="Blue-green Algae Graph" /></td>
<td>$r^2=0.1$ RMSE=19.7</td>
<td></td>
</tr>
<tr>
<td><img src="image4.png" alt="Algal functional dynamic Graph" /></td>
<td>$r^2=0.0$ RMSE=19.7</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 65** Simulation results of SALMO-OO for algal functional groups in Roodeplaat Dam (2003). X-axis in days, ——— = simulated epilimnion; o = measured data. For algal functional groups simulation (bottom, right), ——— = Diatoms; - - - = Green Algae; x = Blue-green Algae.
CHAPTER 5

5.0 DISCUSSION AND CONCLUSIONS

5.1 General Discussion

This research aimed to make an original contribution to the knowledge domain Ecology of Tropical Lakes by successfully applying both process based models (SALMO-OO) and data driven models (HEA) for the first time to tropical lake ecosystem. Outcomes of this research extend the applicability of the model SALMO-OO to the new category of ‘mesotrophic tropical-stratified lakes’. Despite the limitations of available data it has been demonstrated by this research that the SALMO-OO model allows to simulate the nutrient cycling and food web dynamics of the Malaysian lakes Penang and Kenyir with reasonable accuracy. Validated simulation results achieved by this study allowed determining the water quality status and the level of eutrophication of the two lakes.

This research has also contributed to the testing of SALMO-PLUS that supports multi-objective parameter optimisation based on the integration of particle swarm optimisation (PSO) and SALMO-OO, and the promotion of benefits of process-based and data-driven modelling techniques for the analysis and synthesis of complex water quality data of tropical lakes.

The following research questions have been addressed by this research project:

1. Can the models SALMO-PLUS and HEA be validated for tropical lakes with different morphometry and trophic state with regards to food web dynamics, nutrient cycles and phytoplankton growth?

   In the context of this research, HEA and SALMO-PLUS have been validated for three tropical lakes with different morphometry and trophic state regarding to food web dynamics, nutrient cycles and phytoplankton growth. SALMO-PLUS has been validated for the simulation of phosphate and nitrate concentrations, phytoplankton biomass, chlorophyll-a concentrations and algal functional group dynamics of the lakes Penang and Kenyir.

   HEA has been applied to develop forecasting models for chlorophyll-a and algal biovolume concentrations in the lakes Penang and Kenyir, and to develop and validate a forecasting model for Chl-a concentrations in the mesotrophic shallow-polymictic tropical Malaysian lake Putrajaya.

2. Can a generic model structure of SALMO-PLUS be found for mesotrophic tropical-stratified lakes?

   After a rigorous testing a model structure of SALMO-OO has been identified that proved to be generic for the two mesotrophic and tropical-stratified lakes Penang
and Kenyir. The resulting generic model structure has been successfully validated and its causality for algal dynamic has been explained.

3. Can generic rule-set models be developed by HEA for the prediction of algal growth in tropical lakes with different morphometry and trophic states?

The application of HEA by means of the bootstrap method to merged time-series limnological data of the lakes Penang and Kenyir resulted in rule-sets being predictive for both tropical lakes. Since both lakes represent the same category of mesotrophic and tropical-stratified lakes and their Chl-a and biovolume concentrations can now be forecasted by the same models, the question for future research arises whether the models will also be valid for data of other lakes belonging to this category.

5.2 The SALMO-PLUS Model as a Tool for Tropical Lake Ecosystem Analysis

The SALMO-OO lake simulation library was designed to be generic for a range of different lake conditions such as different trophic state, morphometry or climate conditions. The generality of SALMO-OO for achieving an acceptable level of accuracy for a wide variety of lake conditions has been well documented (Cetin, 2007). Generic models provide an alternative to the development of ad hoc models for each specific ecosystem or organism under study, therefore, reducing the need to build and test new models from scratch.

For the purpose of this research, SALMO-PLUS has been successfully applied to a variety of lakes including Penang and Kenyir. The SALMO-PLUS is a functional extension of SALMO-OO that has been designed and implemented by Martin & Recknagel (2010) to automatically determine the optimum model structure from the model library and perform multi-objective parameter optimisation by means of the particle swarm optimisation (PSO) (Kennedy & Eberhart, 1995). SALMO-PLUS allows identification of both the optimum model structure and optimum parameter setting for the simulation of nutrient concentrations, phytoplankton and zooplankton biomasses. It has been rigorously tested and validated for selected water bodies during this research. Following validation methods were used: 1) visual comparison of the calculated and measured outputs, 2) statistical analysis using RMSE and $r^2$ values.

Results from this research have shown that SALMO-PLUS simulates reasonably well seasonal dynamics of phosphate and nitrate concentrations as well as total phytoplankton biomass in Lake Kenyir and nitrate and chlorophyll-a concentrations in Lake Penang. The simulation of the succession of algal functional groups in both lakes was successful as well.

Table 29 shows successful applications of SALMO-PLUS for the identification of generic model structures for different lake categories including the new category of mesotrophic tropical-stratified lakes.
Table 29: Four lake categories and their best performing model structures for algal processes identified by SALMO-PLUS

<table>
<thead>
<tr>
<th>Lake categories</th>
<th>Lake examples</th>
<th>Best performing structure for algal biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warm monomictic and hypertrophic</td>
<td>Hartbeespoort (South Africa)</td>
<td>AGRO[D]–AGRA[D]</td>
</tr>
<tr>
<td></td>
<td>Klipvoor (South Africa)</td>
<td></td>
</tr>
<tr>
<td>Dimictic and eutrophic</td>
<td>Bautzen (Germany)</td>
<td>AGRO[B]–AGRA[C]</td>
</tr>
<tr>
<td></td>
<td>Arendsee (Germany)</td>
<td></td>
</tr>
<tr>
<td>Dimictic and mesotrophic and tropical-stratified</td>
<td>Saidenbach (Germany)</td>
<td>AGRO[B]–AGRA[D]</td>
</tr>
<tr>
<td></td>
<td>Weida (Germany)</td>
<td></td>
</tr>
<tr>
<td>Mesotrophic and tropical-stratified</td>
<td>Kenyir (Malaysia)</td>
<td>AGRO[A]–AGRA[A]</td>
</tr>
<tr>
<td></td>
<td>Penang (Malaysia)</td>
<td></td>
</tr>
</tbody>
</table>

Where:

AGRO[i]D = algal growth of the mixed water body/epilimnion/hypolimnion (g wet weight m⁻³ day⁻¹)

AGRA[i]D = algal grazing of the mixed lake, epi- or hypolimnion by herbivorous zooplankton (g wet weight m⁻³ day⁻¹)

5.2.1 Parameter Optimisation by SALMO-PLUS

Martin & Recknagel (2010) has integrated the PSO method into SALMO-PLUS to perform multi-objective parameter optimisation and improve the models simulation accuracy. SALMO-PLUS aims to overcome following limitations of models with conventional parameter identification and calibration:

- fixed and rigid where models can barely reflect the changes of ecological processes according to the prevailing conditions of the lake ecosystem
- calibrations of these parameters are often difficult since a huge number of uncertain parameters need to be simultaneously tested within a wide range of possible values.

The following procedure for the parameter optimisation with SALMO-PLUS has been conducted: (1) Expert-based settings of parameter values for MOMIN and MOT for lakes Kenyir and Penang by Recknagel (pers. comm.) based on repeated
model simulations for the succession of algal functional groups and zooplankton dynamics. (2) Fine tuning of the pre-set parameter values by using the multi-objective parameter optimisation option of SALMO-PLUS.

The expert-based parameter settings for Lake Kenyir resulted in MOMIN = 0.0075 and MOT = 0.01218 respectively in order to reflect a zooplankton dynamics that correspond with the algal functional groups dynamic in Lake Kenyir. Like in most ecological models, fish influences are not explicitly modelled in SALMO-PLUS. However, the predation pressure of fish is quantified by increasing the death rate of zooplankton. Since there are no seasonal dependence of this rate, a constant predation pressure is realistic for the situation in tropical lakes.

The expert-based parameter settings for Lake Penang resulted in MOMIN = 0.005 and MOT = 0.001 respectively in order to reflect a zooplankton dynamics that correspond with the algal functional groups dynamic in Lake Kenyir. The MOMIN value in Lake Penang was set lower than in Lake Kenyir to reflect its larger community of piscivorous fish compared to Lake Penang. The resulting increase in the zooplankton abundance in Lake Penang was the key to improve simulation results for the algal functional groups.

The SALMO-PLUS default parameter value for TOPTZ is 20°C. The TOPTZ values for Lake Kenyir and Lake Penang obtained after multi objective parameter optimisation is 27.35°C and 29.9°C respectively. This parameter that represent optimal temperature for feeding activity of the zooplankton in both lakes are comparable between each other lake. Experimentation with TOPTZ value in Lake Kenyir did not show any significant changes on phytoplankton biomass simulation by the lowering the TOPTZ value thus, the values obtained from multi objective parameter optimisation is maintain for the simulation in Lake Penang.

Tables 30 and 31 describe parameters value selected for multi objective parameter optimisation and values after optimisation for Lake Kenyir and Lake Penang respectively.
Table 30 Comparison on parameters value selected for multi objective parameter optimisation and values after optimisation for Lake Kenyir (X1=diatoms, X2=green algae, X3=blue-green algae)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Expert-based pre-set parameter values of SALMO-PLUS</th>
<th>Fine-tuning of the pre-set parameter values by multi objective optimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X1</td>
<td>X2</td>
</tr>
<tr>
<td>KP</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>YX</td>
<td>0.8</td>
<td>0.41</td>
</tr>
<tr>
<td>KI</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>YZP</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>TOPTX</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>0.17</td>
<td>0.35</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>2.37</td>
<td>3.3</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>MOT</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>
Table 31 Comparison on parameters value selected for multi objective parameter optimisation and values after optimisation for Lake Penang. (X1=diatoms, X2=green algae, X3=blue-green algae)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Expert-based pre-set parameter values of SALMO-PLUS</th>
<th>Fine-tuning of the pre-set parameter values by multi objective optimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X1</td>
<td>X2</td>
</tr>
<tr>
<td>KP</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>YX</td>
<td>0.8</td>
<td>0.41</td>
</tr>
<tr>
<td>KI</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>YZP</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>YZN</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>TOPTX</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>PHOTXMIN</td>
<td>0.17</td>
<td>0.35</td>
</tr>
<tr>
<td>PHOTXMAX</td>
<td>2.37</td>
<td>3.3</td>
</tr>
<tr>
<td>MOMIN</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>MOT</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>TOPTZ</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

5.2.2 Initial Value Identification

Different initial values were tested for during each of the simulation trial run. Changing value of one variable can cause another to increase or decrease as described by the process equation in the model. Facts about the nutrient availability in the tropical lake ecosystem and observation on the measured data available were used to approximate the initial value for running the SALMO-PLUS model simulation. Another important consideration is the well-balanced simulation between initial/startiong data point and ends point of simulation. The SALMO-PLUS was feed with the initial values for each state variable for 360 days simulation. The initial
value for state variables used for running the simulation for both Lake Kenyir and Lake Penang are presented in Table 32.

The initial values for state variables of PO\textsubscript{4}-P and NO\textsubscript{3}-N were set lower in Lake Penang compared to Lake Kenyir (see Tab. 5.3). This is to reflect better the nature of nutrient rich hypolimnion in Lake Kenyir that provides nutrient supply to the epilimnion during pre-monsoon seasons. These nutrients were originated partly from submerged rainforest during the construction of the hydroelectric project in the 1980s. No macrophytes were submerged during the formation of Lake Penang and sources of high nutrient were unlikely to inflow into the water body thus, the value for PO\textsubscript{4}-P and NO\textsubscript{3}-N state variables were set lower.

In contrast, the zooplankton initial value was set higher in Lake Penang compared to Lake Kenyir. As mentioned in section 5.2.1, the MOMIN value was set lower in Lake Penang than Lake Kenyir to reflect the impact of piscivorous fishes available in greater number in Lake Penang. Therefore, the higher zooplankton initial value in Lake Penang serves to emphasise the assumption that less impact of fishes on zooplankton biomass was observed in Lake Penang.

In SALMO-OO, phytoplankton production in the hypolimnion is distinguished by the use of input data from the hypolimnion rather than from the mixed or epilimnion layers. Hypolimnion initial values for state variables, even though were listed as well in Table 32, does not contribute much to the dynamic of this stratified tropical lake. However, hypolimnion values especially for oxygen state variables were made sure to maintain lower than epilimnion values during simulation to reflect the real process in lakes.
Table 32 Initial values for diatoms, green algae and blue-green algae in Lake Kenyir and Lake Penang used for this research

<table>
<thead>
<tr>
<th>State variables</th>
<th>Initial values interactively fine-tuned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lake Kenyir</td>
</tr>
<tr>
<td><strong>PO$_4$-P Epilimnion</strong></td>
<td>60 μg/l</td>
</tr>
<tr>
<td><strong>PO$_4$-P Hypolimnion</strong></td>
<td>120 μg/l</td>
</tr>
<tr>
<td><strong>NO$_3$-N Epilimnion</strong></td>
<td>15.0 mg/l</td>
</tr>
<tr>
<td><strong>NO$_3$-N Hypolimnion</strong></td>
<td>12.0 mg/l</td>
</tr>
<tr>
<td>Diatoms Epilimnion and Hypolimnion</td>
<td>2.0 and 0.2 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Green Algae Epilimnion and Hypolimnion</td>
<td>6.0 and 0.2 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Blue-green Algae Epilimnion and Hypolimnion</td>
<td>2.0 and 0.2 cm$^3$/m$^3$</td>
</tr>
<tr>
<td>Zooplankton Epilimnion and Hypolimnion</td>
<td>3.0 and 0.3 cm$^3$/m$^3$</td>
</tr>
</tbody>
</table>

5.3 Addition of the New Process Model from Law et al. (2009) into SALMO-OO Simulation Library

One of the objectives of this research was to identify another literature model as an extension of the process model library of SALMO-OO, and its testing for a variety of lakes. As a result the model of Law et al. (2009) has been selected and implemented in SALMO-OO because of its use of alternative process equations for algal growth and grazing as well as zooplankton growth and mortality.

The SALMO-OO is the core generic model and the modularisation of the model structure by means of object-oriented design has allowed an efficient inclusion of Law et al. (2009) alternative model structure. Common parameter values and names to both Law et al. (2009) and those from SALMO-OO were maintain according to the overall generic design.

The model structure (i.e. the mathematical function) and constant parameter values for Lake Kenyir and Lake Penang are kept unchanged for each simulation. Therefore, it is the measured input data that distinguishes between the two lakes. In
keeping with the objective, all of the simulation library parameters were kept constant for the simulation of new process model from Law et al. (2009)

Cetin (2007) has developed the SALMO-OO simulation library consisting of three phytoplankton growth and grazing process models that have been adopted from literature models developed for specific lake systems. However, none of these process models was formulated for tropical lake ecosystem. Literature review searching for algal and zooplankton process model that was developed specifically for tropical lake ecosystem was unsuccessful. The same criteria and rational for phytoplankton growth models as adopted by Cetin (2007) were used during this literature review. One reason suggested for this could be lack of modelling research as well as scarce limnological study on tropical lake as elaborated in section 5.8. The literature review however, found a model developed by Law et al. (2009) that after minimal adjustment, was chosen for addition into SALMO-OO process library. Addition of the new process model from Law et al. (2009) has introduces another perspective into the lake category to the current simulation library and has broadened the scope of SALMO-OO.

The framework for testing the new process model from Law et al. (2009) involves optimising the best structure for test lakes and gradually replacing each algal growth, algal grazing, zooplankton growth and zooplankton mortality of the optimised structure with the alternative from the new process model. In order to verify the applicability of the Law et al. (2009) process model in SALMO-OO simulation library, 6 lakes from 5 different trophic state and mixing conditions were selected and tested with Law et al. (2009) from the simulation library to improve the simulation of lakes within each category. Combinations of different process models were tested within the simulation library as given in Table 28. The idea on experimenting with this new process model is to achieve a better approximation of a particular lake’s dynamic compared to the current optimal structure. The simulation results from combinations of optimised model structure with gradual replacement of new process model by Law et al. (2009) were compared quantitatively and qualitatively. Different water bodies possessing various trophic states and mixing conditions were tested and suitable model structures combination were selected from this exercise for further assessment against the present optimum structure found by means of particle swarm optimisation for respective lakes. Table 33 summarises the findings presented in section 4.9 for each lake tested.

Law et al. (2009) has been successfully implemented into SALMO-OO model library. However, results from this experiment shows that even though the model is working, no improvements on the results for water bodies tested were observed for this implementation

5.3.1 Evaluation of SALMO-OO Results

The evaluation of model quality is an essential aspect of the modelling activity in order to know how much confidence one can have in the model results and also in order to choose between alternative models (Wallach & Goffinet, 1989). This
involves both quantitatively and qualitatively validation of the results. Schlesinger et al. (1979) defined validation as meaning ‘substantiation that a computerised model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model. This indicates applicability of model for targeted objectives even though it might not be the best model yet. The criteria and analysis procedure developed by Cetin (2007) for assessing different combinations of process model was adopted for this experiment. The key criteria were based on the realistic simulation of phosphate concentration and reasonable algal functional dynamics reflecting respective lake ecosystem trophic level.

Table 33  Summary of process model combinations selected for further assessment from experiment with lakes of different trophic state and mixing conditions. A=Benndorf and Recknagel (1982); B=Park et al. (1974); C=Hongping and Jianyi (2002); D=Arhonditsis and Brett (2005); E=Law et al. (2009)

<table>
<thead>
<tr>
<th>Trophic state and mixing conditions</th>
<th>Algal growth</th>
<th>Algal grazing</th>
<th>Zooplankton growth</th>
<th>Zooplankton mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesotrophic Tropical-stratified (Lake Kenyir)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>E</td>
</tr>
<tr>
<td>Mesotrophic Tropical-stratified (Lake Penang)</td>
<td>D</td>
<td>E</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Dimictic and mesotrophic (Saidenbach Reservoir)</td>
<td>B</td>
<td>D</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Warm monomictic and hypertrophic (Roodeplaat Dam)</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>Warm monomictic and mesotrophic (South Para Reservoir)</td>
<td>A</td>
<td>B</td>
<td>E</td>
<td>E</td>
</tr>
</tbody>
</table>

Quantitative analysis methods by means of linear regression ($r^2$) and root mean square error (RMSE) were used in combination with the qualitative analysis by means of comparing realistic and logic simulated versus measured data for a given lake during validation purposes. Both $r^2$ and RMSE reflect associated error of predicted against measured data and allow inferences on the reliability of the outputs. Visual appraisal of predictions by expert was adopted to determine whether the model output is able to predict the timing and magnitude of algal growth. Expert knowledge input was used to discern whether simulation results have accurately and reasonably predicted given lake conditions (such as eutrophic or mesotrophic lake status). Cetin (2007) emphasises on the qualitative analysis conforming to current expert knowledge of lake ecosystem dynamics in evaluating simulation results and argues that the use of $r^2$ and RMSE assessment did not always give very informative answers to how well the model was performing and how adequately the model described lake dynamics. The visual assessment gave the following important information which cannot be
obtained by means of $r^2$ and RMSE 1) how well the model performs 2) areas of poor simulation and 3) the improvement of the present against previous output. The same approach was adopted by Atanasova et al. (2011) with procedure for model evaluation includes expert’s visual estimation. The final selection of an optimal model prediction is made by taking $r^2$, RMSE and expert knowledge of lake ecosystem interactions into account.

In term of quantitative statistics, RMSE gives a better assessment than $r^2$ particularly when comparing the results of different model combination. However, it was observed that certain model combination yielded low RMSE values indicating high model performance but do not fulfil the expert’s expectations in terms of a realistic simulation of algal group dynamics. Reasons were the combination output do not adequately simulating peaks in the measured data, the simulation has lower or higher magnitude than measured data. For this research, it was decided that a model output with higher RMSE on the same experiment performs better according to the expert’s opinion.

As shown in Table 33, the zooplankton mortality function from Law et al. (2009) contributed to the potential results improvement for 5 out of 6 lakes trophic state and mixing conditions tested while the zooplankton growth function contributed to the half of the lakes tested. However, the priority in determining an accurate SALMO-OO simulation relies firstly on a realistic simulation of key nutrients state variables (phosphate and nitrate). Further assessment of the new growth model showed that the chosen combinations could not improve the current optimal model results. In all 5 water bodies tested, no major improvement to the simulation of total phytoplankton biomass and algal functional groups dynamic were achieved by replacing the present optimal SALMO-OO model structure with those from the new process model by Law et al. (2009). Addition of the Law et al. (2009) process model into SALMO-OO simulation library that was anticipated to improve the accuracy of simulating key state variables such as phosphate concentration, total phytoplankton biomass and algal functional groups dynamics from the present optimal structure of the six water bodies tested does not materialise. However, it must be noted that for the purpose of having a uniform treatment for all test lakes, these exercises were conducted without applying parameter optimisation. Therefore, there is still possibility that this new process model could improve the results from current simulation in the future by means of parameters optimisation and expert input for fine-tuning the parameter values according to respective lake conditions.

An accurate simulation of key state variables such as phosphate concentration, phytoplankton biomass and algal functional dynamics might have not been fully achieved by the addition of Law et al. (2009) process model into the SALMO-OO library, however, this process model had offered additional model alternatives for the simulation of key state variables for various lakes. Besides, this addition also provides more combinations of growth and grazing process function to the current library. The scope of the SALMO-OO simulation library has also been broadened by this addition. The list of current process model in the SALMO-OO simulation library is given in Table 34.
<table>
<thead>
<tr>
<th>Source</th>
<th>Algal Growth</th>
<th>Algal Grazing</th>
<th>Zooplankton Growth</th>
<th>Zooplankton Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benndorf and Recknagel (1982)</td>
<td>A1</td>
<td>A2</td>
<td>A3</td>
<td>A4</td>
</tr>
<tr>
<td>Park et al. (1974)</td>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
</tr>
<tr>
<td>Hongping and Jianyi (2002)</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
</tr>
<tr>
<td>Arhonditsis and Brett (2005)</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
</tr>
<tr>
<td>Law et al. (2009)</td>
<td>E1</td>
<td>E2</td>
<td>E3</td>
<td>E4</td>
</tr>
</tbody>
</table>

Even though the process model from Law et al. (2009) was applied on the same Lake Washington as Arhonditsis and Brett (2005) process model, the different between the two lies in their respective objectives. The process model by Law et al. (2009) was meant to examine the relative importance of the ecological mechanisms underlying plankton seasonal variability in Lake Washington with specific reference to effects of local climate change (air temperature) on the phenology of lake processes. The objective of Arhonditsis and Brett (2005) process model was to simulate plankton dynamics in Lake Washington for testing alternative managerial schemes where the model addresses eutrophication scenarios associated with nutrient loadings and compositional shifts in the plankton community. Therefore, the mathematical description representing both process models differs between each other.

Why Law et al. (2009) fails?

Adaptation of Law et al. (2009) process model for application on tropical lakes was unsuccessful. The Law et al. (2009) process model was found to be not a feasible solution to produce accurate simulation of state variables from tropical lakes ecosystem in this research. Reasons to this are that Law et al. (2009) was developed for a temperate mesotrophic lake itself. The reference water temperature used in Law et al. (2009) formulation is 20°C, whereas general temperature data from tropical lake in this research suggest that the average water temperature is around 29°C. Even though changes was made to the reference temperature to reflect tropical lake conditions of Lake Penang and Lake Kenyir, no satisfying results were obtained for nutrient simulation and other key state variables of respective lakes. Due to strong interdependence between growth rate, irradiance and temperature (Dauta et al., 1990), effect of temperature difference between temperate and tropical is also suspected to cause the inapplicability of Law et al. (2009) process model for tropical lakes especially the growth rates. An increase in water temperature can lead to increased algal growth rates. Biological effects of high temperature can be anticipated in view of the increase of photosynthesis and respiration rates with temperature (Talling, 1957a). However, as water temperature increase, growth may be slowed or terminated depending on the optimal temperature of algal species. Only light saturated photosynthesis is temperature-dependant whereas temperature effects are absent under limiting light (Talling, 1957b). This has been explained by the fact that
light-saturated photosynthesis is controlled by enzymatic processes while light-limited photosynthesis rates depend on photochemical reactions which are temperature independent (Tilzer et al., 1986). Goldman and Carpenter (1974) have found that algal growth rates involve many biochemical reactions and would follow the van’t Hoff rule: that is, biological reactions should approximately double for each 10°C rise in temperature.

Finding a process model developed purposely for tropical lakes is a challenging task. Apart from the very limited modelling works on tropical lake ecosystem observed in the literature, the present monitoring exercise and limnological research conducted in tropical lake ecosystem are normally on short term basis. This may hamper the effort of long term simulation model such as SALMO-OO. Instead, results from this research suggest that works should be concentrated on optimising parameters for input that are prominent in tropical lake ecosystem such at temperature and light conditions and differ from temperate zones ecosystem. This was also because the ability of phytoplankton biomass to increase is subject to a sustaining nutrients supply, light availability and optimum temperature in any water body.

The review on literature for phytoplankton and zooplankton process model was also limited by the time frame available for selection on models. Cetin (2007) study had contributed to the development of simulation library in SALMO-OO by evaluating numerous models on literature until year 2007. Therefore, this study had a limited year left and number of literature papers beginning from 2007 to consider for tropical lakes simulation.

5.4 Data-driven Forecasting Models Developed by HEA

The majority of process-based ecological lake models have been developed for temperate and Mediterranean lakes using parameter settings for chemical and biological process that are representative for these climate conditions. To apply these models to tropical lakes requires an adjustment of the process representations and parameter values to tropical conditions. By contrast the parameterisation of data-driven models results automatically by synthesising models from lake or category-specific data patterns. Therefore models developed by the hybrid evolutionary algorithms HEA for the tropical lakes Putrajaya, Kenyir and Penang are ad hoc and rely on the ‘pattern-richness’ of the data.

The Tab. 35 indicates that models developed by HEA for the lakes Kenyir and Penang reached quite a good validity reflected by $r^2$-values in the range of 0.76 to 0.95 as time series data of two years were available of these permanently stratified lakes with minor seasonal dynamics. The HEA models simulated well effects of monsoon seasons on algal biovolumes in Lake Kenyir and performed satisfying 7-days-ahead forecasts of Chl-a concentrations and algal biovolume in both lakes Kenyir and Penang.

However the time series data of the ‘young’ and shallow Lake Putrajaya proved to be highly uncertain for modelling by HEA reaching an unsatisfying $r^2$ of 0.33 only.
Therefore the key for improved modelling results by HEA will be more consistent and extended time series data of this lake.

RMSE is a measure of differences between values predicted by a model and the values actually measured. Even though the lower values of RMSE indicate better fit and shows how accurately the model predicts the response, the main purpose of the model is prediction. Simulation results are reported in the root mean squared error (RMSE) because the RMSE is measured in the same units as the data, rather than in squared units and is representative of the size of a "typical" error. For example, the RMSE values for Chl-a data (unit is µg/l) from Lake Penang showed lower value compared to RMSE results for biovolume (unit is µm³) of similar lake due to the conversion aspect of Chl-a to biovolume which requires a factor of 10×8. Therefore, the Table 35 should be read according to their respective Chl-a or biovolume grouping results and corresponding RMSE.

The \( r^2 \) and RMSE regression showed the closeness of predicted values to the observed data values. Even though lower RMSE values indicate better fit and show how accurate the model predicts ecosystem responses, another important criterion for ecological models are the accurate timing of predictions. In order to compliment these regression models for this research, visual observation aided by expert knowledge was also included. High RMSE value from generic forecasting model for algal biovolume was due to the variance between measured data from merged lakes Kenyir and Penang data. Both lakes differ principally in morphometry and management operations. Therefore, a different algal growth pattern and lower magnitude of algal biovolume exist in Lake Penang compared to Lake Kenyir.

<table>
<thead>
<tr>
<th>Table 35 List of ( r^2 ) and RMSE values obtained from HEA model experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment using Chl-a</strong></td>
</tr>
<tr>
<td>Predictive Chl-a model for Lake Putrajaya</td>
</tr>
<tr>
<td>Predictive Chl-a model for Lake Penang</td>
</tr>
<tr>
<td><strong>Experiment using algal biovolume</strong></td>
</tr>
<tr>
<td>Predictive algal biovolume model for Lake Kenyir</td>
</tr>
<tr>
<td>Predictive algal biovolume model for Lake Penang</td>
</tr>
<tr>
<td>Generic forecasting model for algal biovolume in mesotrophic stratified-tropical lakes</td>
</tr>
</tbody>
</table>

Valuable insights into relationships between physico-chemical input variables and algal abundances in the three tropical lakes have been gained by means of
sensitivity analyses and the frequency of input variables selection during the process of model evolution by HEA.

5.4.1 Input Sensitivity Analyses

The tropical climate in Malaysia is characterised by a distinct monsoon season and a dry season with little rainfall. Seasonal fluctuations of water temperature, transparency, storage levels and nutrient loadings in lakes like Kenyir, Penang and Putrajaya are restricted to the rainy season. As pointed out by Ochumba and Kibaara (1989) synergetic effects of these factors result in peak events of algal growth during the dry season and moderate algal growth during the monsoon season affected by flushing, temperature reduction and low transparency. Results of sensitivity analyses of all data-driven models developed in this research by HEA reflect well these findings by selecting variables related to these conditions and showing relationships with algal abundance measurements accordingly.

Lake Putrajaya

The model for Chl-a in lake Putrajaya applies the criterion of pH $\leq 7.9$ to the THEN-branch of the rule (see Fig. 15a) that relates to low Chl-a concentrations and the criterion pH $> 7.9$ to the ELSE-branch that relates to higher Chl-a concentrations. The fastest growth rates of Chl-a coincide with the pH range from 7.1 to 7.7 in the THEN-branch and from 7.8 to 8.4 in the ELSE-branch.

Shapiro (1973) has shown that pH is important in controlling algal dominance especially in determining the relative abundance of green and blue-green algae and believes this is associated with CO$_2$ availability. CO$_2$ concentrations can become high as a result of algal respiration. According to Stumm and Morgan (1970), the pH of freshwater is determined by its CO$_2$ budget. The free CO$_2$ released during respiration reacts with water, producing carbonic acid (H$_2$CO$_3$) and pH is lowered. High concentration of CO$_2$ lowers lake water’s pH and limits the capacity of algae to perform photosynthesis, therefore, lower Chl-a concentrations was observed when pH $\leq 7.9$ at the THEN-branch of the rule. Later, Shapiro et al. (1975) stated that high CO$_2$ or low pH stresses the algae and thus producing low level of chlorophyll level. Studies by Shapiro (1973) also indicate that blue-green algal frequency and abundance tend to be higher with increasing pH values. When CO$_2$ is taken up by photosynthesis, free H$^+$ ions are associated with HCO$_3$ and CO$_3$ ions leading to fewer free H$^+$ and thereby a higher pH (Reynolds, 1984). The findings from this lake confirm to these postulate where the criterion of pH higher than 7.9 in the ELSE-branch relates to higher Chl-a concentrations due to increase in photosynthesis. From this result, it can be concluded that this tropical lakes is naturally highly dynamic ecosystem and any perturbation may have far reaching effects on the relative dominance of biota.

With regards to nutrient dynamics in Lake Putrajaya, the ELSE-branch reflects optimal conditions for phytoplankton growth outside the monsoon season where highest Secchi depths of up to 2.4m and lowest NO$_3$-N inflow concentrations coincide with highest Chl-a concentrations (see Fig. 15b). By contrast, the THEN-branch reflects sub-optimal growth conditions for phytoplankton during the monsoon season.
in terms of underwater light so that phytoplankton seems not to utilise fully the relatively high PO₄-P concentrations (see Fig. 15a).

The unsatisfying low $r^2=0.33$ of the Chl-a model for Lake Putrajaya (see Fig. 16) is largely influenced by the mismatch of predicted and observed Chl-a dynamics in the years 2003 and 2004. This mismatch can be explained by the fact that these were the first 2 years of the lake after its construction characterised by extreme hydrodynamic patterns and highly fluctuating phytoplankton concentrations in the lake. However the 7-days-ahead forecasting results for the remaining years 2005 till 2008 demonstrated that predictive modelling of a shallow tropical lake like Putrajaya by HEA can be successful if reasonable time-series data of typical lake conditions are available.

A question arises whether excluding the years 2003 and 2004 would improve HEA results for Lake Putrajaya. Therefore, another HEA experiment on Lake Putrajaya (experiment 2) was performed in order to test input data for the years 2005 till 2008. The four years of daily interpolated water quality data consists of data from the years 2005 until 2008 excluding data from the years 2003 and 2004 which represent early period of lake inundation. The previous aim which was to develop a 7-days-ahead forecasting model for Chl-a of Lake Putrajaya was maintained for this experiment. Attempt was also made to include recent data of the years 2009 and 2010 from Lake Putrajaya to expand the scope of current modelling experiment. However, this was not possible largely due to change of consultant team within that year which resulted in differences of sampling preferences, methods and locations among the different years of monitoring.

All input variable used in experiment 1 were retained in this experiment 2. The resulted model of experiment 2 for Chl-a in Lake Putrajaya (see Fig. 18) applies the criterion of NO₃-N$_{\text{inflow}} < 0.22$ mg/l to the THEN-branch of the rule that relates to slightly higher Chl-a concentrations and the criterion of NO₃-N$_{\text{inflow}} > 0.22$ mg/l to the ELSE-branch that relates to slightly lower Chl-a concentrations. This findings is in agreement with Lee et al. (2003) where high chlorophyll-a levels are often associated with low levels of the limiting nutrient concentration (due to nutrient uptake by algae).

The sensitivity analysis for the THEN- and ELSE-branches also indicate that the sensitivity of Chl-a is always high to Secchi depth, but with opposite trends for the THEN- and ELSE-branches respectively. Optimal phytoplankton growth at highest Secchi depths and lowest NO₃-N$_{\text{inflow}}$ concentrations was observed at the THEN-branch (see Fig. 18a) while sub-optimal conditions phytoplankton growth during the warmer temperature was reflected in the ELSE-branch (Fig. 18b). The rule set suggested that pH, DO, N$_{\text{inflow}}$, PO₄-P, water temperature and Secchi depth were important criteria to stimulate the growth of algal. The causal relationship between Chl-a concentrations and these variables were further revealed by their sensitivity curves shown in Fig. 18. At lower value of NO₃-N$_{\text{inflow}}$ (NO₃-N$_{\text{inflow}} < 0.22$ mg/l, Fig. 18a) growth of algal biomass depends largely on Secchi depth level where increasing Secchi depth will support higher algal growth by allowing more photosynthetic active light to penetrate the water column. This has led to higher consumption of NO₃-N$_{\text{inflow}}$, causing the nutrient’s level to decrease. Even in lesser NO₃-N$_{\text{inflow}}$ value, algal biomass still showed increasing trend due to increasing level of Secchi depth coupled
with higher light penetration to the water column. The increasing algal growth is fast consuming NO$_3$-N nutrient and causes decrease in DO level. Eventually, the higher algal growth lead to decrease in dissolved oxygen level due to respiration.

However, at higher NO$_3$-N$_{inflow}$ value of more than 0.22 mg/l, slightly lower level of Chl-a was observed. The lower Chl-a a level was observed even though PO$_4$-P and pH showed strong positive relationship towards Chl-a concentrations (see Figure 18b). This could be partly due to the shading effect from increased algal biomass in the water column (Fig. 18b) resulted from the increasing level of NO$_3$-N$_{inflow}$ nutrient. However, increasing Secchi depth as observed in Fig. 18b does not lead to increase in algal growth. One possible explanation is that the rising temperature level observed for this rule is negatively correlated with algal biomass casing reduced algal growth when the water temperature is within the range of 29.9 to 31.1°C. It is well known that algal photosynthesis is regulated by water temperature (Kappers, 1980; Reynolds, 1984). As water temperatures increase more, growth of a particular algal species may be slowed or terminated (Goldman & Carpenter, 1974). Ochumba and Kibaara (1989) found optimum water temperature regime for algal growth in the open waters of Lake Victoria within the range of 25.9-28°C.

The influence of extreme hydrodynamic patterns and highly fluctuating phytoplankton concentrations of the years 2003 and 2004 in the lake was eliminated. As a result, the huge gap between prediction and measured data observed in experiment 1 was reduced and a slightly better result was obtained from experiment 2. The $r^2=0.41$ of the Chl-a model for Lake Putrajaya (see Fig. 18) from experiment 2 is an improvement compared to results from experiment 1 of the same lake ($r^2=0.33$). The root mean square error (RMSE) of 2.28 for this experiment was also slightly better than results achieved from experiment 1 (RMSE=2.35). Improvement was also made on the timing of major Chl-a peaks where most of the prediction managed to pick out period of high Chl-a level. Generally a better accuracy between predicted and observed Chl-a dynamics was obtained by experiment 2 as demonstrated by the 7-days-ahead forecasting results for the years 2005 until 2008 in this intensely managed lake. Despite the highly fluctuating phytoplankton dynamics of this shallow tropical lake, the availability of additional time-series data in the future will promise a better predictive modelling by HEA.

Lake Kenyir

The 7-days-ahead forecasting model for algal biovolumes in Lake Kenyir (Fig. 22) uses the criterion SD > 2.94 meter to activate the THEN-branch that stands for slightly higher algal biovolume than the otherwise activated ELSE-branch. The decrease in Secchi depth is a response to the start of raining seasons in the lake. At the commencing of monsoon seasons, heavy rain creates incoming water full of debris, organic litter and soil particles from the surrounding virgin forest causing increased turbidity in the lake and thus reducing the Secchi depth value. However, greater water clarity exists outside the monsoon seasons due to more stable water conditions and less water intrusion from surroundings catchment. This has increased the Secchi depth and consequently the algal growth. Gaedke et al. (1998) also reported that phytoplankton growth is tightly coupled to increased light availability.
With regards to nutrient dynamics the sensitivity of the functions of both the THEN- and ELSE-branch indicated the expected relationship that growing algal biovolumes cause diminishing PO$_4$-P concentrations. The algal of the lake consumed the PO$_4$-P nutrient in the water. The connection between phosphates and algae growth is well documented and it is a popular opinion that phosphorus is the key to algal blooms (Carter et al., 1970). In deep stratified lakes where stagnant water body covers the bottom, phosphorus release mainly occurs as a results of redox controlled phosphorus mobilisation and molecular diffusion (Bostrom as cited in Bostrom et al., 1988). Long duration of phosphorus release events can compensate for the slow rate of transport (Bostrom et al., 1988).

Sensitivity of algal biovolumes to both pH and water temperature in the THEN- and the ELSE- functions shows that increasing biovolumes correspond with increasing pH and water temperature. The sensitivity results indicate fastest growth at pH ranges from 6.7 to 7.4 for both branches inclusive. As the results ranges around almost neutral pH preferences, these pH results are comparable with the algal preferences at about pH 6.5 reported by Mayo (1997). With regards to water temperature, algal in Lake Kenyir was postulated to thrive in the range between 24.4 and 29.9°C. These results are also within the range of 23-26°C in the bloom reported by Hammer (1964).

Lake Kenyir

The algal biovolume model for Lake Kenyir has been developed by means of data only of the year 1992 and therefore the validation result of $r^2=0.86$ applies only to 1992 and the model cannot be judged as a valid model for Lake Kenyir until it has been tested for at least 3 years. The visual comparison between measured and forecasted data (see section 4.1.2, Figure 22) shows that the model matches well the seasonal dynamics of algal biovolumes in terms of timing of fast and slow growth as well as magnitudes of growth i.e. marginally lower algal biovolumes during monsoon seasons from October till December and April to May.

Lake Penang

The 7-days-ahead forecasting model for Chl-a concentrations in Lake Penang (see section 4.1.3, Fig. 27) uses the criterion $\exp(\text{SD}^2) > 41.2$ to activate the THEN-branch that stands for higher Chl-a concentrations than the otherwise activated ELSE-branch.

The function of the THEN-branch of the model indicates rapid NH$_4$ consumption along with phytoplankton growth while the function of the ELSE-branch suggests a decreasing chemical oxygen demand at increasing Chl-a concentrations. One possible explanation for an increase in Chl-a when COD level is decreasing is the C:N:P ratio in the lake. Cromar & Fallowfield (1997) found that a high C:N:P ratio increased phosphorus and COD removal and increased the concentration of algal biomass (chlorophyll a).

The relationship between COD and phytoplankton concentration may reflect the fact that water from river Muda is occasionally pumped into the lake that is believed to contain high level of oxidising potassium originating from fertilizers drained from the paddy fields along the River Muda. The rule regarding Chl-a obtained for Lake Penang which shows Chl-a decrease when COD is increasing
suggests that the oxidisable inorganic substance is supplied from the rivers pumped into the lake since there was no point source observed within or around the lake’s surrounding during site visit to the water body.

Even though it is common for an agriculture land use of any watershed to have a significant level of COD runoff as reported by Nakasone (2003), the vast area of Muda Paddy Agricultural Scheme is applying fertilizers and may cause impacts on water quality of downstream water bodies such as Lake Penang. The finding from this study should be given attention for the sustainable management of Lake Penang. One possible management option may be to dilute the high COD concentrations by increasing the flow in the river.

The Chl-a model for Lake Penang has been developed only by means of 12 months of data between 2005 and 2006 and therefore the validation result of $r^2=0.89$ applies only to these 12 months. Therefore the model cannot be judged as a valid model for Lake Penang until it has been tested for at least 3 years. The visual comparison between measured and forecasted data (Figure 27) shows that the model matches well the seasonal dynamics of Chl-a in terms of timing of fast and slow growth as well as magnitudes of growth i.e. peaks in early September 2005, early February and early May 2006.

The 7-days-ahead forecasting model for algal biovolumes in Lake Penang (see section 4.1.4, Fig. 32) uses the criterion COD $\leq 26.3$ to activate the THEN-branch that calculates lower biovolumes than the otherwise activated ELSE-branch.

In terms of nutrient dynamics in Lake Penang, a weak positive relationship was observed for algal biovolume against NO$_3$-N and water temperature respectively (THEN-branch) indicating consumption of NO$_3$-N at optimal pH (between 7.6 and 8.8) and water temperature (between 30.2 and 31.1$^\circ$C) ranges. The result for optimal pH for this experiment was found to be fairly similar with the pH-optimum for nitrate consumption in the range of 7.5 and 8.0 observed by Estuardo et al. (2008). A pH optimum within the range 6.5-7.0 has also been reported by Nilsson et al. (1980). These optimum values are lower than the one found in this work; the difference might be due to the different climate as in this work a tropical lake data was used. Optimal temperature for NO$_3$-N consumption in this experiment (between 30.2 and 31.1$^\circ$C) is comparable with the optimum temperature of 35$^\circ$C reported by Nilsson et al. (1980). Cyanobacteria are known to prefer high water temperature (Reynolds, 1984; Shapiro, 1990). However, this conclusion is based on research in the temperate zone and the have been few direct comparisons of nutrient and temperature effects in tropical freshwaters.

Another model has also been developed for algal biovolume in Lake Penang. The algal biovolume model for Lake Penang has been developed only by means of 12 months of data between 2005 and 2006 and therefore the validation result of $r^2=0.95$ applies only to these 12 months. Therefore the model cannot be judged as a valid model for Lake Penang until it has been tested for at least 3 years. The visual comparison between measured and forecasted data (see Figure 32, section 4.1.4) shows that the model matches well the seasonal dynamics of algal biovolumes in terms of timing of fast and slow growth as well as magnitudes of growth i.e. peaks in early September 2005, early February and early May 2006. The function of the
THEN-branch indicated that increasing algal biovolumes coincide with low BOD values and vice versa. For Lake Penang, conditions ideal for the rapid growth of algal biovolume occurred between pH from 7.4 to 8.0 (ELSE-branch) and minor increase in algal biovolume was observed pH range of 7.6 to 8.8 (THEN-branch). The consumption of NH$_4$ by this slightly lower concentration algal biovolume (ELSE-branch) contributed to the algal growth. This growth has contributed to the increasing TSS value (ELSE-branch) as phytoplankton also contributes to the turbidity level and subsequently TSS.

5.4.2 Generic Forecasting Model of Algal Biovolumes for the Lakes Penang and Kenyir

The forecasting model for algal biovolumes (see Fig. 35 and Fig. 36) resulted from the application of the hybrid evolutionary algorithm HEA to merged data sets of the lakes Penang and Kenyir and has been proven to be valid for both lakes. It therefore can be considered as a first step towards generic forecasting models for phytoplankton dynamics in mesotrophic stratified tropical lakes. Even though the validation results for the two years of both lakes are encouraging future research should provide data of additional years from both lakes and ideally new data of other ‘mesotrophic stratified tropical’ lakes in order to update the model and prove its validity for a broader range of seasonal pattern and properties of lakes belonging to this lake category.

Recknagel et al. (2008b) defined the lake categories by trophic state and circulation type, assuming that circulation type reflects climate conditions and morphometry to some extent, whilst the trophic state indicates habitat properties and community structure. Recknagel et al. (2008b) and Welk et al. (2008) have successfully tested this paradigm for the categories ‘warm monomictic and hypertrophic’, ‘warm monomictic and mesotrophic’ as well as ‘shallow polymictic and eutrophic’. This research has successfully extended the testing of the paradigm to the category of ‘mesotrophic stratified tropical’ lakes.

Advantages of having generic models for lakes of the same category are to gradually enable model sharing for lakes with limited data sets and to improve comparisons of lakes based on sensitivity analyses and forecasting results from the same model.

5.4.3 Generic Forecasting Model of Algal Biovolumes for the Lakes Penang and Kenyir using Electronically Measured Data

After utilising upon the generality of HEA for developing forecasting model, it is helpful to find the other capability that the HEA can offer. Advances in electronic technology and networking have allowed real time measurements of selected water quality parameters. In such cases, the development of forecasting model through electronically measured data and/or telemetry is the preferred option. This research
attempted to explore the prospect of initiating an early warning system of algal biovolumes for lakes Penang and Kenyir on real time basis, generated by model developed by means of HEA. These models are applicable only for those parameters that can be measured directly from the measuring instrumentation such as meteorological sensors, data loggers and telecommunications allow for routine measurement of critical parameters at fine temporal scales. The decision to used only electronically measureable input variables to determine the algal biovolumes at any particular time was made to ensure that they could be applied to real-time situations for these tropical lakes.

It was found that the models generated using electronically measured data (see section 4.3) were able to match capacity of model generated by HEA from experiment using both manual and electronically measured methods (see section 4.1 and 4.2). Models trained with only electronically measured inputs of water temperature, DO and pH were able to successfully forecast algal biovolume level 7 days in advance. This produced models that are considered as compatible with result from section 4.1 and section 4.2 despite the fact that only electronically measured input data were used. Welk (2007) showed that HEA could contribute toward improved water management by real-time forecasting using HEA with promising results with regards to the models application to real-time data provided on-line monitoring of relevant variables is implemented. The results from this research have proven that the electronically measured dataset is on par with the data gained from both manual and electronic ways. Another advantage of using electronically measured data is that the data is highly mobile and this facilitates electronically storage, access, transfer, documentation and archiving thus, minimizing both the reading errors and the writing errors. Research on water quality modelling ecology is placing increasing demands on long-term measurement collection systems. The approach taken by this research in utilising the electronically measured data will help to cater for this demand and is highly recommended to be enlarged to include more feasible parameters in the current lake monitoring regime. On top of that, the lakes management authorities need for eutrophication and algal bloom forecasting and subsequent quick decision making in preparation for the recurrent algal bloom in the region would be taken care of.

5.5 Combination of Process-based and Data–driven Modelling Techniques

Lakes and drinking water reservoirs are highly complex ecosystems. To model the behaviour of these systems realistically requires sophisticated modelling techniques. As mentioned by Reichert et al. (2008), lake models provide a link from concentrations to fluxes and transformation rates that are much more difficult to measure than concentrations. Each of these models has their own strength and limitations. More than one modelling paradigms are often needed to capture the minimal description of a complex ecosystem, to address the complexity and multidisciplinary nature of environmental problems, the spatial and cultural de-localisation of co-operating research groups and the recognition of cross-scale effects with different phenomena operating at different scales (Villa, 2001).

Thus the integration of modelling options, whether they are in the form of hybridised technologies or alternative model structures, into a single framework can
assist in the development of more informative models and make the computerised decision-making process easier (Cetin, 2007). The application of the data-driven modelling by HEA and process-based modelling by SALMO-PLUS in this research has provided both insights into the causality and forecasting of phytoplankton dynamics of two tropical lakes ecosystem.

While SALMO-PLUS simulates abiotic and biotic state variables such as nutrient concentrations and phyto- and zooplankton abundance based on key ecological processes of lakes, HEA models forecast phytoplankton abundance and reveal causal relationships between phytoplankton and abiotic lake parameters by sensitivity analysis. The combined information from both modelling techniques supports improved understanding and sustainable management of tropical lake ecosystem.

The approach of this research conforms to Ford and Thornton (1992) who identified two additional criteria that should be considered in evaluating and selecting water resources model: model complexity and environmental compatibility. Even though complexity and incorporation of details do not necessarily lead to the construction of a better model, both complex and simple models should be complimentary to each other with the ability to relate to environmental endpoints.

It is also unlikely that any single experimental study will demonstrate the causal relationship between the nutrient input, algal population dynamics and eutrophication onset especially with a limited dataset from tropical lakes such as in this study. Thus, in order to achieve greater confidence over the outcome of this experiment results, combination of more than one approach will provide reliable conclusions. While HEA was proven to successfully developed rule sets as prediction tools for highly complex ecological time series (Cao et al., 2006b), Walter et al. (2001) applied two different models for Lake Burrinjuck in Australia using both SALMO-OO and ANNA (recurrent feed-forward neural network) and concluded that different types of models are needed in order to meet required predictions for the management of freshwater lakes.

The implementation of both HEA and SALMO-OO models will allow the proactive forecasting and appropriate implementation of operational control measures to prevent or reduce the potential of algal bloom and eutrophication in tropical lakes.

5.6 The DYRESM Model for Tropical Lakes

This research was the first attempt to demonstrate the applicability of DYRESM model for predicting the vertical distribution of temperature, salinity and density for hydrodynamic studies of Lake Putrajaya. DYRESM had been used for lake simulation periods extending from weeks to decades worldwide, this study also attempted to use the model as a means of predicting seasonal and inter-annual variation in Lake Putrajaya.
What has been done?

The thermal profile of Lake Putrajaya was simulated by DYRESM utilising vertical water temperature, salinity and density with horizontal Lagrangian layers that vary in thickness and number. This author assumes that the light extinction coefficient of water remains constant during the simulation period. The application of DYRESM for Lake Putrajaya does not produce reasonable output as expected. The nine possible text files used for input to the model with a brief description of each is shown in Appendix B. In this experiment, it was assumed that Lake Putrajaya comply with the one-dimensional approximation in that the destabilising forcing variables (wind, surface cooling, and plunging inflows) do not act over prolonged periods of time.

What has happen?

Being a relatively small in size, Lake Putrajaya was expected to exhibit one-dimensionality and subsequently one-dimension model can be used to predict the physical structure of the lake. However, the results from initial test with DYRESM on Lake Putrajaya do not successful. Validation and calibration testing of DYRESM was not been able to be successfully performed with the model. Comparison with default input data showed the error was most likely due to process representation in the model or error in the field data, such as probe positioning while calibrating the temperature profile measurements on site. To justify the assumption for using one-dimensional numerical model, investigation on the lateral mixing mechanism for Lake Putrajaya should be conducted first.

Several explanations are suggested for this:

**One-dimensional assumption**

The assumption that the water bodies comply with the one-dimensional approximation requires further investigation. One-dimensional models are most commonly used in rivers but can also be used in lakes with large length-to-width ratios and this one-dimensionality assumption could also be a limitation of this model for usage in Lake Putrajaya. The one-dimensionality approach in DYRESM assume that first order balances of energy, momentum and mass are determined by vertical property variations and fluctuations in lateral directions are small compared with vertical variations (Hamilton & Schladow, 1997). However, destabilising forcing functions such as wind is causing surface cooling and is believed to act for a longer period of time in Lake Putrajaya. Also, three-dimensional effects such as circulation associated with convective heating and cooling may disrupt vertical density structure implicit in the one-dimensional assumption such as adopted by DYRESM (Tanentzap et al., 2007).

Wind is an important meteorological factor in the behaviour of shallow lakes (Shanahan et al., 1993). The wind affects lakes by creating a complex pattern of horizontal circulation and a strong seiche, which creates frequent re-suspension of the sediments from the bottom of the lake. This coupled with large daily variation of air temperature in Lake Putrajaya, has caused dynamic fluctuations of algal abundance as shown in experiment with HEA model results from this research. This corroborate the finding by Horne and Goldman (1994) that daily changes play a more important role in the plankton community pattern in tropical lake. The closer a lake is to the equator, the more daily cycles dominate its ecology (Horne & Goldman, 1994).
The validation of one-dimensionality of Lake Putrajaya should also be tested by means of Lake Number $L_N$ (Imberger & Patterson, 1989) and even with this assumption, the vertical density structure is generated by a combination of incompletely understood processes (Patterson et al., 1984). The Lake number is categorised by the stratification stability and the magnitude of the wind. The water stratification is subjected to wind field that provides friction on the lake surface. Wind stress on the water surface layer provides a forcing function to overturn the density structure of the water column. Due to the unusual shape of the lake – with an island in the middle and narrow path for water movement, the wind stress cannot be assumed to be constant over the entire water surface. Multiple stratifications were observed during diurnal sampling where stratifications of the lakes are believed to be overturned by imposing surface wind stress.

Vertical mixing in DYRESM is controlled by the wind speed. However, test on reducing the wind speed by a factor in DYRESM to account for wind sheltering in Lake Putrajaya in the .par file (effective surface area coefficient) does not improve the model’s run. Turbulence mixing in the metalimnion and/or hypolimnion also contributed to the breaking up of the stratification. The average depth of study site is 12 meters and is also the deepest part of Lake Putrajaya. This site is close to the inflow and the underwater current was observed to be relatively high during the sampling period. There was dynamic sedimentation observed during the sampling period where construction from bordering area is causing silted pathway for boat in the lake. This could also contribute to the unusual pattern of underwater current from month to month during sampling period. Topographic is also another aspect to consider where sheltering effect from the prevailing winds may results in a shallow surface mixed layer.

Thus, a horizontal 2D model might offer more insight in the factors determining local water quality including the description of the hydrodynamic and ecological processes to overcome the limitation assumption in vertical 1-dimensional model.

**Local climate conditions**

This could be explained by looking at the local climate conditions. According to Tanentzap et al. (2007) regional environmental influences, i.e groundwater intrusions or wind sheltering may confound mixing and forcing processes, potential influencing model simulations. Previous application of DYRESM in Dam Terip (a tropical water bodies in Malaysia) by Kassim et al. (1998) may not be broadly applicable to the morphometry and local climatic of young water body such as Lake Putrajaya due to its dominance by human input in terms of hydrological and drainage aspect. The Terip Dam is a strongly stratified reservoir which requires the use of bubble plumes to break the thermal strata to alleviate the problems of high levels of iron and manganese (ASM & NAHRIM, 2010). The scenario is different with Lake Putrajaya where wind induces turn over in the lake produces a highly dynamic stratification almost daily. Sahoo (as cited in Sahoo & Luketina, 2006) demonstrated that tropical reservoirs experience weak stratification compared to temperate and subtropical reservoirs.
**Input data**

As the meteorological data used for the simulation was collected from a nearby weather station, a sensitivity analysis is needed to evaluate the effect of meteorological data (such as air temperature, shortwave solar radiation, wind speed and vapour pressure) bias on the simulation results. Etemad-Shahidi *et al.* (2010) strongly recommended that parameters such as air temperature and short wave radiation to be measured in situ. Also, accurate measurements of light extinction coefficient are recommended to be carried out because this coefficient affects the thermocline position.

**Conclusions**

Consideration should be given in the future studies to the use of measured parameter values to reduce source of error in DYRESM. This first time research on using DYRESM for Lake Putrajaya has shown the large gaps in input data with a good quality. Yet this should not impede further research on the hydrodynamic of this lake and it is believed that the DYRESM has the potential to describe physical processes in this lake if the forcing functions are accurately given.

**5.7 The Limitation of Input Data Set from Tropical Lakes in this Research**

Assumptions and concept in models developed for temperate environments can be very different in terms of physico-chemical and biological parameters if applied to the tropics. Thus it is necessary to validate whether parameters values and algal growth assessment from temperate studies are applicable to tropical lakes ecosystem in Malaysia.

In comparison to other managed water bodies in Malaysia, a considerable research efforts have been invested to Lake Putrajaya, Lake Kenyir and Lake Penang (listed according to the most extensively investigated order first). Among the three tropical lakes studied in this research, only Lake Putrajaya has a wide-ranging and structured monitoring effort to monitor the lake’s health status. A comprehensive and continuous monitoring exercise at Lake Putrajaya begins since the lake construction was started in 1999. However, dynamic and rapid sedimentation by surface run-off due to mass land development on the surrounding area posed major problems for water quality monitoring program in all these water bodies. In Lake Putrajaya, several water quality sampling stations had to be relocated due to the inaccessibility of the areas. The problem with Lake Kenyir is the opposite. Lake Kenyir continues to receive attention from scientific community from time to time due to its huge size and strategic importance. However, these interests were constantly limited by the logistical factors where Lake Kenyir is located far from the closest town in the area. Among the three lakes, Lake Penang is the least to receive attention from researchers probably due to its small size. However, adequate funding is believed to be another major obstacle for the continuation of an established limnological research in Malaysia. Sources of funding that are limited to selected government resources (Sharip & Zakaria, 2008) are mere comparison to the huge number of water bodies in Malaysia. Although there has been basic water quality monitoring of Lake Putrajaya by the local authority dating back to the 2000, nation-wide lake monitoring in
Malaysia are costly, logistically challenging and inconsistent. There are too few published studies of lake dynamics with regards to algal blooms and little is known in detail of seasonal and spatial trends in water chemistry and biological assemblages across the lakes in Malaysia.

Limited water quality data certainly is the greatest constraint on the accuracy of models. Nevertheless, results from this research have shown that the process-based (SALMO-PLUS) modelling and data driven (HEA) approach can be applied to tropical lakes ecosystem in Malaysia but the results must be treated cautiously because of the limited data. In relation to data, this research on HEA has used bootstrapping method which is suitable for a limited data study. The merged data from same lake ecosystem category approach was selected in order to develop rule-based model generic for tropical lake ecosystem category and HEA has proven their suitability and applicability to a wide range of lake conditions. If the monitoring can be continued and extended for the future, a better prediction on the algal population dynamics in these lakes is possible. Present results had indicated factors influencing these tropical lakes ecosystems and additional analyses in the future will enhance the understanding the dynamic of these lakes better.

The limitation of available data also posed the possibility of under-represented major events such as monsoon seasons. The forecasted phytoplankton dynamic under these conditions is more uncertain than under the common conditions that are likely to be represented in the available data. However, despite the limited long-term data available, it is believed that the results discussed in previous chapter conveyed much acquired confidence in using the SALMO-PLUS and HEA models. This study has shown to improve the versatility of both models to simulate phytoplankton dynamics in lake ecosystems far from those for which its use was originally designed. Comparison of different forecasting techniques has been able to be explained even though only on snapshot results due to limited data available. These testing methods were able to provide a model that would be able to be used operationally in tropical lakes.

Models for tropical lakes

Tropical lakes model are rare. Jaffe (1988) constructed a dynamic model to predict temporal changes in the algae biomass as a function of nutrient loads to a tropical eutrophic Lake Valencia. Parameter determination and prioritisation for his model were based on F values where a high and significant F value means that the nutrient dynamics in Lake Valencia are very sensitive to the corresponding parameter and that the value of this parameter should be obtained from site-specific field measurements. The most sensitive parameter found by his research is the chlorophyll-to-nitrogen ratio for algae. However, Jaffe (1988) model simulations had a high degree of uncertainty associated with them. No similar modelling of nutrient input and phytoplankton dynamics for tropical lake as attempted by this research was found in the literature.

Indeed, the fluxes and transformation rates of tropical lakes are high and variable (Marotta et al., 2009) and information on nutrient cycling from temperate lakes cannot necessarily be extrapolated to tropical lakes, because of the fundamental difference in the physical and biological dynamics of the two types of systems.
(Gardner et al., 1998). Higher temperatures in tropical lakes tend to enhance biological activity and enzyme reaction (Lewis, 1987).

However, results from SALMO-PLUS showed that nutrient state variables PO$_4$-P and NO$_3$-N are capable to be simulated using models that have never been used in tropical lake ecosystem. Therefore, in can be concluded that the functions related to nutrients cycle in SALMO-PLUS is valid for tropical lakes scenario subject to parameter optimisation. Therefore, future research could focus on the development of process model specific for tropical lakes ecosystem categories with the objective to predict algal bloom for early warning system and for sustainable water quality management purposes.

5.8 Closing Gaps on the Tropical Lake Ecosystem Analysis: Role of SALMO-PLUS and HEA Models

The application of both process-based approach model (SALMO-PLUS) and the data driven computational modelling techniques (HEA) to tropical lake ecosystem time series data in this research complimented each other in terms of prediction and elucidation of short-term and long-term dynamics of the plankton community and water quality. These two different but related approaches present a bigger picture on the current phytoplankton growth status of Lake Kenyir, Lake Penang and Lake Putrajaya in Malaysia. Research requirements elucidated from modelling approaches in this research contributed to the better understanding of the ecosystem processes in tropical lakes. These elucidations are also needed to compliment the current monitoring exercise especially in Lake Putrajaya. Without clear and proper direction gained from modelling exercises, any monitoring exercise would not be useful in term of predicting the future trend in algal blooms and potential eutrophication.

While there are many policies and institutional issues related to man-made water bodies and water quality needs to be addressed in Malaysia, technical issues such as the capability to predict phytoplankton growth and algal bloom timing are the most important aspects to be addressed for achieving the target in water resources management. Both SALMO-PLUS and HEA models were shown to be capable of addressing this basic issues of eutrophication prediction for integrated water resources management in Malaysia.

A dogma of aquatic science is that marine and estuarine phytoplankton tends to be nitrogen limited while freshwater phytoplankton tends to be phosphorus limited (Hecky & Kilham, 1988). Phosphorus in most cases is considered as the limiting factor of eutrophication in temperate lakes (Vollenweider et al., 1980) but nitrogen limitation of phytoplankton appears to be very common in tropical lakes and evidence for widespread occurrence of nitrogen limitation in tropical lakes involves several kinds of indicators that are known to correlate will with experimental evidence (Hecky & Kilham, 1988; Jaffé, 1988; Lewis, 2002).
During the last century it was increasingly recognised that tropical lakes differ from temperate lakes in some basic features such as (Barbosa & Padişáč, 2002);

1. N-limitation of phytoplankton growth is more frequent than in temperate lakes
2. anoxic hypolimnion in the stratified period do not indicate eutrophic status
3. they need specific ranges for the classification of their trophic status

The nitrate limitation in these study lakes has raised doubt with regards to physico-chemical index since the Carlson index will not work if the phytoplankton are limited by a nutrient other than phosphorus or where Secchi depth is influenced by humic acids, sediments or other non-algae-related factors (Horne & Goldman, 1994). On another note, the limiting effect of nitrogen on phytoplankton growth in tropical lakes is more difficult to describe than phosphorus limitation because they are some phytoplankton species that are able to fix nitrogen from dissolved molecular nitrogen. Virtually all stratified lakes at low latitude can be expected to lose a substantial portion of their nitrogen inventory on an annual basis through denitrification and thus have greater likelihood of nitrogen depletion by phytoplankton growing in the upper water column (Lewis et al., 2002).

5.9 Algal Functional Groups Dynamics in Tropical Lake Ecosystem

The most commonly acknowledge factors affecting phytoplankton growth and reproduction are temperature, nutrient availability (nitrogen and phosphorus) and light conditions. Other factors which determine their growth include water temperature, thermal stratification of the water column and zooplankton grazing as well as resource competition (Whigham & Recknagel, 2002). Growth and other process activities of phytoplankton are in turn affected by ecological and hydrological conditions of the water body especially the water movement. Phytoplankton is transported both vertically and horizontally by undercurrent and therefore is exposed to a gradient of temperature, light and nutrient availability.

In SALMO-PLUS, different functional groups of algal have similar process formulation but differ in model parameter values. Algal functional groups differ according to different trophic level of the water bodies which relate to their ecological properties and nutrient loadings. Ecological key processes in tropical lakes are likely to be characterised by faster process rates as a result of higher water temperature conditions. However, very few studies on tropical lake ecosystems have discussed the seasonal fluctuations of algal functional dynamics or nutrient loadings (Pinto-coelho, 1998; Jeppesen et al., 2007; Descy & Sarmento, 2008; Ndebele-Murisa et al., 2010) which restrict comparisons across a trophic gradient.

The reliance on light for growth produces a seasonal pattern to the phytoplankton community dynamic although this signal may be altered by changes in any if the previously mentioned factors. Study on phytoplankton seasonality in Thai freshwaters and review of studies in other tropical areas showed that the phytoplankton dominance by different divisions depends on season-related weather conditions and water chemistry (Peerapornpisal as cited in Daam et al., 2009;
Algal successions in tropical aquatic systems are generally characterised by a sharp contrast between the two main seasons (dry/wet-flood season) with cyanobacteria dominating dry season (Bouvy et al., 2006).

5.10 Conclusions

Although the study with bigger database from larger number of tropical lakes may be required to draw a final conclusion, present results in this research are very encouraging and suggest that both HEA and SALMO-PLUS model are well-suited for eutrophication and algal bloom management in tropical lakes. Use of process based (SALMO-PLUS) and data driven (HEA) model in this research have shown that both models can be effective tools for water quality management in tropical lakes. The main conclusions are presented based on the aims outline in the Introduction as follows:

1. Can the models HEA and SALMO-OO be validated for tropical lakes with different morphometry and trophic levels with regards to food web dynamics, nutrient cycles and phytoplankton growth?

A generic SALMO-OO model in its SALMO-PLUS form contains specific attribute capable of searching for the optimal model structure and parameters values and performs parameter optimisation while maintaining the ability to simulate a wide variety of lake conditions and scenarios. The original SALMO-OO model has only been applied for simulation on non-tropical lake before. These new features in SALMO-PLUS were utilised to find a model structure that performs best for mesotrophic tropical stratified lakes. Results showed that SALMO-PLUS was able to 1) discover the most suitable model structure and optimum parameters for mesotrophic tropical stratified lakes and 2) identify a generic model structure and optimum parameters for a specific lake category. The causal relationship of eutrophication and algal growth dynamic for study lakes were also evaluated and explained by means of rules provided by HEA. The rule-based HEA models were verified to be both explanatory and predictive with sensitivity analysis describing input and output causal relationship.

Results from SALMO-PLUS model prediction in this research confirms that state variables of tropical lake ecosystem follows the same dynamics quantitatively and qualitatively as in temperate lakes but differ in the specific parameters value as discussed in section 5.2. However, prevailing influence of local climatic such as monsoon seasons as observed in HEA results was also obvious as expected.

Both HEA and SALMO-PLUS models can simulate the growth of phytoplankton reasonably well and produce acceptable results. Therefore, the issue of validating data driven and process based model for forecasting and understanding the dynamic of algal bloom in tropical lakes ecosystem has also been successfully addressed by this research. The scenario analysis option in SALMO-PLUS model is an ideal tool for quantitative testing of hypothesis about tropical lake ecosystem dynamics. In terms of quantitative results, the accuracy of both models to predict key
state variables in tropical lakes is comparable with results from others. The forecast for key events such as high algal growth seasons was also well observed. The outcomes from both models provide additional foundation for the successful implementation of the Catchment Development Management Plan (CDMP) for respective study sites as well as other lakes and reservoirs in Malaysia in order to protect the integrity of tropical lake ecosystem.

2. Can a generic model structure of SALMO-OO be found for mesotrophic tropical stratified lakes?

The key characteristics of SALMO-OO as a generic model that allows the simulation of different lake conditions using the same model structure was tested for tropical lakes. As a result, structure that performs the best in terms of predicting algal growth for mesotrophic tropical stratified lakes conditions has been able to be identified by submitting combinations of alternative phytoplankton growth and grazing process models within the simulation library of SALMO-OO for testing in this study. This best structure has demonstrated and confirmed the flexibility and generality of SALMO-PLUS in forecasting state variable dynamics for lakes beyond its conventional domain such as in tropical lake ecosystem.

This research has confirmed the capability of SALMO-PLUS in predicting the state variables dynamics in mesotrophic tropical stratified lakes in Malaysia. The computational efficiency of SALMO-PLUS makes it possible to analyse the model performance through changing its parameter values so that better simulation are obtained compared to those of the original SALMO-OO model. The SALMO-PLUS model utilise in this research has managed to simulate state variables and phytoplankton biomass with reasonable range and magnitude. According to Reichert et al. (2008) getting the important mass fluxes correct are vital for models meant for management purposes. This has the potential of assisting local decision makers in choosing the best management practises for their respective lakes management objectives.

3. Can a generic predictive rule-set be developed by HEA for algal growth in tropical lakes with different morphometry and trophic states?

Each of the water body studied in this research has unique qualities that require specific management input to maintain the water quality. A generic rule-set model is a standard rule which can facilitate model adaptation to a particular lake ecosystem category. Results from this research have demonstrated and developed predictive rule-set for algal growth in tropical lakes ecosystem with different morphometry and trophic states and produces generic rule-based model for tropical lakes ecosystem category by means of HEA. The study has successfully shown that using merged limnological data of lakes from the same ecosystem category is another potential way for generalising rule-based model in HEA. By utilising generic rule-set, local authorities managing these lakes is be able to reduce the time and money required to develop and maintain several different models for different study sites and can be applied to several different sites with slight adaptation. These generic rule-set
are also useful when forecasting for lakes with inadequate data such as the situation in the tropical countries. Steps need to be taken in the future research to overcome this weakness in order for gaining confidence on the conclusions made especially as the basis of comparisons among lakes. Despite that, these results have gained some insight into the current monitoring exercise on respective tropical lakes under study and are capable of enhancing the current lake management regime. The physico-chemical data associated with phytoplankton dynamics and contributing input factors can be used to identify bloom sensitive scenarios and should be highlighted in the current management strategies for tropical lake ecosystem.

Even though this is the first research on producing generic HEA models for tropical lake, the generic HEA model can be validated by time-series data from other tropical lake ecosystem with different trophic status and hydrological conditions for future research.

4. Can the model DYRESM be validated for dynamic, shallow-polymictic tropical lake such as Lake Putrajaya?

DYRESM model was applied to simulate the hydrodynamic of Lake Putrajaya in this study. This model has rarely been applied to tropical water bodies in Malaysia. The main objective for applying DYRESM was to determine if the model could be used as an effective management tool for water quality in Lake Putrajaya. Prediction of thermal structure for Lake Putrajaya by means of DYRESM was expected to provide guidance for lake authority in simulating the changes in thermal regime of the lake in response to climate change and predict its future behaviour. However, despite the attempt using one-year hydrodynamic and water quality data from Lake Putrajaya, the experiment was not successful. Hydrological input factor was suspected to cause this and other probable causes were explained in discussion section.

5.11 Future Works

By adapting the SALMO-PLUS ecological model for simulating tropical lake ecosystem and by conducting two different modelling approaches for algal growth assessment in tropical lakes, this research has overcome the two challenges for making further progress in the field of lake ecosystem model as listed by Mooij et al. (2010). The two challenges are 1) to avoid developing more models largely following the concept of other (reinventing the wheel) and 2) to avoid focusing on only one type of model, while ignoring new and diverse approaches that have become available (having tunnel vision). On top of that, the three main objectives for constructing and using lakes model (Reichert et al., 2008) which are 1) improving the understanding of lake ecosystem function 2) summarising and communicating knowledge about lake ecosystem and 3) supporting lake ecosystem management were obtained by this research. Despite this accomplishment, further works to improve current models should not be overlooked.

All the improvements made to date and most of the additions likely to be made in the future probably concern the addition of bigger pool of simulation library in
SALMO-OO which the respective lakes are set to operate. Some of the specific areas that have been developed recently include adding sediment nutrient flux routine predictions to the SALMO-OO model. The new process model from Law et al. (2009) that has been included in the SALMO-OO had a degree of uncertainty associated with them due but, the models can improved if more validation on different set of input values and parameter values optimisation from broader lake range with different morphometry and trophic status were conducted. Aspects to be considered for further improvement on HEA includes 1) implementation of particle swarm optimisation (PSO) method and 2) another method for optimisation known as ant colony optimisation (ACO). Further studies on the use of the process based and data driven modelling shown here will have to focus on models incorporating more variables related to tropical conditions and constraints related to the management of respective water bodies such as water withdrawal. More parameter optimisation studies for process based model will ensure better reliability and increase acceptability for use in tropical lake ecosystem.

The two modelling approaches (SALMO-PLUS and HEA) for analysing the relationship among the input and tropical lake ecosystem response were working well but the pattern showed was limited due to data availability. It is good to note that all of these tropical lakes are controlled to some extent. Thus, additional sampling directed towards the measurement of the longer time span is desirable. The impact of management is probably largest for Lake Putrajaya due to its smaller surface area and closer proximity to the local population. This will be the challenge for future studies in which the sampling exercise could be tailored for more meaningful objective with data archive needs to be expanded.
References


Blenckner, T. (2008). Models as tools for understanding past, recent and future changes in large lakes *European Large Lakes Ecosystem changes and their ecological and socioeconomic impacts* (pp. 177-182).


Kim, D.-K., Cao, H., Jeong, K.-S., Recknagel, F., & Joo, G.-J. (2007a). Predictive function and rules for population dynamics of Microcystis aeruginosa in the
regulated Nakdong River (South Korea), discovered by evolutionary algorithms. *Ecological Modelling*, 203(1-2), 147-156.


of the 1979 Cornell University Conference Phosphorus Management Strategies for Lakes.


NOTE:
This appendix is included in the print copy of the thesis held in the University of Adelaide Library.
## APPENDIX B

**DYRESM - input text files**

<table>
<thead>
<tr>
<th>No.</th>
<th>File Title</th>
<th>File Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Configuration</td>
<td>.cfg</td>
</tr>
<tr>
<td>3.3</td>
<td>Physical data &amp; lake morphometry</td>
<td>.stg</td>
</tr>
<tr>
<td>3.4</td>
<td>Initial profile</td>
<td>.pro</td>
</tr>
<tr>
<td>3.5</td>
<td>Meteorological data</td>
<td>.met</td>
</tr>
<tr>
<td>3.6</td>
<td>Stream inflow</td>
<td>.inf</td>
</tr>
<tr>
<td>3.7</td>
<td>Withdrawals</td>
<td>.wdr</td>
</tr>
<tr>
<td>3.8</td>
<td>Parameter</td>
<td>.par</td>
</tr>
<tr>
<td>3.9</td>
<td>Artificial mixing*</td>
<td>.mix</td>
</tr>
<tr>
<td>3.10</td>
<td>Field data</td>
<td>.fld</td>
</tr>
</tbody>
</table>

*not applicable

# wherever no data is available, data from closest site was chosen as an alternative.

### Simulation run

Simulation 1: unchanged parameters of DYRESM

Simulation 2: activation of the non neutral atmospheric stability to test and enhance mixing in the surface layer (air-water)

Simulation 3: activation of the non neutral atmospheric stability & an accurate value of the light extinction coefficient (field measurement)

Simulation 4: alteration on shortwave radiation and wind speed (sheltering factor)

Comparison of field and simulation result
3.2 DYRESM configuration file (.cfg)

Dyresm Configuration File for DYCD V4.0.0

2008120 # Simulation start day: YearDayNum
300 # Simulation length (days)
.FALSE. # Run CAEDYM (.TRUE. or .FALSE.)
1 # Output Interval (in days, or -9999 for every time step)
0.85 # Light extinction coefficient \([\text{m}^{-1}]\)
1.5 # Min layer thickness [m]
3 # Max layer thickness [m]
3600 # Time Step [s]
3 # Number of Output Selections
TEMPTURE SALINITY DENSITY # List of Output Selections
.FALSE. # Activate bubbler (.TRUE. or .FALSE.)
.TRUE. # Activate non-neutral atmos. stability (.TRUE. or .FALSE.)

* .TRUE. = promotes heat transfer between the atmosphere and the surface of the lake
3.3 Physical data and lake morphometry file (.stg)

Comment line: Reservoir morphometry

(+2) # latitude
100 # height above MSL
2 # number of inflows
SURF 88.8 0.24 0.016 Stream1 # entry height, 1/2-angle, slope, drag coefficient name
0 # zero-height elevation (i.e., bottom elevation)
33.1 # crest elevation [m]
1 # number of outlets
19.31 # outlet heights
11 # number of stg survey points after header line

<table>
<thead>
<tr>
<th>Elev [m]</th>
<th>SurfArea [m^2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>127200</td>
</tr>
<tr>
<td>6</td>
<td>200000</td>
</tr>
<tr>
<td>9</td>
<td>289000</td>
</tr>
<tr>
<td>12</td>
<td>389300</td>
</tr>
<tr>
<td>15</td>
<td>501000</td>
</tr>
</tbody>
</table>
3.4 Initial Profile (.pro)
CWR DYRESM-CAEDYM Initial profile
Initial vertical profile of water temperature and salinity specified at the deepest point in the water body

4 # number of initial profile points

<table>
<thead>
<tr>
<th>Height (m)</th>
<th>Temperature (Celsius)</th>
<th>Salinity (pss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>30.0</td>
<td>0.700</td>
</tr>
<tr>
<td>5.0</td>
<td>31.0</td>
<td>0.300</td>
</tr>
<tr>
<td>8.0</td>
<td>30.0</td>
<td>0.700</td>
</tr>
<tr>
<td>12.0</td>
<td>31.0</td>
<td>0.700</td>
</tr>
</tbody>
</table>
3.5 Meteorological data (.met)

<#3
Comment line: dy_basic test case - daily met forcing
86400 # met data input time step (seconds)
CLOUD COVER # longwave radiation type (NETT_LW, INCIDENT_LW, CLOUD_COVER)
FIXED_HT 10 # sensor type (FLOATING; FIXED_HT), height in m (above water surface; above lake bottom)

<table>
<thead>
<tr>
<th>YearDayNum</th>
<th>SW_{[W/m2]}</th>
<th>Cloud-Cover</th>
<th>Tair_{[C]}</th>
<th>Vapour_Press_{[hPa]}</th>
<th>Wind_Speed_{[m/s]}</th>
<th>Rain_{[m]}</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008120</td>
<td>1977</td>
<td>0.5</td>
<td>34.0</td>
<td>8.7</td>
<td>0.68</td>
<td>0.03</td>
</tr>
<tr>
<td>2008121</td>
<td>2225</td>
<td>0.6</td>
<td>34.2</td>
<td>9.1</td>
<td>0.55</td>
<td>0.01</td>
</tr>
<tr>
<td>2008122</td>
<td>1177</td>
<td>0.5</td>
<td>34.0</td>
<td>6.7</td>
<td>0.32</td>
<td>0.05</td>
</tr>
<tr>
<td>2008123</td>
<td>625</td>
<td>0.7</td>
<td>32.2</td>
<td>10.4</td>
<td>0.53</td>
<td>0.0</td>
</tr>
<tr>
<td>2008124</td>
<td>2334</td>
<td>0.3</td>
<td>33.4</td>
<td>11.2</td>
<td>0.75</td>
<td>0.0</td>
</tr>
<tr>
<td>2008125</td>
<td>809</td>
<td>0.5</td>
<td>30.2</td>
<td>9.9</td>
<td>0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>2008126</td>
<td>1864</td>
<td>0.6</td>
<td>31.0</td>
<td>7.4</td>
<td>0.41</td>
<td>0.01</td>
</tr>
<tr>
<td>2008127</td>
<td>1876</td>
<td>0.6</td>
<td>29.5</td>
<td>8.8</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>2008128</td>
<td>1681</td>
<td>0.7</td>
<td>33.0</td>
<td>8.5</td>
<td>0.47</td>
<td>0.0</td>
</tr>
<tr>
<td>2008129</td>
<td>1542</td>
<td>0.8</td>
<td>32.0</td>
<td>8.8</td>
<td>0.47</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Cloud-cover fraction: valid range is [0,1]
### 3.6 Stream Inflow (.inf)

CWR DYRESM-CAEDYM Inflow file

2  # number of inflow streams

<table>
<thead>
<tr>
<th>CW</th>
<th># stream 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>UB</td>
<td># stream 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>YearDayNumber</th>
<th>Inflow Number</th>
<th>Volume</th>
<th>Temperature</th>
<th>Practical Salinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008120</td>
<td>1</td>
<td>72672</td>
<td>30.30</td>
<td>0.040</td>
</tr>
<tr>
<td>2008121</td>
<td>2</td>
<td>15048</td>
<td>30.01</td>
<td>0.039</td>
</tr>
<tr>
<td>2008122</td>
<td>1</td>
<td>88416</td>
<td>30.37</td>
<td>0.043</td>
</tr>
<tr>
<td>2008123</td>
<td>2</td>
<td>1572.8</td>
<td>30.30</td>
<td>0.039</td>
</tr>
<tr>
<td>2008124</td>
<td>1</td>
<td>84160</td>
<td>30.30</td>
<td>0.045</td>
</tr>
<tr>
<td>2008125</td>
<td>2</td>
<td>5897.6</td>
<td>30.54</td>
<td>0.033</td>
</tr>
<tr>
<td>2008126</td>
<td>1</td>
<td>99904</td>
<td>30.30</td>
<td>0.043</td>
</tr>
<tr>
<td>2008127</td>
<td>2</td>
<td>16122.4</td>
<td>30.30</td>
<td>0.041</td>
</tr>
<tr>
<td>2008128</td>
<td>1</td>
<td>96790.8</td>
<td>30.20</td>
<td>0.041</td>
</tr>
<tr>
<td>2008129</td>
<td>2</td>
<td>16747.2</td>
<td>30.63</td>
<td>0.044</td>
</tr>
<tr>
<td>2008130</td>
<td>1</td>
<td>77677.7</td>
<td>30.03</td>
<td>0.043</td>
</tr>
</tbody>
</table>

pss : practical salinity scale

Stream 1 = Upper Bisa
Stream 2 = Central Wetlands
3.7 Withdrawals (.wdr)

Comment line: Daily withdrawals [m³/day]
1 # number of withdrawal outlets

<table>
<thead>
<tr>
<th>YrDayNum</th>
<th>Langat_River</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008120</td>
<td>182491</td>
</tr>
<tr>
<td>2008121</td>
<td>182491</td>
</tr>
<tr>
<td>2008122</td>
<td>182491</td>
</tr>
<tr>
<td>2008123</td>
<td>182491</td>
</tr>
<tr>
<td>2008124</td>
<td>182491</td>
</tr>
<tr>
<td>2008125</td>
<td>182491</td>
</tr>
<tr>
<td>2008126</td>
<td>182491</td>
</tr>
<tr>
<td>2008127</td>
<td>182491</td>
</tr>
<tr>
<td>2008128</td>
<td>182491</td>
</tr>
<tr>
<td>2008129</td>
<td>182491</td>
</tr>
</tbody>
</table>

Note: Withdrawal was set to a constant value to reflect the consistency of water withdrawal that was in operation at Lake Putrajaya.
3.8 Parameter file (.par)

Dyresm Parameters File for DYCD V4.0.0

1.30E-03  # bulk aerodynamic mmt. transport coeff  (priv. comm. [Imberger, 1998])
0.08  # mean albedo of water
0.96  # emissivity of a water surface  (Imberger & Patterson [1981, p316])
3  # critical wind speed [m s^-1]
54000  # time of day for output (secs from midnight) (54000 s = 15:00 HR)
0.012  # bubbler entrainment coefficient  (priv. comm. [Alexander, 2000])
0.083  # buoyant plume entrainment coefficient [Fischer et al. 1979]
0.06  # shear production efficiency  (eta_K)
0.2  # potential energy mixing efficiency  (eta_P)
0.4  # wind stirring efficiency (eta_S)
1.00E+07  # effective surface area coefficient (priv. com. [Yeates, 2002])
1.40E-05  # BBL detrainment diffusivity  (priv. com. [Yeates, 2002])
200  # vertical mix coefficient  (priv. com. [Yeates, 2002])

^a does not play role in Putrajaya Lake - no bubble plume diffusers or submerged inflow pipes are present
^b effective surface area coefficient represents wind sheltering (lake surface contour)

Rule of thumb for selecting the minimum and maximum layer thicknesses:

The maximum layer thickness must be at least 2x min thickness. The min thickness is really based on the overall depth of the lake and the resolution you require-usually between 0.5-1.0m, thinner for wetland type environment though. Generally we have found (min, max) tuple of approx. (1, 2) meters to produce reasonable results. There are some issues relating to the energetic if the layer thickness limits are 'too' low (e.g., min=0.5m). Small thicknesses mean more layers and more 'layer-activity' during model operation (splitting and merging).
### 3.10 Field data file (.fld)

<table>
<thead>
<tr>
<th>Yr Jul day</th>
<th>layers</th>
<th>2008120</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>WTR_ TEMP</td>
<td>DO</td>
<td>PH</td>
</tr>
<tr>
<td>0.3</td>
<td>17.2</td>
<td>8.5</td>
<td>8.44</td>
</tr>
<tr>
<td>2.8</td>
<td>17</td>
<td>8.3</td>
<td>8.48</td>
</tr>
<tr>
<td>5.7</td>
<td>16.7</td>
<td>7.9</td>
<td>8.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yr Jul day</th>
<th>layers</th>
<th>2008121</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>WTR_ TEMP</td>
<td>DO</td>
<td>PH</td>
</tr>
<tr>
<td>0.3</td>
<td>19.4</td>
<td>7.5</td>
<td>7.88</td>
</tr>
<tr>
<td>2.8</td>
<td>18.1</td>
<td>7.4</td>
<td>7.86</td>
</tr>
<tr>
<td>5.7</td>
<td>17.9</td>
<td>6.3</td>
<td>7.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yr Jul day</th>
<th>layers</th>
<th>2008122</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>WTR_ TEMP</td>
<td>DO</td>
<td>PH</td>
</tr>
<tr>
<td>0.3</td>
<td>23.2</td>
<td>8.9</td>
<td>7.97</td>
</tr>
<tr>
<td>2.8</td>
<td>21.8</td>
<td>7.2</td>
<td>7.7</td>
</tr>
<tr>
<td>5.7</td>
<td>20.3</td>
<td>3.5</td>
<td>7.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yr Jul day</th>
<th>layers</th>
<th>2008123</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPTH</td>
<td>WTR_ TEMP</td>
<td>DO</td>
<td>PH</td>
</tr>
<tr>
<td>0.3</td>
<td>23.2</td>
<td>8.9</td>
<td>7.97</td>
</tr>
<tr>
<td>2.8</td>
<td>21.8</td>
<td>7.2</td>
<td>7.7</td>
</tr>
<tr>
<td>5.7</td>
<td>20.3</td>
<td>3.5</td>
<td>7.19</td>
</tr>
</tbody>
</table>