INVESTIGATING THE APPLICABILITY OF ECOLOGICAL INFORMATICS MODELLING TECHNIQUES FOR PREDICTING HARMFULL ALGAL BLOOMS IN HYPERTROPHIC RESERVOIRS OF SOUTH AFRICA

CE VAN GINKEL

Thesis submitted for the degree of
Doctor of Philosophy
North-West University
Potchefstroom

Promotor: Dr. S du Plessis

November 2007

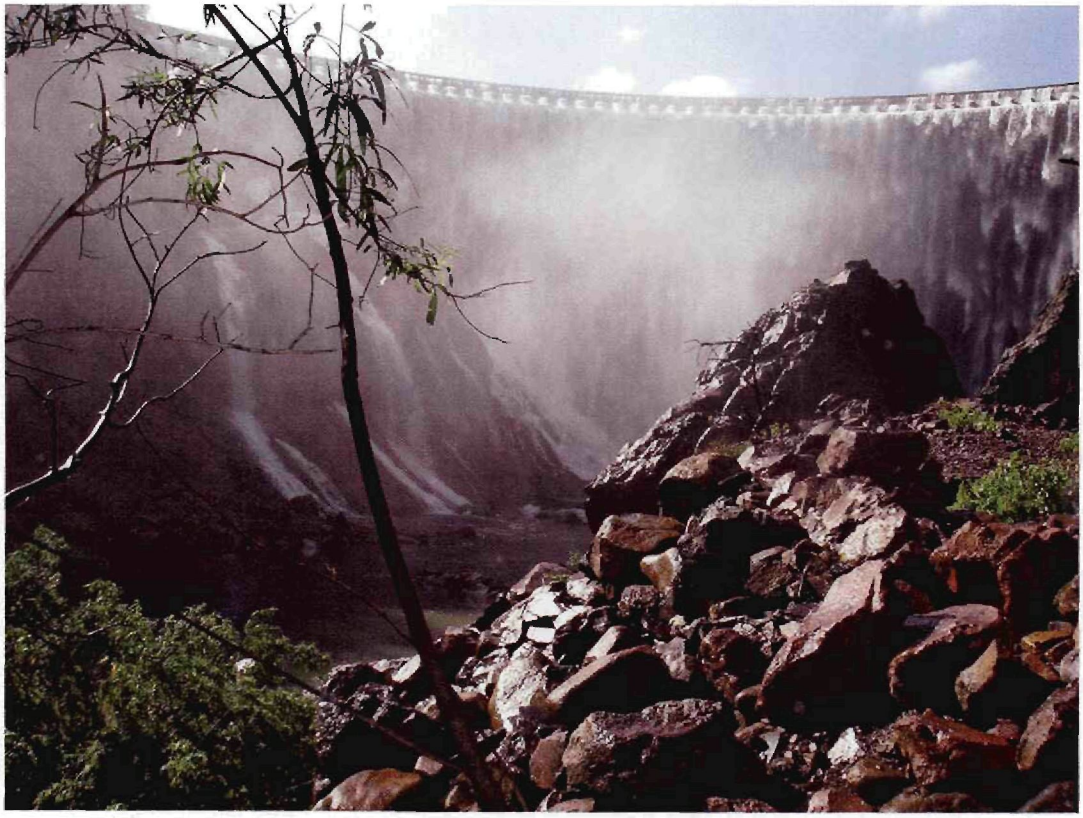


TABLE OF CONTENTS

Summary i

Opsomming vi

Acknowledgements xi

List of Figures xii

List of Tables xxiii

List of Abbreviations xxvi

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: LITERATURE SURVEY 4

2.1 Eutrophication Impacts and Management 4

2.1.1 Sources and flow characteristics of nutrients 5

2.1.1.1 Phosphorous 6

2.1.1.2 Nitrogen 8

2.1.2 Cyanobacterial Bloom impacts 10

2.1.2.1 Impacts on Human Health 11

2.1.2.2 Impacts on Water Treatment Works 14

2.1.2.3 Impacts on the Environment 14

2.1.2.4 Impacts on Recreation 15

2.1.2.5 Impacts on Agriculture 15

2.2 Characteristics of Cyanobacteria 15

2.2.1 Morphology and Physiology of Cyanobacteria 15

2.2.2 Bloom Forming Conditions for Cyanobacteria 19

2.2.2.1 Low mixing 21

2.2.2.2 High Levels of Phosphorus 21

2.2.2.3 Low z_{eu}:z_{mix} Ratios 22

2.2.2.4 Redox Processes 22

2.2.2.5 High Ammonium Conditions 23

2.2.2.6 High Temperature 24

2.3 Impacts of Climate and Weather on Freshwater Systems 25

2.3.1 Climate, Weather and Changes 25

2.3.2 Climate and Weather Impacts on Phytoplankton 26
2.3.3 Modelling Climate and Biological Processes 27

CHAPTER 3: SAMPLING SITES 30
3.1 Location of the Sampling Sites 30
3.2 Climatic conditions of the Sampling Sites 31
3.3 Previous Studies and Background of the Sampling Sites 32
3.3.1 Bon Accord Dam 34
3.3.2 Hartbeespoort Dam 35
3.3.3 Klipvoor Dam 37
3.3.4 Rietvlei Dam 39
3.3.5 Roodeplaat Dam 40

CHAPTER 4: DATA COLLECTION AND ANALYSIS 42
4.1 Limnological and climatological Data Collection Methods 42
4.1.1 Limnological Sampling 42
4.1.2 Determination of the Depth Distribution of Cyanobacterial Toxins 44
4.1.3 Climatological Data Collection 44
4.1.4 Multivariate Statistical Analysis 45
4.2 Results and Discussion 46
4.2.1 Limnological Characteristics and Climatic Conditions of the Sampling Sites 46
4.2.2 Phytoplankton Composition of the Sampling Sites 49
4.2.3 Total microcystin occurrence in the Reservoirs 56
4.2.4 Temporal and Spatial Depth Distribution of Total Microcystins 62
4.2.5 Multivariate Analysis 65
4.3 Conclusions 75

CHAPTER 5: MODELLING TECHNIQUES 77
5.1 Modelling Methods 77
5.1.1 Vollenweider Model 77
5.1.2 SALMO-00 Model 78

5.1.3 Artificial Neural Networks 81
5.1.3.1 Multilayered Feed forward Neural Network (MFNN) 81
5.1.3.1.1 NeuroSolutions 5.04 83
5.1.3.1.2 Forecaster XL 83
5.1.3.2 Self-Organising Map (SOM) 84
5.1.4 Rule Development by Hybrid Evolutionary Algorithms (HEA) 86

5.2 Modelling Results 91
5.2.1 Vollenweider Model 91
5.2.2 Lake Simulation Library of SALMO-00 93
5.2.3 Artificial Neural Networks 97
5.2.3.1 Result with Multilayered Feed forward Neural Networks (MFNN) 97
5.2.3.1.1 NeuroSolutions 5.04 97
5.2.3.1.2 Forecaster XL 99
5.2.3.1.3 Comparison of the Forecaster XL experiments 129
5.2.3.2 Self-Organising Map (SOM) 133
5.2.4 Rule Ste Development by Hybrid Evolutionary Algorithms 137
5.2.4.1 Experiment 1: Real-time prediction simulation and rule-set discovery for *Microcystis* including Chl *a* as an environmental factor 137
5.2.4.2 Experiment 2: Real-time prediction simulation and rule-set discovery for *Microcystis* excluding Chl *a* as an environmental variable 142
5.2.4.3 Experiment 3: 7-Days-forward prediction simulation and rule set discovery for *Microcystis* 146
5.2.4.4 Experiment 4: 14-Days-forward prediction simulation and rule set discovery for *Microcystis* 150
5.2.4.5 Experiment 5: 21-Days-forward prediction simulation and rule set discovery for *Microcystis* 154
5.2.4.6 Experiment 6: 28-Days-forward prediction simulation and rule set discovery for *Microcystis* 158

5.2.4.7 Experiment 7: Real-Time Simulation and Rule Set Discovery for *Ceratium* 162

5.3 Conclusions and Evaluation 170

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS 175

CHAPTER 7: REFERENCES 180

APPENDICES See attached CD

SUMMARY

Introduction

The occurrence and extent of harmful algal blooms (especially cyanobacteria and dinoflagellates) seem to be increasing worldwide. For water resource managers to be prepared for these events, and to enable managers to warn potable industry and recreational users of these incidents, it is essential to develop predictive capability as a part of the management actions.

Hypothesis

A number of hypothesis were set before the commencement of the study to be included in the investigation and the development of a predictive tool for harmful algal blooms:

1) A holistic approach is required to accommodate different levels of external (climate) and internal (trophic linkages and benthic-pelagic coupling) variability and their interactions. Therefore, the integration of biological and meteorological information is needed to understand and estimate the consequences of climate variations on the development of toxic cyanobacterial blooms.

   A detailed literature survey was done to determine the most important physical, chemical and biological factors in the development of harmful algal blooms. The literature survey also investigated the potential impact of climatic conditions on the development of harmful algal blooms, with special reference to cyanobacteria, which is the largest noxious phytoplankton problem within South African fresh water systems.

   This hypothesis also led to the inclusion of a number of climatic variables in the statistical analysis of the collected data, to determine the potential use of climatic variables in predicting harmful algal blooms.

2. A water body is one entity and when a cyanobacterial bloom does occur, there are toxins to be found in the system.

This hypothesis aimed to investigate the distribution and potential for toxic cyanobacteria that may impact on recreational users, even though the regular sampling site does not indicate any or low cyanobacterial toxicity within the water column.

3. Ecological Modelling methods are available to be tested and used to predict harmful algal blooms.

   This hypothesis was tested and proved to be quite applicable in the hypertrophic reservoirs of South Africa.

Results and Discussion

The hypertrophic reservoirs, Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat, situated in South Africa are, warm monomictic reservoirs, downstream of the most populated areas of South Africa. These reservoirs are known to experience annual cyanobacterial blooms of especially, the toxin producing cyanobacterium, *Microcystis aeruginosa*. These reservoirs lie within similar climatic conditions, with warm wet summers and dry, fairly cold winters. With this study in mind, five years of monitoring to determine the phytoplankton community trends and the presence of cyanobacterial toxins was initiated. During these five years Hartbeespoort and Roodeplaat were dominated primarily by *Microcystis aeruginosa*. Bon Accord, Klipvoor and Rietvlei reservoirs experienced both *Microcystis aeruginosa* and *Ceratium hirundinella* blooms annually.

Total microcystin (TM) concentrations were found to be orders of a magnitude higher in South Africa (> 10 000ug/L) than in other countries (between 10ug/L and 100ug/L). The presence of TM's in all the reservoirs was primarily associated with the dominance and blooms of *Microcystis*. The depth distribution of the toxins measured in the Hartbeespoort and Roodeplaat Reservoirs indicated that during the periods of excessive *Microcystis* biovolume, toxins are often found all the way through the water column.

Multivariate analyses of the reservoir's data indicated that the five reservoirs are similar in both algal community and physico-chemical variables. The multivariate analyses showed that of the environmental factors, temperature is the most important factor, and can be used as an indicator of climatic conditions, in the development of cyanobacteria biovolume in these systems. Other environmental variables important to the development of algal blooms are

dissolved inorganic phosphorous (DIP), dissolved inorganic nitrogen (DIN), the DIN:DIP ratio; total phosphorous (TP), total nitrogen (TN) and Chl *a* concentration.

A number of generic and deterministic ecological models were tested on the data to determine their applicability for predicting harmful algal blooms in hypertrophic reservoirs in the northern central parts of South African. A summary of the key findings are as follows:

1) The relatively simple Vollenweider model is easy to apply and provides a manager with a quick answer. Relatively little information is needed to apply the model. It also provides the manager with the possibility of testing different management scenarios.

2) The simulation library SALMO-OO allows forecasting abundances of cyanobacteria, green algae and diatoms in response to eutrophication control scenarios. It takes the complex limnological characteristics of reservoirs into consideration and it supplies the manager with a tool to test different management scenarios to assist in decision-making. The results were, however, only partly successful with large over and under predictions of nutrients and algal groups, even after the growth equations were adapted.

3) Artificial neural network modelling techniques, both the supervised multilayered feed forward neural network and the non-supervised self-organising map were tested for the applicability to predict algal blooms in South African hypertrophic reservoirs.

   The multi-layered feed forward neural network model tested, provides visual predicted successes, but the strict tolerances set by the model, to determine acceptable prediction as part of the outcome of the model, may be a problem to validate the results and ensure that an acceptable amount of good predictions were found.

   The Self Organising Mapping (SOM) method technique is applicable to investigate before and after impact scenarios. This is more of a knowledge development tool than a predictive tool.

4) Finally,the Hybrid Evolutionary Algorithm (HEA) method was used to develop algorithms for algal bloom prediction. The RULE set discovery by HEA was tested on longterm data in three reservoirs using different scenarios of real time, 7-days

forward, 14-days forward, 21-days forward and 28-days forward forecasting respectively of the abundance of the cyanobacterium, *Microcystis*. The developed rule sets are highly applicable to the hypertrophic reservoirs of South Africa. These methods however need, to be tested in other South African reservoirs to determine the applicability under different trophic status and different climatic conditions.

The same method was used to develop a real time algorithm for the dinoflagellate, *Ceratium hirundinella*. The developed RULE set was then tested on the data of two "unseen" reservoirs (Bon Accord and Klipvoor) that both experienced extreme *Ceratium hirundinella* blooms during the study period. This application was found to be highly applicable to these reservoirs. This suggests that the developed RULE set may potentially also be applicable to reservoirs in other climatic areas of South Africa.

Recommendations:

The study showed that eutrophication and the associated problems is a real threat to South African fresh water resources but that modelling methods do exist to assist in managing the problem. The list of recommendations need to be taken further by a number of stakeholders, e.g. the Department of Water Affairs and Forestry, future CMA's, Universities and other researchers:

a) Determine the necessary variables and monitor these for future modelling exercises.
b) Include total microcystin monitoring in impacted fresh water resources at least during the summer periods to enable resource managers to issue warnings to all potential impacted stakeholders.
c) Initiate and test available management options to minimise serious eutrophication levels in South Africa.
d) Manage the risk imposed by the cyanobacterial blooms and the associated toxins produced in the water resources, on drinking water facilities and the health of recreational users.
e) In view of the successes of the modelling results develop short-term forecasting tools, for the algal blooms of *Microcystis* and *Ceratium,* with on-line water quality monitoring for early-warning and real-time forecasting for reservoir managers.
f) The cause and effects for the changing composition of the phytoplankton of these five reservoirs need to be investigated.

g) Monitor at different depths to determine the best depth for abstraction for treatment purposes.

h) Develop capacity in South Africa to use the Hybrid Evolutionary Algorithm (HEA) RULE set development in all research spheres, as the method is applicable to any type of numerical data.

OPSOMMING

Inleiding

Die voorkoms en intensiteit van skadelike alg opbloeie (veral blou-groen bakterie en dinoflagellate) is skynbaar aan die toeneem wereldwyd. Dit is dus noodsaaklik vir waterbron bestuurders om gereed te wees vir hierdie insidente sodat die drinkwater industrie en ontspanning gebruikers gewaarsku kan word. Daar is dus 'n behoefte om voorspelling kapasiteit te ontwikkel.

Hipoteses

'n Aantal hipotesise is gestel voor die begin van die studie om die ontwikkeling van 'n voorspellings model vir skadelike alg opbloeie te ondersoek:

1) A holistiese benadering is nodig om veskillende vlakke van eksterne (klimaat) en interne (trofiese skakels en benties-pelagiese koppelings) veranderlikheid en hulle interaksies te akkomodeer. Die integrasie van biologiese en meteorologiese inligting is dus nodig om die gevolge van klimaats veranderinge op die ontwikkeling van toksiese blou-groen bakteriee opbloeie te bepaal en te verstaan.

   'n Gedetaileerde literatuur studie is gedoen om die belangrikste fisies, chemies en biologiese faktore in die ontwikkeling van skadelike alg opbloeie te bepaal. Die literatuur studie het ook die moontlike impakte van klimaats toestande op die ontwikkeling van skadelike alg opbloeie, met spesifieke fokus op die blou-groen bakteriee, wat die enkel, mees belangrikste probleem fitoplankton spesie in Suid Afirka se varswatersisteme is, ondersoek.

   Die hipotese het ook gelei tot die insluiting van 'n aantal klimaats veranderlikes in die statistiese verwerking van die data, om die belangrikste klimaats veranderlikes te bepaal in die voorspelling van skadelike alg opbloeie.

2) 'n Waterliggaam is een entiteit en wanner 'n blou-groen bakterie opbloei teenwoordig is, is daar toksiene in die sisteem.

Hierdie hipotese is gemik om die verspreiding van toksiese blou-groen bakteriee wat 'n inpak mag he op ontspannings gebruikers te bepaal, selfs al toon die gereelde moniterings punt geen of min toksisiteit nie.

3) Ekologiese modellerings metodes is beskikbaar om getoets te word in die gebruik om skadelike algal opbloeie te voorspel.

   Hierdie hipotese is getoets en die metodes was baie suksesvol in die hipertrofe sisteme van Suid Afrika.

Bespreking en Gevolgtrekkings

Die hiper-eutrofe damme, Bon Accord, Hartbeespoort, Klipvoor, Rietvlei en Roodeplaat, in Suid Afrika, is warm, monomiktiese, damme, stroom-af van die digs bewoonde areas in Suid Afrika. Hierdie damme is bekend daarvoor dat hulle jaarliks blou-groen bakterieë opbloeie van veral die toksien produserende cyanobakterium, *Microcystis aeruginosa.* Hierdie damme is geleë in ooreenstemmende klimaats toestande, met nat somers en droeë, koel winters. Vyf jaar se monitering is geinisieer om die komposisie veranderinge van die fitoplankton en die teenwoordigheid van blou-groen bakterieë toksiene te bepaal. Tydens die vyf jaar studie was Hartbeespoort en Roodeplaat gedomineer met hoofsaaklik *Microcystis aeruginosa.* Bon Accord, Klipvoor en Rietvlei damme het jaarliks dominansie van beide *Microcystis aeruginosa* en *Ceratium hirundinella* ondervind.

Dit is bevind dat totale mikrosistiene (TM) konsentrasies in Siud Afrika orde grootes hoër is (> 10 000 ug/L) as in ander lande (tussen 10 ug/L en 100 ug/L). Die teenwoordigheid van TM was in al die damme hoofsaaklik geassosieer met die dominansie en opbloeie van *Microcystis*. Die diepte verspreiding van die TM is in Hartbeespoort en Roodeplaat bepaal. Dit het aangedui dat tydens intense *Microcystis* opbloeie, die toksiene soms dwarsdeur die water kolom aangetref word.

Veelvuldige veranderlike analises van die damme se data het aangetoon dat die vyf damme soortgelyke alg gemeenskappe asook vergelykbare fiesies-chemiese toestande het. Die veelvuldige veranderlike analises het getoon dat van die omgewingsfaktore, temperatuur die belangrikste faktor is, en as indikator van klimaat kondisies, bydra tot die ontwikkeling van blou-groen bakterieë biovolumes in hierdie sisteme. Ander omgewingsfaktore wat belangrik is in die ontwikkeling van die alg opbloeie, is die opgeloste, anorganiese fosfaat (DIP),

opgeloste anorganiese stikstof (DIN), die DIN:DIP verhouding; totale fosfaat (TP), totale stikstof (TN) en die Chl *a* konsentrasie.

'n Aantal generiese en deterministiese modelle is getoets om die toepaslikheid vir die voorspelling van skadelike alg opbloeie in Suid Afikaanse kondisies te bepaal. Die resultate is opsommend as volg:

(1) Die redelik eenvoudige Vollenweider Model is maklik om te gebruik en voorsien die bestuurder met 'n vinnige andwoord. Redelik min informasie is nodig om die model te gebruik. Dit voorsien die bestuurder ook van 'n metode om verskillende scenarios te toets.

(2) Die simulasie biblioteek SALMO-OO stel die bestuurder instaat om voorspelling te maak aangaande die blou-groen bakterieë, groen alge en diatome dinamika na aanleiding van eutrofikasie bestuurs scenarios. Die model neem die komplekse limnologiese eienskappe van die damme in aanmerking en dit voorsien die bestuurder met 'n model om verskillende bestuurs opsies te toets vir besluitnemings doeleindes. Die resultate was egter slegs gedeeltelik suksesvol, met groot onder en oor voorspellings, self nadat die groei kurwes aangepas was.

(3) Artifisieële neurale netwerk modellerings tegnieke, beide die gekontroleerde veelvuldige-laag voerentoe neurale netwerk en die ongekontroleerde self-organiserende ontwerp is getoets.

Die veelvuldige-laag voerentoe neurale netwerk model wat getoets is het visuele voorspellings suksesse, maar die streng toleransies wat deur die model gebruik word om die sukses van die voorspellings model te bepaal, kan 'n probleem wees in die validasie van die model en of daar 'n aanvaarbare aantal voorspellings gevind word.

Die ongekontroleerde self-organiserende ontwerp (SOM) metode is toepaslik vir die toetsing van voor en na impak scenarios. Dit is meer 'n metode om kennis te ontwikkel as wat dit gebruik kan word vir voorspellings.

(4) Laastens is die Hebriede Evolutionere Algoritme (HEA) gebruik om algoritmes te ontwikkel om alg opbloei voorspellings te maak. Die REËL ontwikkeling van die HEA is getoets op verskillende tyd scenarios. Werklike tyd, 7-dae vooruit, 14-dae vooruit, 21-dae vooruit en 28-dae vooruit voorspellings vir die biomassa ontwikkeling van die blou-groen

bakterieë, *Microcystis.* The ontwikkelde REËL is toepaslik op die hiper-eutrofe damme van Suid Afrika. Die metode metode moet egter ook in ander damme getoets word om die toepaslikheid daarvan onder ander trofiese toestande en ander klimaats toestande te bepaal.

Deiselfde metode is ook gebruik om 'n werklike tyd algoritme vir die dinoflagellaat, *Ceratium hirundinella,* te ontwikkel. Die ontwikkelde REËL is toe getoets op die data van twee 'ongesiende' damme (Bon Accord en Klipvoor), wat beide ook groot opbloeie van *Ceratium hirundinella,* ondervind het tydens die studie. Hierdie toepassing was baie suksesvol op die twee damme.

Hierdie bevindinge dui daarop dat die ontwikkelde REËL, ook potensieel toegepas kan word op ander damme in ander klimaats toestande.

Aanbevelings

Die studie het getoon dat eutrofikasie en verwante probleme 'n wesenlike probleem in Suid Afrikaanse water sisteme is, maar dat daar wel modellerings tegnieke beskikbaar is vir gebruik in die bestuur van hierdie sisteme. Die lys van aanbevelings behoort verder ondersoek te word deur die groep belanghebbendes, nl. Die Departement of Waterwese en Bosbou, toekomstige CMA's, Universiteite en ander navorsers:

i) Bepaal die belangrike en nodige veranderlikes wat gemonitor moet word vir verdere modellerings studies.
j) Sluit die monitering van totale microcystiene in wanner geimpakteerde wars water sisteme gemonitor word, veral gedurende die somer. Dit sal water beheer personeel in staat stel om vroegtydige waarskuwings aan alle moontlike gebruikers deur te gee.
k) Inisieer en toets beskikbare bestuurs opsies om ernstige eutrofikasie vlakke in Suid Afrika te verminder.
l) Bestuur die risiko effek van blou-groen bakteriee en die geassosieerde toksiene in die water sisteme, vir suiwerings werke asook om die impak op rekreasie gebruikers te verminder.
m) Die suksesse van die modellerings resultate kan lei tot die ontwikkeling van korttermyn en huidige tyd voorspellings en besluitnemings metodes vir *Microcystis* en *Ceratium,* deur gebruik te maak van aan-lyn monitering stelsels. Dit kan dien as 'n vroegtydige waarskuwings stelsel vir water hulpbron bestuurders.

n) Die oorsaak en impak van die veranderinge in die samestelling van die fitoplankton in die vyf damme behoort verder ondersoek te word.

o) Monitering op verskillende dieptes moet gedoen word om die beste ontrekking diepte te bepaal vir die beste water gehalte in die suiweringswerke.

p) Ontwikkel kapasiteit in Suid Afrika vir die gebruik van Hebriede Evolutionere Algoritme (HEA) REËL ontwikkeling in alle navorsings aspekte, aangesien die metode toegepas kan word met enige numeriese data.

ACKNOWLEDGEMENTS

This is to acknowledge all the people that gave support, comments and advice during the studies:

- The Department of Water Affairs and Forestry's management, who provided a bursary for the studies.
- Dr Sandra du Plessis, at North West University, who gave guidance and support during the project.
- The WRC for funding of the Australian visits.
- Dr Jaco Bezuidenhout, for his assistance with the statistical analysis of the data.
- The Steering Committee of the WRC Project, for guidance and advice given during the run of the project.
- Dr Friedrich Recknagel and his students, with assistance with the modelling techniques used for the study.
- Dr Honqging Cao, for assistance with the use of the HEA modelling techniques.
- Chris Carelsen, Alice le Grange, Doris le Roux and Rika Schillack, the biological laboratory personnel of Resource Quality Services, for their continued dedication in providing the data for the project.
- The monitoring personnel, Alfred Seloana and Johannes Phetla, for their monitoring of all the sampling sites of the study, on a regular basis.
- Marica Easmus, that was always willing to provide the requested data during the Australian visits.
- Elsa Stofberg, my mother, which was a driving force in the initiation of the studies.
- Last, but not least, Chris, Oswald, Cobus and Elsa, my family, who put up with me and for allowing me to continue and complete my studies.

LIST OF FIGURES

Page No

CHAPTER 2

Figure 2.1 The nitrogen cycle in freshwater ecosystems. The thick lines indicate the main pathways in terms of mass transfer; dotted lines, those involving recycling and mineralisation in the water column (Redrawn from Horne & Goldman 1994). 9

Figure 2.2 Cyanobacterial hyper-scum from the Hartbeespoort Dam in South Africa showing the distinct colours that depend on the pigments present in cyanobacteria. a) Green colour due to Chl *a*; b) Blue colour due to phycocyanin; and c) Red colour due to phycoerythrin. 16

Figure 2.3 The redox scale and equilibrium phases for selected nutrients and metals (redrawn from Lawrence *et al.* 2000). 23

CHAPTER 3

Figure 3.1 The location of the five Reservoirs that were used as sampling sites for the study. NW = North West Province and GP = Gauteng Province. 30

Figure 3.2 The monthly mean solar radiation in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997) 31

Figure 3.3 The mean daily air temperature in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997) 31

Figure 3.4 The means of the median monthly rainfall in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997). 32

Figure 3.5 The mean daily relative humidity in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997). 32

Figure 3.6 The depth to water volume graph of the Bon Accord Dam that is related to the depth to area relationship for the Reservoir. 35

Figure 3.7 The depth to water volume graph of the Hartbeespoort Dam that is related to the depth to area relationship for the Reservoir. 37

Figure 3.8 The depth to water volume graph of the Klipvoor Dam that is related to the depth to area relationship for the Reservoir. 38

Figure 3.9 The depth to water volume graph of the Rietvlei Dam that is related to the depth to area relationship for the Reservoir. 39

Figure 3.10 The depth to water volume graph of the Roodeplaat Dam that is related to the depth to area relationship for the Reservoir. 41

CHAPTER 4

Figure 4.1 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Bon Accord Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB 50

Figure 4.2 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Hartbeespoort Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB. 51

Figure 4.3 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Klipvoor Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB. 52

Figure 4.4 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Rietvlei Dam from October 2000 to September 2005. Cyanophyta biovolume (CyaB), Chlorophyta biovolume (ChlB) Diatom biovolume (DiatomB) and Dinoflagellate biovolume (DinoB). 54

Figure 4.5 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Roodeplaat Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB. 55

Figure 4.6 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Bon Accord Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB and *Oscillatoria* biovolume: OscB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB. 57

Figure 4.7 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Hartbeespoort Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB,

Oscillatoria biovolume: OscB and *Merismopedia* biovolume: MerB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB. 58

Figure 4.8 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Klipvoor Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, *Merismopedia* biovolume: MerB and *Pseudoanabaena* biovolume: PseuB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB. 59

Figure 4.9 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Rietvlei Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, and *Merismopedia* biovolume: MerB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB. 60

Figure 4.10 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Roodeplaat Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, *Merismopedia* biovolume: MerB and *Pseudoanabaena* biovolume: PseuB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB. 61

Figure 4.11 The temporal and spatial distribution of a) Total Microcystin concentration (ug/L), b) Chl *a* concentration (ug/L), c) Microcystis biovolume (cm^3/m^3) and d) Anabaena biovolume (cm^3/m^3) in the Hartbeespoort Dam. 63

Figure 4.12 The temporal and spatial distribution of a) Total Microcystin concentration (ug/L), b) Chl *a* concentration (ug/L), c) Microcystis

biovolume (cm^3/m^3) and d) Anabaena biovolume (cm^3/m^3) in the Roodeplaat Dam. 64

Figure 4.13 The PCA ordination diagram, showing the ordination of all the data from the five reservoirs. The sum of all the EV = 0.916. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat). 66

Figure 4.14 The PCA ordination diagram of algal community data in relation to the ordination of all the dam's data. The sum of all the EV = 0.916. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat & *Ana_B* = *Anabaena* biovolume; *ChlA* = chlorophyll *a*; *CryptoBi* = cryptophyte biovolume; *CyanoBio* = cyanophyte biovolume; *Mic_B* = *Microcystis* biovolume; *PhyrrBio* = Phyrrophyta biovolume (dinoflagellates); *Pse_B* = *Pseudoanabaena* biovolume) 67

Figure 4.15 The PCA ordination diagram of environmental data (physico-chemical data) in relation to the ordination of all the dam's data. The sum of all the EV = 0.695. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat). 68

Figure 4.16 The RDA ordination diagram, showing environmental variables and their relationships to each other. The sum of all the EV = 0.695. 69

Figure 4.17 The RDA ordination diagram, showing environmental variables and their relationships to one another. The RDA ordination diagram, showing environmental variables and their relationships to one another. Test of significance for 1st canonical axis: EV = 0.075; F ratio = 43.479; P-value = 0.0020. Test of significance of all canonical axxes: Trace = 0.142; F ratio = 2.897; P-value = 0.0020. 70

Figure 4.18 The RDA ordination diagram, with standardized data and with the algal groups and species biovolume, as response variables set and the environmental conditions as environmental variables. Test of significance for 1st canonical axis: EV = 0.073; F ratio = 43.226; P-value = 0.0020. Test of significance of all canonical axxes: Trace = 0.119; F ratio = 3.363; P-value = 0.0020. 71

Figure 4.19 The ordination diagram of PCA with non-standardized data with the algal groups and species biovolume as environmental variables and the cyanotoxin concentrations as the species. Test of significance for 1st canonical axis: EV = 0.004; F ratio = 0.551; P-value = 0.7800. Test of

significance of all canonical axxes: Trace = 0.004; F ratio = 0.192; P-value = 0.8200. 72

Figure 4.20 The RDA ordination diagram showing the reduced environmental variables and the cyanotoxin concentrations, cyanobacterial biovolume and species biovolume as the response variables. Test of significance for 1st canonical axis: EV = 0.069; F ratio = 10.432; P-value = 0.0180. Test of significance of all canonical axxes: Trace = 0.094; F ratio = 3.667; P-value = 0.0140. 73

Figure 4.21 The RDA ordination diagram of species and environmental data. All environmental data collected from the five reservoirs during the study period, was used. Test of significance for 1st canonical axis: EV = 0.075; F ratio = 43.479; P-value = 0.0020. Test of significance of all canonical axxes: Trace = 0.142; F ratio = 2.897; P-value = 0.0020. 74

CHAPTER 5

Figure 5.1 The structure of the supervised MFNN as used in this study (redrawn from Recknagel *et al.* 2006). 82

Figure 5.2 The structure of the non-supervised SOM for ordination and clustering of inputs (redrawn from Recknagel *et al.* 2006). 84

Figure 5.3 The non-supervised SOMs ordination and clustering maps shown as a) a distance matrix map and b) a partitioned map (Redrawn from Recknagel *et al.* 2006). 85

Figure 5.4 Conceptual diagram of HEA for the discovery of a predictive rule set for *Microcystis* biovolume in three hypertrophic South African reservoirs. The same model was used for the real time prediction of the dinoflagellate biovolume predictive rule set. 87

Figure 5.5 Flow chart showing the process of HEA (redrawn from Cao *et al.* 2006). 88

Figure 5.6 Comparison of output results (line) with real measured data (markers) for the Hartbeespoort Reservoir with the original SALMO-00 growth and grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available. 94

Figure 5.7 Comparison of output results (line) with real measured data (markers) for the Klipvoor Reservoir with the original SALMO-00 growth and

grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available. 95

Figure 5.8 Comparison of output results (line) with real measured data (markers) for the Roodeplaat Reservoir with the original SALMO-00 growth and grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available. 96

Figure 5.9 Results from the NeuroSolutions 5.04 ANN for the cyanotoxin producing group (CTTot exp), showing real time forecasting on the Roodeplaat Dam data. 98

Figure 5.10 Results from the NeuroSolutions 5.04 ANN for the Non-CT producing group (TotNONCT), showing real time forecasting (pink line) versus the actual data (blue line). 99

Figure 5.11 Actual (orange line) versus forecasted (blue line) results for Non-CT producing group bio-volume using only environmental variables and Chl *a* concentration as input with the Forecaster XL ANN software package. 100

Figure 5.12 Percentage contributions of measured environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group in the Hartbeespoort Dam. 101

Figure 5.13 Actual (orange line) versus forecasted (blue line) results for CT`producing group using only environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package on the Hartbeespoort Dam data. 102

Figure 5.14 Percentage contributions of measured environmental parameters and Chl *a* concentration to ANN prediction of the CT group. 104

Figure 5.15 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package. 105

Figure 5.16 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group. 106

Figure 5.17 Actual (orange line) versus forecasted (blue line) results for the CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package. 108

Figure 5.18 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the CT group. 109

Figure 5.19 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package. 110

Figure 5.20 Percentage contributions of selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group. 112

Figure 5.21 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables as inputs for the Forecaster XL ANN Software package. 113

Figure 5.22 Percentage contributions of selected environmental parameters and Chl *a* concentration to ANN prediction for the CT group. 114

Figure 5.23 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only environmental variables as inputs inputs for the Forecaster XL ANN Software package. 116

Figure 5.24 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT producing group in all the dams. 117

Figure 5.25 Actual (orange line) versus forecasted (blue line) results for the CT producing algae group using environmental variables and algae counts as inputs for the Forecaster XL ANN Software package. 118

Figure 5.26 Percentage contributions of algal dominance data, selected environmental parameters and Chl *a* concentration as result of the Forecaster XL forecasting of the CT producing group. 119

Figure 5.27 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only selected environmental variables as inputs for the Forecaster XL ANN Software package. 120

Figure 5.28 Percentage contributions of the selected environmental parameters to Forecaster XL forecasts for the Non-CT group. 121

Figure 5.29 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables as inputs for the Forecaster XL ANN Software package. 122

Figure 5.30 Percentage contributions of elected environmental parameters to Forecaster XL forecast for the CT producing group. 124
Figure 5.31 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only environmental variables and algal counts as inputs for the Forecaster XL ANN Software package. 125
Figure 5.32 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group. 126
Figure 5.33 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables and algal counts as inputs for the Forecaster XL ANN Software package. 127
Figure 5.34 Percentage contributions of the algal dominance and the selected environmental parameters to ANN prediction for the CT group.a 128
Figure 5.35 The ordination and clustering map of the PO_4-P (DIP) and TP concentrations in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WWTW that show the changes in the water quality. 133
Figure 5.36 The ordination and clustering map of the Chl *a* compared to the Chlorophyta dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WWTW that show the changes in the water quality. 134
Figure 5.37 The ordination and clustering map of the Chl *a* compared to the Cyanophyta dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WWTW that show the changes in the water quality. 135
Figure 5.38 The ordination and clustering map of the Chl *a* compared to the diatom dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WWTW that show the changes in the water quality. 136
Figure 5.39 Training results of the real-time forecasting of *Microcystis* biovolume including Chl *a* as an environmental variable in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the rule set for HEA Experiment 1. 138
Figure 5.40 Testing results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the rule set as shown in Figure 5.41 139

Figure 5.41 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for real-time forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 141

Figure 5.42 Training results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 2 that excluded Chl *a* as an environmental factor. 142

Figure 5.43 Testing results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.44 excluding Chl *a* as an environmental factor. 144

Figure 5.44 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for real-time forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables (excluding Chl *a*) to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 145

Figure 5.45 Training results of the 7-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the developed rule set of experiment 3. 146

Figure 5.46 Testing results of the 7-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.43. 147

Figure 5.47 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 7-days forward forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 148

Figure 5.48 Training results of the 14-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 4, including Chl *a* as an environmental factor. 150

Figure 5.49 Testing results of the 14-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat

reservoirs using the best developed RULE set as shown in Figure 5.50 including Chl *a* as an environmental variable. 151

Figure 5.50 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 14-days forward forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 152

Figure 5.51 Training results of the 21-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 5 including Chl *a* as an environmental factor. 154

Figure 5.52 Testing results of the 21-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best developed rule set as shown in Figure 5.53 including Chl *a* as an environmental variable. 155

Figure 5.53 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 21-days ahead forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 156

Figure 5.54 Training results of the 28-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 6 including Chl *a* as an environmental factor. 158

Figure 5.55 Testing results of the 28-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best developed rule set as shown in Figure 5.56 including Chl *a* as an environmental variable. 159

Figure 5.56 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 28-days ahead forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable. 160

Figure 5.57 Training results of the real-time forecasting of *Ceratium* biovolume (DinoB), in the Hartbeespoort, Roodeplaat and Rietvlei reservoirs during the development of the best rule set for HEA Experiment 7 including Chl *a* as an environmental variable. 163

Figure 5.58 Testing results of the real-time forecasting of *Ceratium* biovolume (DinoB), in the Hartbeespoort, Roodeplaat and Rietvlei reservoirs using the rule set as shown in Figure 5.59, including Chl *a* as an environmental variable.. 164

Figure 5.59 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the rule set for real-time Dinoflagellate biovolume forecasting. 165

Figure 5.60 Testing of the developed rule set on Bon Accord Dam for real-time Dinoflagellate (*Ceratium*) biovolume forecasting. 167

Figure 5.61. Testing of the developed rule set on Klipvoor Dam for real-time *Ceratium* biovolume (DinoB) forecasting. 168

LIST OF TABLES

Page No.

CHAPTER 3

Table 3.1 Morphological and hydrological characteristics of Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs (Department of Water Affairs 1980, Stofberg 1983, Department of Water Affairs 1986) 33

CHAPTER 4

Table 4.1 Limnological properties of the Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs during 200 to 2005 48

CHAPTER 5

Table 5.1 Method to determine the constant parameters included for the Vollenweider Model. 78

Table 5.2. Methods to determine the constant parameters included in SALMO. 79

Table 5.3: Methods to determine the input variables included in SALMO. 80

Table 5.4 The classification of seasons for the SOM modelling. 86

Table 5.5 Parameter settings of the hybrid evolutionary algorithm rule set discovery for the Hartbeespoort, Rietvlei and Roodeplaat reservoirs. 89

Table 5.6 Input variables for the Vollenweider Model with the results found by applying three different management scenarios in the hypertrophic Hartbeespoort, Klipvoor and Roodeplaat dams. 92

Table 5.7 Summary of the initial neural network results using NeuroSolutions for prediction of the cyanotoxin producing group (CTTot exp) and the non-cyanotoxin producing group (TotNONCT). 98

Table 5.8 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using only environmental variables and Chl *a* concentration as inputs. 101

Table 5.9 A summary of the modelling result output with the Forecaster XL ANN software package for the cyanotoxin (CT) producing group using only environmental variables and Chl *a* concentration as input. 103

Table 5.10 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs. 106

Table 5.11 A summary of the modelling result output with the Forecaster XL ANN software package for the CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs. 109

Table 5.12 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using selected environmental variables d Chl *a* as inputs. 111

Table 5.13 A summary of the modelling result output with the Forecaster XL ANN software package for the CT producing group using selected environmental variables and Chl *a* as inputs. 114

Table 5.14 A summary of the modelling results for the Non-CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs. 116

Table 5.15 A summary of the modelling results for CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs. 119

Table 5.16 A summary of the modelling results for the Non-CT producing group using only environmental variables as inputs. 121

Table 5.17 A summary of the modelling results for the CT producing group using only the selected environmental variables as inputs. 123

Table 5.18 A summary of the modelling results of the Non-CT producing group using the selected environmental variables, algal dominance and Chl *a* concentration as inputs. 126

Table 5.19 A summary of the modelling results for the CT producing group using algal dominance and selected environmental variables as inputs. 128

Table 5.20 Comparison of the experiments done with the MFF ANN modelling results of ForecasterXL showing the average actual errors, the good and bad forecasts and the contribution of the environmental, biomass and Chl *a* towards the output of each experiment. 130

Table 5.21 Summary of the average actual error (AE), the mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume including Chl *a* as an environmental variable in real time rule set development. 139

Table 5.22 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume real time rule set development excluding Chl *a* as an environmental variable. 143

Table 5.23 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the 7-days forward forecasting of *Microcystis* biovolume RULE set development. 147

Table 5.24 Summary of the average actual error (AE), the mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume, 14-days ahead RULE set development. 151

Table 5.25 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume 21-days ahead rule set development. 155

Table 5.26 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume 28-days ahead rule set development. 159

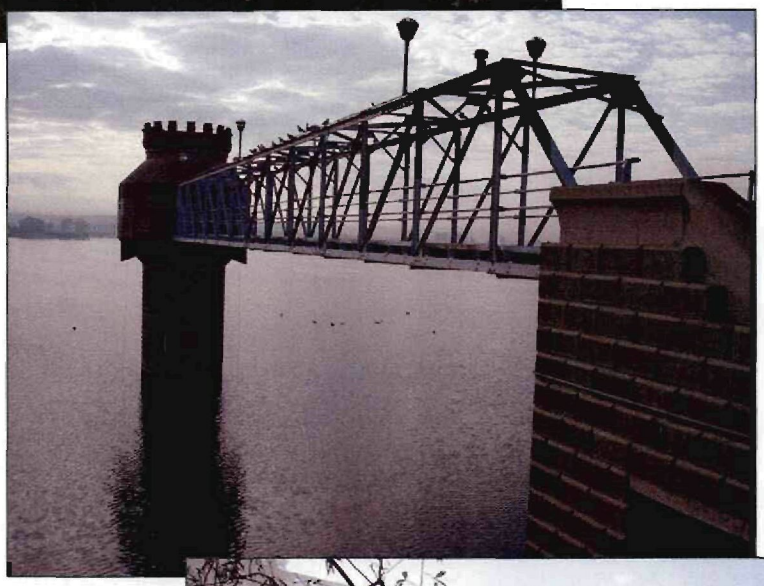
Table 5.27 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the real time training and testing data sets of the *Ceratium* biovolume rule set development. 164

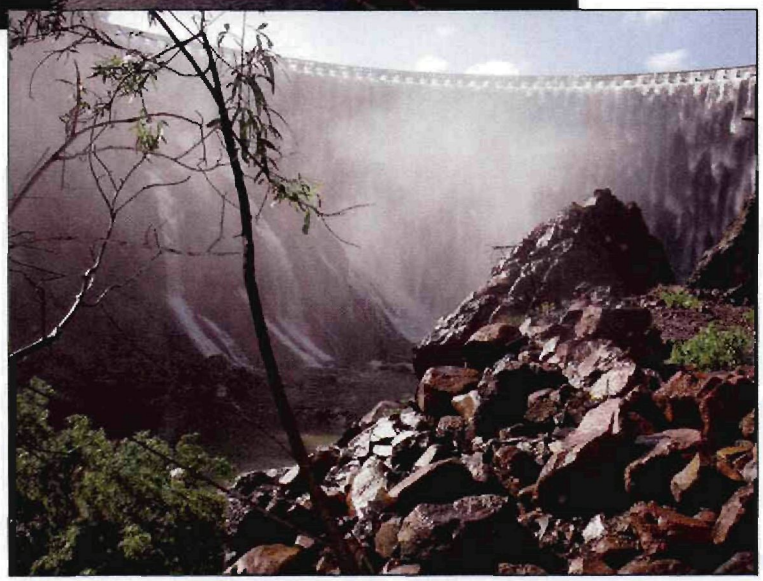
Table 5.28 The summarised results from the HEA Experiments, showing the correlation coefficient (R2), Root Mean Square Error (RMSE), the relative tolerances (10% for training and 30% for testing), the Environmental Variables important in Conditions to apply the THEN or ELSE Rule set, and the Environmental variables (in order of importance) that was used in the developed RULE sets. Experiment 1-6 was the development of RULE sets for *Microcystis*, and Experiment 7 was done for the development of a RULE set for the dinoflagellate, *Ceratium*. 169

LIST OF ABBREVIATIONS

Alk	Total alkalinity as $CaCO_3$
Ana_B	*Anabaena* biovolume
Asl	above sea leavel
ATMax	Maximum air temperature on day of sampling
ATMin	Minimum air temperature on day o sampling
BA	Bon Accord
Ca	calcium
ChlA	chlorophyll *a*
ChloroBi	Chlorophyte biovolume
ChrysoBi	Chrysophyte biovolume
Cl	chloride
CPV	Cumalitive percentage variance
CryptoBi	cryptophyte biovolume
CTMax	Maximum cyanobacterial toxin concentration
CTMean	Mean cyanobacterial toxin concentration
CTMin	Minimum cyanobacterial toxin concentration
CTTot	Total cyanotoxin producing group biovolume
CyanoBio	cyanophyte biovolume
Cyl_B	*Cylindrospermopsis* biovolume
DIN	Dissolved inorganic nitrogen ($NO_3 + NO_2 + NH_4$-N)
DIN:DIP	Dissolved inorganic nitrogen to dissolved inorganic phosphorous ratio
DIP	Dissolved inorganic phoshorous (measured as PO_4-P)
EC	Eletrical conductivity
EV	Eigenvalue
EuglenoB	Euglenophyte biovolume
z_{eu}:z_{mix}	euphotic depth: mixing depth ratio
fsl	full supply level
HBP	Hartbeespoort
HEA	Hybrid evolutionary algorithm
K	potassium
KV	Klipvoor
MATMax14	Mean maximum air temperture 14 days before sampling

MATMin14	Mean minimum air temperture 14 days before sampling
MIB	2-methylisoborneol
Mg	magnesium
Mic_B	*Microcystis* biovolume
MT	Mean daily temperature
MT5	Mean temperature 0-5m in water column
Na	sodium
NH_3	Ammonia
Osc_B	*Oscillatoria* biovolume
OxyDV	Oxygen depleted volume
PCA	Primary component analysis.
pH	Is the logarithm of the reciprocal of the conentration of free hydrogen ions. It measures the the activity of the hydrogen ion.
PhyrrBio	Phyrrophyta biovolume (dinoflagellates)
Pse_B	*Pseudoanabaena* biovolume
Rad	Radiation on day of sampling
Rad14	Mean Radiation 14days before sampling
RDA	Redundancy analysis
RDP	Roodeplaat
RV	Rietvlei
Secchi	Secchi disc reading
SO_4	Sulphate
Spec/Env Corr	Species and environmental variable correlation
TDS	Total dissolved salts
Tin14	Total inflow 14 days before sampling
TM	Total microcystin
TMO	Total monthly outflow
TN	Total nitrogen ($KN + NO_2 + NO_3$)
TN:TP	Total nitrogen to total phoshporous ratio
TotNONCT	Total non-cyanotoxin producing group biovolume
TotVol	Total volume on day of sampling
TP	Total phosphorous
TR14	Total rainfall 14 days before sampling
TV	Total variance
WST	Water surface temperature





CHAPTER 1

INTRODUCTION

Worldwide and in South Africa, there seem to be an increase in the frequency of cyanobacterial and dinoflagellate blooms recorded (Codd *et al.* 2005, Guven & Howard 2006, Harding & Paxton 2001, Van Ginkel *et al.* 2001). This may be the result of increased eutrophication and/or increased and more effective monitoring of phytoplankton in the water resources. A number of factors can contribute to the phenomenon, of which the impact of global warning (Burroughs 2003, Guven & Howard 2006, IPCC 1992, Mitchell 1991 as quoted by Kunz *et al.* 1995) and the then likely increases in water temperatures (Van Ginkel & Silberbauer 2007), especially the increases in minimum temperatures (Kunz *et al.* 1995, Mühlenbruch-Tegen 1992) is but one.

Cyanobacteria blooms, as a symptom of eutrophication, have a notorious reputation to develop rapidly in biovolume from obscurity to massive blooms in nutrient enriched systems. It can then just as abruptly recede. Some waters exhibit frequent and even sustained aggregations of cyanobacteria, while others experience evanescent, but often extremely noxious growth events. A significant proportion of cyanobacteria genera produce one or more of a range of cyanotoxins (WHO 1999) and many are associated with taste and odour problems for Water Treatment Works (Swanepoel *et al.* 2007). If water containing high cyanobacterial toxin concentrations is ingested (in drinking or in recreational waters), they present a risk to human and animal health (Pouria *et al.,* 1998, WHO 1999).

The hypertrophic reservoirs, Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat, in South Africa are warm, monomictic reservoirs, known to experience annual cyanobacterial blooms of the toxin producing *Microcystis aeruginosa* (Robarts 1984; NIWR 1985; Chutter 1989; Zohary & Robarts 1989; Chutter & Rossouw 1991; Van Ginkel *et al.*2000a; Van Ginkel *et* al. 2000b, Harding *et al.* 2004a). Hartbeespoort, Rietvlei and Roodeplaat reservoirs release water to downstream Water Treatment Works that withdrew water for potable use, and this presents a potential threat to the health of users. The extent of geosmin, MIB and cyanobacterial toxins associated with the cyanobacterial blooms highlighted the need to develop trustworthy short and long-term forecasting for hypertrophic systems so that Water Treatment Works can prepare for these algal bloom events.

These harmful algal/cyanobacterial blooms, often experienced in South African hyper-eutrophic fresh water systems, and its unpredictability, usually catch the water resource manager off-guard. Toxicity of these blooms is equally unexpected as blooms are not necessary toxic according to the literature. The water resource manager needs a modeling tool that can predict harmfull bloom formation based on physical and chemical characteristics of the impoundment, including climatic conditions.

The main aim of this project was to investigate the applicability of ecological infomatics modelling techniques to develop a predictive tool for harmful algal blooms in South Africa. This will enable water resources managers, managers of water treatment fascilities and managers of recreational water use to be prepared for harmful algal blooms, e.g 1) In planning management action timeously, 2) in purchasing activated carbon, and 3) To plan recreational activities and warn recreational users.

A number of hypothesis were set before the commencement of the study to be included in the investigation and the development of a predictive tool for harmful algal blooms:

1.1 In ecological studies a holistic approach is required to understand how ecosystems accommodate different levels of external (climate) and internal (trophic linkages and benthic-pelagic coupling) variability and their interactions. The integration of biological and meteorological information is needed to understand and estimate the consequences of climate variations on the development of toxic cyanobacterial blooms.

A detailed literature survey was done to determine the most important physical, chemical and biological factors in the development of harmful algal blooms. The literature survey also investigated the potential impact of climatic conditions on the development of harmful algal blooms, with special reference to cyanobacteria, that is the largest noxious phytoplankton problem within South African fresh water systems.

This hypothesis also led to the inclusion of a number of climatic variables in the statistical analysis of the collected data, to determine the potential use of climatic variables in predicting harmful algal blooms.

1.2 A water body is one entity and when a cyanobacterial bloom does occur, there are toxins to be found in the system.

This hypothesis aimed to investigate the distribution and potential for toxic cyanobacteria that may impact on recreational users, even though the regular sampling site does not indicate any or low cyanobacterial toxicity within the water column.

1.3 Various Deterministic and Ecological Modelling methods are available to be tested for the applicability to predict harmful algal blooms.

A number of different types of models were tested and variable successes were achieved. The Hybrid Evolutionary Algorithms proved to be quite useful to develop algorithms that can be used in the prediction of cyanobacterial and algal bloomforming species in the hypertrophic reservoirs of South Africa.

Five years (October 2000 – September 2005) of data was collected on the above mentioned five hypertrophic reservoirs to test the hypothesis and to use in the ecological modelling of blooms of harmful algal genera. Climatic variables were included to determine the potential impact of the weather patterns on the development of the problem algae/cyanobacteria. Another fourteen years of additional data was extracted from the Water Management System (WMS) of the Department of Water Affairs and Forestry and was used in the Hybrid Evolutionary Algorithm method to develop predictive capability.

CHAPTER 2

LITERATURE SURVEY

The aim of this literature survey is to give a broad overview of 1) eutrophication impacts and mangement, 2) to investigate the characteristics of harmful algal species, 3) to determine important physical and environmental variables in the development of harmful algal blooms, and 4) to compile a review of the literature on the application of models used to predict specific algal bloom conditions.

2.1 Eutrophication: Impacts and Management

Eutrophication:

"...the nutrient enrichment of waters which result in the stimulation of an array of symptomatic changes, amongst which increased production of algae and aquatic macrophytes, deterioration of water quality and other symptomatic changes are found to be undesirable and interfere with water uses"

OECD (1982)

Eutrophication is primarily associated with the natural ageing process of lakes (Wetzel 2003). During this process, organic matter (both living and decaying) is accumulated within the water system. Progressively the character of a deep water system is changed to that of a wetland and ultimately to that of a terrestrial system. Under normal and natural conditions this process take place over tens of thousands of years (Walmsley 2003). However, human impacts enhance this eutrophication process to occur over a much shorter period of time and not only in lakes. Therefore, two types of eutrophication have been identified, namely a) natural eutrophication and b) cultural eutrophication. Natural eutrophication occurs at a slow rate but is not reversible, while cultural eutrophication can be reversed, and occurs in a relatively short time. The latter refers to human-induced eutrophication due to anthropogenic activities. Eutrophication became associated with numerous aquatic resource problems (Dunst *et al.* 1974) of which toxic cyanobacterial blooms are one of the major problems.

2.1.1 *Sources and flow characteristics of nutrients*

The interactive linkage between plant production, nutrients and human activities has been noted since the early 20th century. Knowledge and understanding of especially nitrogen and phosphorus cycles in the aquatic medium is important to understand eutrophication.

A nutrient is a chemical compound or element that must be used directly by plant cells (both phytoplankton and macrophytes) for growth. Nutrients are usually inorganic elements that are assimilated by plants via metabolic processes and are utilised to produce and accumulate organic material in the aquatic ecosystem. Aquatic phytoplankton and macrophytes require 20 different nutrients. However, the rate and extent of aquatic plant growth is dependent on concentrations and ratios of nutrients in the water medium.

The overall composition of aquatic plant tissue is $C_{106}H_{263}O_{110}N_{16}P_1$ yielding a C:N:P weight ratio of 41:7:1, indicating the most important nutirents for plant development. The ambient optimal N:P ratio (w:w) for algal growth in surface water is in the range of 8:1 to 12:1 (Walmsley 2003). Because of nutrient supply and demand in nature, it has been observed that phosphorus (P) and nitrogen (N) are the most frequent limiting nutrients in freshwater systems (Walmsley 2003).

Biological growth and decomposition processes in the aquatic ecosystems are important for regulating available nutrient levels. There are varying rates for plant nutrient accumulations and bacterial mineralization of nutrients, which may lead to the incorrect understanding of nutrient concentrations such as C, N and P when real time observations are considered. That is why, ortho-phosphate (PO_4-P) may not be the right measure of phosphorus as an indicator of the trophic state, as it only indicates the excess phosphorus that has not been used by the standing crop. It is therefore better to measure the total phosphorus, as it gives a better understanding of the potential for phosphorus availability and the degree of nutrient enrichment. For example Lawrence *et al.* (2000) confirmed that the availability of nutrients largely determines algal biovolume as found in Burrinjuck Reservoir in Australia.

There is considerable debate about the most important nutrient, N or P. According to Harris (1986), Pick & Lean (1987), Elser *et al.* (1990), Jensen *et al.* (1994) and Scheffer *et al.* (1997) there is little evidence to support the contention that TN:TP ratios are an important determinant

of cyanobacterial dominance (Oliver & Ganf 2000). Yet, the impact of the N:P ratio has been considered as a potential management option for eutrophication in South Africa (Chutter & Rossouw 1991). The problem appears to be that the majority of experiments and lake manipulations are done without first assessing whether major nutrients are growth limiting. Reynolds (1992) have pointed out that the ratio is immaterial if the nutrient concentrations are in excess of those limiting to growth. It must be considered that the N:P ratio, when the inorganic forms of the nutrients are considered, might not be the driving force, but the result of the biological processes taking place in the system. Xie *et al.* (2003) concluded that their study proved that the low TN:TP ratio is not a cause but rather a result of *Microcystis* blooms. It, therefore, only shows the ratio for the excess of nutrients to establish growth in the system. However, when TN:TP is considered, it is a true reflection of the nutrients available in the system that can be utilised for future growth.

2.1.1.1 Phosphorus

Ecologically speaking phosphorus is in many ways the most important element since it is according to Liebig's law more likely to be deficient, and, therefore, to limit biological productivity (Hutchinson 1975, OECD 1982, Atlas & Bartha 1998, Wetzel 2001), than any of the other biologically important elements. According to Sharpley & Rekolainen (1997) eutrophication may occur at or below 0.01 mg/L dissolved P and according to Hesketh and Brookes (2000), Oliver and Ganf (2000) 0.02 mg/L total P. Thus, the ratio fraction of P needed for biological productivity is so small, that a very small quantity is needed to sustain biological productivity.

In some circumstances, moderate amounts of phosphorous, sub-optimal to one species, appear to be inhibitory to others (Hutchinson (1975). Some species (e.g. *Asterionella*) do have the ability to remove phosphorus efficiently from water containing very low concentrations of the element. Yet, numerous studies have shown high correlation between total phosphorus and chlorophyll *a* (Chl *a*) (OECD 1982, Perkins & Underwood 2000, Wetzel 2001).

Hutchinson (1975) noted that although there is little variation in phosphate in unproductive lakes; in very productive lakes with clino-grade oxygen curves, there is an increase in soluble phosphate in the oxygen-deficient part of the hypolimnion. This is partly due to the

decomposition of the sinking plankton, but in most cases is primarily caused by the liberation of phosphate from sediments on reduction.

According to Hutchinson (1975) and Wetzel (2001), an oxidised mud surface holds and prevents diffusion of phosphate and ferrous ions from the deeper mud layers. The ferrous ions are always in excess and when oxidised, precipitates all the phosphate. However, in lakes that are in phase III of stagnation, large amounts of ferric phosphate are formed during autumn turnover. Hutchinson (1975) suggests that this ferric phosphate then slowly hydrolyses and restores phosphate to the lake water and to the littoral zone. Where a large concentration of H_2S is formed in phase IV of stagnation, the phosphate will become more mobile at circulation and increase the potential for regeneration of phosphate.

Fisher & Reddy (2001) found that there has been a good deal of evidence of rapid movement of phosphate from the sediments to the water, at all levels, in small lakes and wetlands. In high summer, there may therefore be a great increase of total phosphorus in the surface waters during times of algal blooms, although the soluble fraction may be almost undetectable. Hutchinson (1975) suggests that one condition for the maximum development of such blooms may be the rapid decomposition and liberation of phosphate in the littoral zones during spells of very warm weather.

The recycling of sediment-bound phosphorus has been the subject of research for many years (Fleisher 1978, Wetzel 2001). A number of transport mechanisms are responsible for the sediment/water interface, namely a) diffusion, b) wind-induced water turbulence, c) advective movements of water and particles mediated by benthic animals (bioturbation), d) gas ebullition, e) attached algal or cyanobacterial uptake and f) rooted aquatic plant uptake (Wetzel 2001). There are, also a number of mechanisms by which the hypolimnion water can move into the epilimnion. These include: a) natural seasonal overturn; b) increase in wind action that leads to internal current formation causing a degree of water circulation; and c) small diurnal thermal changes in the water body that lead to leakage of the bottom water to the surface (Donnelly *et* al 1997). Therefore, wind action and the diffusional release of phosphorus into the freshwater system, supported by anaerobic internal loading can contribute significantly to algal biovolume (Donnelly *et al.* 1997, Perkins & Underwood 2001).

Historically it was believed that 80% of phosphorous influx into the freshwater system originated from point sources, namely waste water treatment works (Grobler & Silberbauer 1984). Since then many international research studies have also shown that extensive influxes (up to 91%) of P originate from diffuse agricultural sources (Jordan *et al.* 1997, Gillingham & Thorrold 2000, Heathwaite *et al.* 2000, Ruan & Gilkes 2000, Schoumans & Groenendijk 2000; Jordan *et al.* 2001, Withers *et al.*2001) and from internal recycling (Donnelly *et al.* 1997, Perkins & Underwood 2001). One has to note that these scenarios may be very site specific.

2.1.1.2 Nitrogen

Although the ultimate source of nitrogen, which plays a fundamental part in the metabolism of organisms, is certainly the molecular nitrogen of the atmosphere (Hutchinson 1975), most biologically available nitrogen is present as nitrate. The nitrate passes from rain to rivers and lakes, where much of it is taken up by algae, used for growth, and then deposited in the sediments (Horne & Goldman 1994). The threshold value, below which nitrogen limitation occurs, was assumed to be, 0.1 mg/L ((Sas 1989, Seip *et al.* 1992, Reynolds 1992 and Cooke *et al.* 1993) as referenced by Oliver and Ganf 2000).

The aquatic nitrogen cycle is shown in Figure 2-1. Nitrogen may enter the water phase in different forms: as dissolved N_2, nitric acid, NH_4^+, and NO_3^- as NH_4^+ adsorbed to inorganic particulate matter, and as organic compounds (either in dissolved or particulate phases). According to Wetzel (2001), there is no direct correlation between the volume of rainfall and the quantity of nitrogen influx per area of land. Dry fallout may contain high concentrations of nitrogen. The form of nitrogen present in lake waters can be grouped as follows:

- Molecular nitrogen N_2 that is in solution. This form of nitrogen causes the least growth in e.g. *Aphanizomenon flos aqua* (heterocystous) and *Microcystis aeruginosa* (non-heterocystous) according to Ward and Wetzel (1980). This form of nitrogen is readily assimilated by species capable of nitrogen fixation e.g. *Anabaena* and *Oscillatoria* (Oliver & Ganf 2000).
- Organic nitrogen compounds, including a great variety of decomposing products ranging from proteins to such simple substances as amino acids, urea, and the methylamines.

- Ammonia, mainly as NH_4^+ and NH_4OH has been shown to be responsible for the highest growth rates for both *Aphanizomenon flos aqua* (heterocystous) and *Microcystis aeruginosa* (non-heterocystous) at all light intensities (Ward and Wetzel, 1980). Tandeau de Marsac & Houmard (1993) also found that cyanobacteria have the highest preference for NH_4^+ (Oliver & Ganf 2000). Therefore, if NH_4^+ is available cyanobacteria and other micro-algae do not assimilate alternative nitrogen sources (Turpin 1991, Ochoa de Alda *et al.* 1996).
- Secondly, nitrate, as NO_3^- is reduced by nitrate reductase to
- Nitrite, as NO_2^-, that is reduced by nitrate reductase to ammonium (Guerrero & Lara 1987).

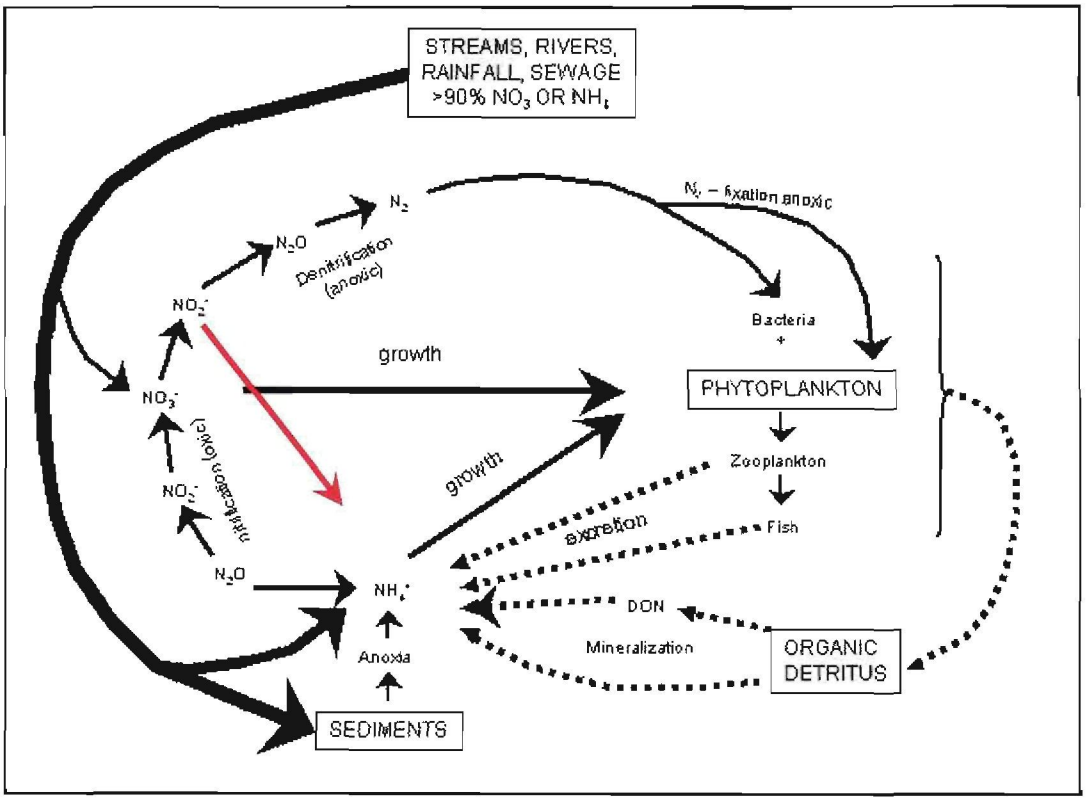


Figure 2.1 The nitrogen cycle in freshwater ecosystems. The thick lines indicate the main pathways in terms of mass transfer; the dotted lines, those involving recycling; and the thin lines indicate the mineralization in the water column (Redrawn from Horne & Goldman 1994)

Olivier & Ganf (2000) stated that cyanobacteria are disadvantaged by using nitrate-nitrogen under low light conditions, rather than non-nitrogen fixing cyanobacteria being favoured by ammonium and eukaryotic algae by nitrate nitrogen as suggested by Blomqvist *et al.* (1994).

All freshwater phytoplankton species can assimilate the latter four nitrogen forms (organic nitrogen compounds, ammonia, nitrate and nitrite). It must be noted that cyanobacteria and other bacteria are important in the anoxic sections of the nitrogen cycle, in contrast to the cycle that occur under oxygenated conditions. There is a direct correlation between N_2 fixation and heterocyst containing cyanobacteria e.g. *Anabaena* (Wetzel 2001). N_2 fixation is primarily light dependent and high concentrations of total phosphorus are necessary for N_2 fixation to occur. There is seasonal and temporal variability in N_2 fixation. According to Howarth *et al.* (1988a, b), N_2 fixation can account for as much as 80 percent of the annual nitrogen inputs in eutrophic lakes (Wetzel 2001) but this should be subject to the cyanobacterial species present.

However, in addition to atmospheric nitrogen, a major source of nitrogen into the freshwater system is from both surface run-off and from groundwater returns flows. The latter depends on the geological soil types, e.g. limestone soils can contribute large quantities of nitrogen.

According to Jordan *et al.* (1997) the transport of nitrate, which is very soluble and leach quite easily from soils, are enhanced by rapid infiltration and groundwater flows originating from fertilizer application on agricultural lands. Organic nitrogen and carbon are transported mainly by surface runoff and shallow subsurface flow.

Oliver and Ganf (2000) noted that nitrogen is of particular significance to the gas-vacuolated cyanobacteria, as it forms an essential component in the synthesis of the gas-vesicles. Thus, nitrogen limitation may impact on the cell growth, cell buoyancy, as well as the ability to regulate cell buoyancy. Therefore, nitrogen limitation is particularly detrimental to the non-nitrogen fixing bloom-forming cyanobacteria, e.g. *Microcystis*, and may be critical in the replacement with other phytoplankton species.

2.1.2 *Cyanobacterial bloom impacts*

The successful net growth of a species depends on its ability to optimise resource usage and to minimise losses (Whitton & Potts 2000). Cyanobacteria do possess the physiological and

ecological characteristics to dominate the phytoplankton of lakes, reservoirs and rivers. The cyanobacteria principally responsible for blooms are gas-vacuolated species in a number of genera and with varying size and shape (Whitton & Potts 2000). Cyanobacterial blooms cause problems that can have both short-term and long-term impacts. These problems can be classed into five different categories as follows:

- Impacts on Human Health

  Increased toxic cyanobacterial events cause increased human health impacts through inappropriate treatment of domestic water, as well as direct and indirect recreational use of freshwaters (WHO 1999).
- Impacts on Water Treatment Works

  Cyanobacterial blooms in raw water cause clogging of reticulation and canal systems by filamentous benthic algae, taste and odour problems and increased water treatment costs (Du Preez & Van Baalen 2006, Graham *et al.* 1998, WHO 1999).
- Impacts on the Ecological System

  The impacts on the ecological system include increased occurrence of de-oxygenation in the hypolimnion with associated chemical effects and releases of phosphorus from the sediments, increased fish and invertebrate mortality and changes to the ecological community structure and loss of biodiversity (Wetzel 2001).
- Impacts on human activities

  There is increased interference in recreational activities (boating, fishing, and swimming), decreased property values, and undesirable aesthetic conditions (Harding *et al.* 2004a, WHO 1999).
- Impacts on agriculture

  The main impacts on agriculture are the mortality of livestock due to cyanobacterial toxins (Kellerman *et al.* 2005). The effect of irrigation of crops consumed raw may be affected.

2.1.2.1 Impacts on human health

In 2000, 1.1 billion people, still did not have access to a safe supply of water, and 2.4 billion people of the world population had inadequate sanitation services (WHO 2000). There is a major drive in South Africa to improve access to drinking water and adequate sanitation

services. Despite these efforts, human health is still at substantial risk due to water quality problems in many areas of the world (Peters & Meybeck 2000).

Eutrophication has caused increased incidents of algal and cyanobacterial blooms. Associated with these blooms are problems with taste and odours (Wnorowski 1992, World Health Organization 1999, Whitton & Potts 2000, Downing and Van Ginkel 2003), and cyanobacterial toxin production that cause user problems to both the potable water industry and the recreational use of surface waters (World Health Organization 1999, Whitton & Potts 2000). Taste and odours as well as cyanobacterial toxins in drinking water supplies have received increasing scientific attention over the past 40 years. The complaints from end-users demanding safe, odourless drinking water, has put pressure on the water treatment authorities.

a) Taste and odours

> "There is a certain, though not universal, correlation between odour production from *Microcystis* blooms and the presence of microcystins"
>
> (Wnorowski 1992)

According to Wnorowski (1995) the most problematic condition regarding taste and odours are algal production of earthy/musty odours from eutrophied water sources. In taste and odours, with specific mentioning of geosmin and MIB, activated carbon was one of the most successful treatment methods (Sigworth 1957, Wnorowski 1995). A study done by Wnorowski (1992) on South African taste and odour producers highlighted the fact that algae, cyanobacteria and actinomycetes are the major biota causing taste and odour problems. Problems with taste and odour can also arise from other chemical compounds (Du Preez & Van Baalen 2006). However, most of the complaints in South Africa concerning bad taste and odours stem from eutrophic systems (Wnorowski 1992).

Wnorowski (1995) found positive correlations between nitrogen concentration and geosmin production and indicated that increases in geosmin are rather the result of increases in algal biovolume than enhanced geosmin production. Algal species, that have been associated with the production of geosmin and MIB, are *Anabaena, Aphanizomenon, Fischerella, Hyella, Lyngbya, Oscillatoria, Phormidium, Pseudoanabaena, Schizothrix* and *Symploca* (Miwa & Morizane 1988, Naes & Post 1988, Negoro *et al.* 1988, Wu & Juttner 1988, Wu & Hsu 1988,

Naes *et al.* 1989, Aoyama *et al.* 1995, Izaguirre & Taylor 1995, Izaguirre & Taylor 1998, WHO 1999). Other microbiological organisms, such as *Streptomyces* sp. J-11 (Yagi *et al.* 2002) also produce geosmin and MIB.

Geosmin does not pose a health hazard (Du Preez & Van Baalen 2006). Wnorowski (1992) found geosmin concentrations ranging from 7.5 ng/L to 3170 ng/L in seventeen (31.5%) of the fifty-four tested dams that are situated in the old Transvaal and Free State. All the odour-producing dams were dominated by cyanobacteria, indicating that cyanobacteria are potentially the main culprit of taste and odours in the final drinking water. There is thus the potential and possible association with cyanotoxins. However, it must be noted that many other algal species and other compounds are also associated with potential taste and odour problems.

b) Cyanobacterial toxins

The cyanobacterial species most commonly associated with cyanobacterial toxin production are *Anabaena, Oscillatoria, Phormidium, Microcystis, Lyngbya* and *Cylindrospermopsis* (WHO 1999). These species produce different types of toxins and there are already more than 60 known cyanobacterial toxins identified internationally. The three main types of toxins are hepatotoxins (e.g. microcystins, nodularins and cylindrospermopsins), neurotoxins (e.g. anatoxins and saxitoxins) and lipopolysaccharides. Of these toxins only a few standards are available for analyses (Carelsen 2003 *Pers. Comm.*). This makes analysis for the toxins problematic. Only microcystin-LR was readily available for HPLC analysis. The cost of the ELISA method prohibits the wide use of this method that test for five different hepatotoxins. This situation is currently being improved as more standards are commercially available.

In contrast to geosmin and MIB, cyanobacterial toxins may pose a serious health hazard as known cases of allergic reactions, poisoning and death have been reported (WHO 1999). Frank & Wolf (2002) did a survey in 155 lakes in Germany to determine the incidence of cyanobacteria and toxins in order to assess the recreational health hazard. They concluded that there is a risk of toxic cyanobacterial blooms during the summer bathing (swimming) season. The only known human mortalities due to cyanobacterial toxins originated from badly treated drinking water (WHO 1999). An internal DWAF survey on the Hartbeespoort and Brits drinking water plants highlighted the potential risk of users (Van Ginkel 2004) when drinking

water is provided from a hypertrophic system. Eynard *et al.* (2000) also showed the potential risk to users of groundwater contamination from an eutrophied lake.

The WHO (1999) has determined a provisional Total Daily Intake (TDI) guideline for drinking water of 1 mg/L microcystin-LR. Wnorowski (1992) determined that in four of the dams where taste and odours were identified toxins were also found. All these incidents were associated with *Microcystis*.

2.1.2.2 Impacts on Water Treatment Works

Downing & Van Ginkel (2003), Codd *et al.* (2005) and Du Preez & Van Baalen (2006) confirmed that the major perceived problem for Water Treatment Works associated with cyanobacterial blooms are taste and odour problems. Very little work has been done in South Africa on the presence of cyanobacterial toxins in the final water of the treatment works. An emergency at the Brits and Hartbeespoort water treatment works during November 2003 (Van Ginkel 2004) highlighted the potential hazard to end users of poorly treated water. In fact, this is possibly the tip of the iceberg regarding the problems experienced by Water Treatment Works, as only four Water Treatment Works in South Africa are presently testing for cyanobacterial toxins in the raw or final water.

2.1.2.3 Impacts on the Environment

Cyanobacterial toxins are known to accumulate in common aquatic vertebrates and invertebrates, including fish, mussels and zooplankton. There is thus a considerable potential for increased toxic effects in the aquatic food chains (WHO 1999). Cyanobacterial toxins have been confirmed to be responsible for fish kills (Zimba *et al.* 2002). However, fish and other aquatic animal deaths may be associated with the anaerobic circumstances that develop during the collapse of a cyanobacterial or algal bloom. According to the WHO (1999), it appears that cyanobacterial toxins may exhibit a deleterious effect on zooplankton, but no definite conclusions on the effect of cyanotoxins on bacteria could be drawn.

2.1.2.4 Impacts on Recreation

Numerous descriptions of adverse health effects for swimmers exposed to cyanobacterial blooms have been described (Dillenberg & Dehnel 1960, Turner *et al.* 1990, WHO 1999) and minor contact with cyanobacteria in swimming waters can lead to skin irritations and increased likelihood for the development of gastrointestinal symptoms (Pilotto *et al.* 1997).

Known symptoms of cyanobacterial toxins vary between headaches, vomiting, skin rashes, mouth ulcers, fevers, eye and ear irritations, muscular pains, painful diarrhoea and severe pneumonia (WHO 1999). The symptoms increased significantly with the duration of water contact and cyanobacterial cell density. It is thus clear that recreational users are at risk to contract health impacts, even with indirect recreational contact, like rowing.

2.1.2.5 Impacts on agriculture

Numerous incidents of livestock and pet deaths have been associated with cyanobacterial bloom conditions in water supply dams (WHO 1999; Harding & Paxton 2001, Kellerman *et al.* 2005). Livestock deaths have a serious financial implication and until now has been associated extensively with cyanobacterial blooms (Kellerman *et al.* 2005). In South Africa almost a complete dairy herd of 500 cows were killed during the Kareedouw event (Harding 1997, Kellerman *et al.* 2006)

2.2 Characteristics of Cyanobacteria

This section of the literature survey was done to understand the morphology, physiology and aquatic bloomforming conditions that may have an impact on the testing and development of models to predict future harmful algal blooms.

2.2.1 Morphology and Physiology of Cyanobacteria

Cyanobacteria are ubiquitous prokaryotic organisms that may be found on moist barren soil, footpaths, tidal flats, cliffs and sidewalks kept wet by moist environments (Whitton & Potts 2000). They occur in hot springs, in intimate association with other plants or animals (Weier *et al.* 1974) and in ice (Whitton & Potts 2000).

Cyanobacteria are unicellular or filamentous organisms. Cyanobacteria are true bacteria with a simple prokaryotic cell structure. These prokaryotic cells lack a nuclear membrane, mitochondria and chloroplasts. Cyanobacteria are, however, in many respects physiologically similar to the algae. The most important characteristic is the autotrophic photosynthesis that is the primary mode of nutrition, and the major process of synthesizing new organic matter. The cells contain cellular inclusions, which in some cases assume similar functions to eukaryotic organelles, e.g. pigment-bearing lamellae (photosynthetic lamellae), a plasma membrane, and a nuclear region containing chromosomal material (Lee 1999, Weier *et al.* 1974, Wetzel 2001). The cyanobacteria differ from bacteria in that they lack mesosomes and they have the ability to liberate oxygen for photosynthesis (Weier *et al.* 1974).

Cyanobacteria have murein, a protein forming up to 50 percent of the cell wall. Reproduction is by binary fission and they do not divide by mitosis, like most other bacteria. Cyanobacteria are distinguished from other bacteria by the enzymatic and pigment (e.g. Chl *a*) capability to photosynthesise, like other higher plants, under aerobic conditions. Water is used as an electron donor in photosynthesis, which is also more advanced than bacterial photosynthesis. Glycogen, not starch or oil, is the product of photosynthesis (Coombe & Bell 1965).

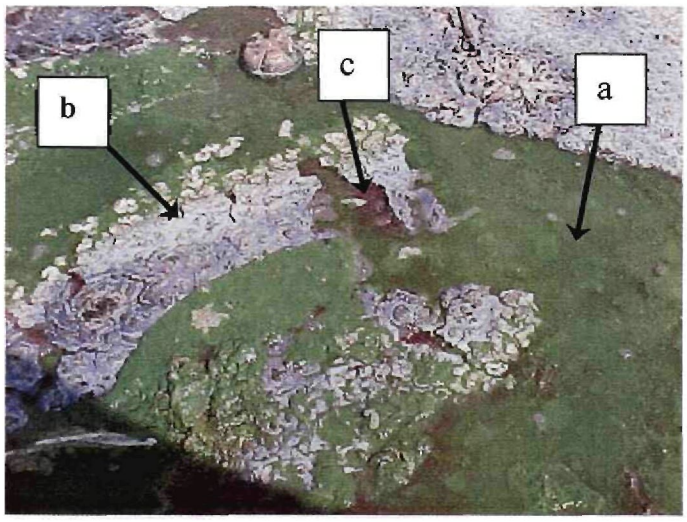


Figure 2.2 Cyanobacterial hyper-scum from the Hartbeespoort Dam in South Africa showing the distinct colours that depend on the pigments present in cyanobacteria. a) green colour due to Chl *a*; b) blue colour due to phycocyanin; and c) red colour due to phycoerythrin.

The pigments present in the cyanobacteria are Chl *a*, β-carotene, lutein, myxoxanthin, myxoxanthophyll, oscilloxanthin, echinenone, phycocyanin and phycoerythrin (Weier *et al.* 1974, Wetzel 2001). Chl *a* forms the primary photosynthetic pigment of all oxygen-evolving photosynthetic organisms and is present in all algae and cyanobacteria. The Chl *a* is situated in the thylakoid membrane in cyanobacteria (Lee 1999). The carotenes are linear unsaturated hydrocarbons, and the xanthophylls are oxygenated derivatives of the carotenes (Goodwin 1974). The phycobiliproteins, phycocyanin and phycoerythrin are water-soluble, pigment-protein complexes that occur in cyanobacteria, and are attached to the thylakoid membranes (Lee 1999).

Cyanobacteria are variable in colour when viewed macroscopically. Colour varies from blue-green, grey-green, violet, brown, purplish to red. These colours are dependent on the relative proportions of their photosynthetic pigments, namely chlorophyll (green), phycocyanin (blue) and phycoerythrin (red). The sheath pigments, e.g. scytonemin produce the brown colouring in some cyanobacteria. According to Coombe & Bell (1965), the colour pigment develops complimentary to the light the cell receives. Lee (1999) noted that cyanobacteria that produce the red phycoerythrin and the blue phycocyanin in white light, suppress phycoerythrin synthesis in red light and phycocyanin synthesis in green light (from Tandeau de Marsac 1977).

Visual blooms of cyanobacteria are often blue-green, yet sometimes shades of purple are found when all three pigments are present (Fig. 2-2). This is because many species have a sheath around individual cells or the whole filament and this sheath is often golden or dark brown, though sometimes a shade of red (John, Whitton & Brook 2003).

Single-celled cyanobacteria range in size from 0.6 μm to well over 30 μm, while filamentous forms range in cell diameter from 0.4 μm to over 100 μm. In the filamentous forms with a sheath, e.g. *Lyngbya*, the cellular structure is known as a trichome, while the filament refers to both the trichome and the sheath (John, Whitton & Brook 2003, Lee 1999).

Small-celled (less than 2 μm) cyanobacteria often occur as single cells, and are known as picoplankton. Slightly larger unicellular forms are mostly arranged in some sort of a colony, e.g. *Chroococcus* and *Microcystis*. The filamentous forms occur more often as single filaments, yet the majority tend to have characteristic colony growth forms.

Many cyanobacteria have the ability to photosynthesize under aerobic and anaerobic conditions. Under aerobic conditions, electrons for photosystem I are derived from photosystem II. Under anaerobic conditions, in the presence of sulphur, electrons are derived by the oxidation of sulphur (Daintith 1991):

$$CO_2 + 2H_2S \xrightarrow[\text{Chlorophyll}]{\text{light}} (CH_2O) + H_2O + S$$

sugar

Lee (1999)

Oscillatoria limnetica are known to use both photosystem I and photosystem II, depending on the oxygen content and the availability of sulphur under anaerobic conditions. Lee (1999) also noted that in many cyanobacteria photosynthesis is stimulated by lowered oxygen concentrations.

Luxury P uptake by *Microcystis* is well known, whereby the cells store nutrients in excess of the immediate growth requirements (Ishikawa *et al.* 2002). Nitrogen can be stored as cyanophycin and phosphorus as polyphosphate bodies (Kromkamp 1987, Simon 1987). This enables the species to continue growth even under nutrient limitation (Ishikawa *et al.* 2002).

Many cyanobacteria contain akinetes (resting spores) and heterocysts. The akinetes are resistant to unfavourable environmental conditions and occur in the orders Nostocales and Stigomentales. New filaments are produced by the germination of akinetes. Mature akinetes are considerably larger than the vegetative cells, contain protoplasm with food reserves and have a normal cell wall (Lee 1999). An increase in cytoplasmic density of the akinetes that causes a loss of buoyancy causes the filaments to sink and over-winter in bottom sediments. When there is germination of the akinetes, the reverse is happening. According to Lee (1999), ranges of physio-chemical factors (e.g. phosphate deficiency, low temperatures, carbon limitation and the reduction in light energy) stimulate akinete formation.

Heterocysts are also larger than the vegetative cells but appear empty under the light microscope when compared to the akinetes that are full of storage products. Heterocysts are surrounded by a thick laminated cell wall that limits interchange of atmospheric gases,

including oxygen. The internal environment of the heterocyst is thus virtually anoxic, the ideal situation for nitrogenase, a known oxygen sensitive enzyme (Lee 1999). Heterocyst formation is influenced by the molybdenum and nitrogen concentration of the medium. Heterocysts are the centre of nitrogen fixation and have cytoplasmic connections to the vegetative cells for conveying fixed nitrogen to the neighbouring cells and for conveying fixed carbon metabolites from the vegetative cells to the heterocysts (Giddings & Staehelin 1981 in Lee 1999).

Orr and Jones (1998) found that there is a linear relationship between microcystin production and cell division rates. Ishikawa *et al.* (2002) found that *Microcystis* cells could divide once or twice during transport from inshore bays to offshore areas in Lake Biwa. Toxins may be produced as the cells divide during transport. Therefore, toxins can be reasonably mobile within a lake or Reservoir. Frank & Wolf (2002) found that toxin concentration is not proportional to total phosphorus or Chl *a*. A suitable monitoring programme should thus include toxin analysis for safety reasons.

2.2.2 Bloom Forming Conditions for Cyanobacteria

Cyanobacterial blooms and the prediction thereof are important for precautionary management to be established in time. One of the important recommendations by the National Rivers Authority (1990) was that 'the use of predictive models, to quantify the development of algal biovolume in relation to changes in environmental variables, should be further evaluated'. A holistic approach is required in ecological studies to understand how ecosystems accommodate different levels of external (climate) and internal (trophic linkages and benthic-pelagic coupling) variability and their interactions. According to Whitton & Potts (2000) the cell densities, species composition, vertical distribution, longevity and timing of maxima differ to a large degree because of changing influences of the stratification and mixing that is determined by chemical, climatic and meteorological conditions. The integration of biological and meteorological research is needed to understand and estimate the consequences of climate variations on the ecosystem.

According to Blenckner, Pettersson & Padisák (2003) warmer winter temperatures, with associated low phosphorus concentrations, as well as earlier spring blooms without an early onset of stratification may lead to higher summer phytoplankton biovolume. Robarts (1984) has already found that photosynthetic capacity is temperature dependent. Lawrence *et al.* (2000)

note that different forms of the same nutrient can also have substantial impacts on the composition of the phytoplankton population. All the cyanobacterial species that cause blooms is primarily gas-vacuolated species (Whitton & Potts 2000). Shapiro (1990) notes that a low availability of CO_2 gives gas-vacuolated cyanobacteria an advantage. This advantage is re-enforced during periods of large cyanobacterial blooms and assist these organisms to stay dominant (Oliver and Ganf 2000).

Under stable and suitable physical environmental conditions, available nitrogen forms influence the growth of large buoyant cyanobacteria. The nitrogen fixing genera have an advantage when nitrogen limitation occurs. When nitrogen is limiting but in the nitrate form, green algae seem to be dominant. However, according to Oliver and Ganf (2000) the main limiting nutrients (N, P and C) will have a major influence through its impacts on buoyancy regulation.

The number of factors contributing to the dominance of cyanobacteria according to Lawrence *et al.* (2002) includes:

- Low wind and hence **low mixing** energy that result in the dominance of Cyanobacteria as a result of its buoyancy abilities;
- **High levels of phosphate** is a driving force in cyanobacterial blooms;
- **Low z_{eu}:z_{mix} ratios** throughout the summer period;
- Potential direct involvement of cyanobacteria (specifically *Microcystis*) in the **redox process**;
- **High NH_3 fraction** to the concentration of inorganic N.
- **High temperature.**

According to Reynolds (1984), the two major environmental variables stimulating phytoplankton species progression towards a cyanobacterial dominated bloom, are changes in the stability of stratification and declining nutrient availability.

In South Africa, turbidity may play an important role in the development of cyanobacterial blooms, by limiting light penetration into the euphotic zone and therefore regulating phytoplankton production (Rossouw 2000). Turbidity is not a major factor that contributes to the development of cyanobacteria as the dominant species, but it is in limiting the biovolume volume through light limitation. However, turbidity may play a role in the specific dominant

cyanobacterial species, as certain species can out-compete other species under lower light conditions (Wetzel 2001).

2.2.2.1 Low Mixing

Algae, which are dependent on mixing currents for circulation through the euphotic zone, are disadvantaged under low mixing conditions that favour cyanobacteria.

A study on the Roodeplaat Dam determined that the effect of wind behaviour of the Reservoir is more significant than the effect of radiation or humidity (Venter & Herold 2001). Physical mixing of a reservoir is important in maintaining and/or increasing turbidity. This can be induced by high inflow and wind action (Rossouw 2000). These two factors need to be considered when determining the stability or low mixing within a system.

Donnelly *et al.* (1997) highlighted the effect that long residence times had on the major cyanobacterial bloom of the Darling-Barwon River in 1991. However, in South Africa most of the Reservoirs have low retention times (Rossouw and Görgens 2005) and it is not comparable to the Australian River system.

2.2.2.2 High Levels of Phosphorus

Eutrophication research has focussed for many years on the importance of phosphorus as the limiting nutrient in phytoplankton production (Vollenweider 1990, Rossouw 2000 Rossouw and Görgens. 2005). The concept of high levels of phosphorus has for many years been subject to the N:P ratio. Wetzel (2001) highlighted the regulating effect of phosphate uptake in cyanobacteria by cation concentrations, particularly Ca^{+2} and Mg^{+2}. Whitton and Potts (2000) highlighted that nutrient limitation stimulates the storage of non-limiting nutrients in e.g. cyanobacteria, so that growth can be maintained during periods of nutrient scarcity. This applies to both phosphorous and nitrogen.

Hutchinson (1975) noted that the onset of cyanobacterial blooms is primarily determined by the onset of rapid decomposition at high temperatures in the littoral zone with a subsequent rapid output of phosphorus into the epilimnion. Chutter & Rossouw (1991) mentioned that the adsorption rate of phosphorus by the sediments is governed by the ambient dissolved

phosphorus concentration. Donnelly *et al.* (1997) and Lawrence *et al.* (2000) highlighted that in Australian waters the 'internal loading related pathway' (release of nutrients from the sediments as a result of elevated organic loading) are the most likely nutrient pathway. This indicate that nutrients, through internal loading, may be a major contributor to the development of cyanobacterial blooms.

Additionally Chriswell *et al.* (1995) determined that sulphate plays a major role in P release from freshwater sediments and Barbanti *et al.* (1995) confirmed a relationship between relatively large pore water P concentrations in anoxic bottom sediments and the activity of sulphate-reducing bacteria (in Donnelly *et al.* 1997). It was found that the influx of sulphate-rich saline water into the Darling-Barwon River during the 1991 cyanobacterial bloom caused higher clarity and increased the intensity of sediment/water interactions in the anoxic bottom sediments due to sulphate-reducing bacteria.

2.2.2.3 Low z_{eu}:z_{mix} Ratios

The euphotic depth (z_{eu}) is the depth where 1 percent of the surface radiation penetrated into the surface water. It is used to measure the depth below which photosynthesis is limited due to light limitation. Secchi depth is commonly used for this determination in the relationship:

$$z_{eu} = 2.7 \times Z_{secchi}$$

Lawrence *et al.* (2000)

Mixing depth can be determined from the temperature profiles. The depth of the surface mixed area was calculated by Lawrence *et al.* (2002) as the maximum depth where the temperature gradient was less than 0.2°C.

2.2.2.4 Redox Processes

In pure inorganic chemical systems, oxidation-reduction (redox) reactions are characterised by the flow of electrons between oxidised and reduced states until equilibrium is reached (Wetzel 2001). The elements C, N, O, S, Fe and Mn are the major elements participating in the redox process in surface water systems. Photosynthesis and externally derived organic matter are the major causes for the dis-equilibrium in redox conditions in natural waters, producing local centres of highly negative redox (Lawrence *et al.* 2000). The generation of NH_3 are biologically

mediated, however, the amount of NH_3 is maintained in the water column is a function of the redox equilibrium in the system. Thus, the nitrogen form present in the system is a function of the redox and pH of the water body.

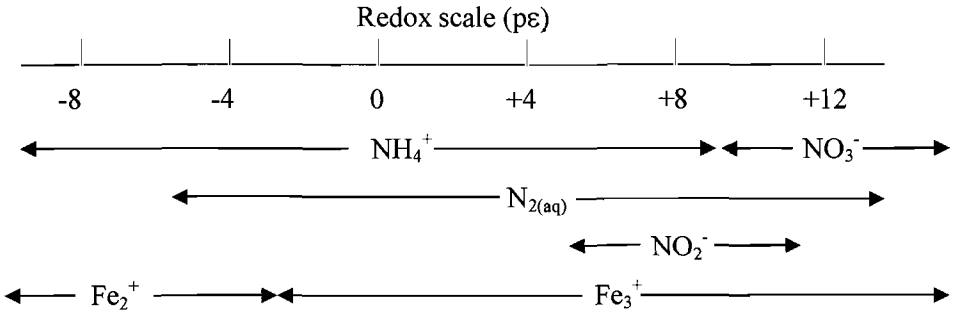


Figure 2.3 The redox scale and equilibrium phases for selected nutrients and metals (redrawn from Lawrence *et al.* 2000).

The redox potential also controls the form of iron present in the hypolimnion (Fig. 2.3), therefore, controlling the solubility of phosphorus in the bottom waters of the water body.

2.2.2.5 High Ammonium composition

Lawrence *et al.* (2000) found that potential nutrient limitation and depressed mixing regimes in late summer causes a shift in dominant algae to: non-fixing cyanobacteria under high ammonia or reducing conditions; or to motile forms of green algae under low ammonia and high nitrate conditions; or to N-fixing cyanobacteria in situations where nitrogen becomes limiting over the summer period. *Microcystis* tend to become dominant when nitrogen is limiting in the surface waters but is present as ammonia in the deeper anoxic layers of the water column. A significant negative correlation between ammonium and Chl *a*, was shown by the study of Perkins & Underwood (2000).

Whitton and Potts (2000) showed that according to Blomqvist (1994) significant cyanobacterial populations did not commence until nitrate was almost depleted. This led to the postulation that non-nitrogen fixing cyanobacteria, like *Microcystis*, are favoured by ammonium, especially under low light conditions (Ward and Wetzel 1980).

2.2.2.6 High Temperature

High temperature was significantly related to the log scale photosynthetic saturation parameter (I_K) according to Talling (1957), and Robarts (1984) in the Hartbeespoort Dam. Many studies have highlighted the importance of temperature to cyanobacterial blooms due to positive correlation between algal biovolume variation and water temperature (Recknagel 1997, Ha *et al.* 2002, Perkins & Underwood 2000, Weyhenmeyer *et al.* 2002)

Andradóttir & Nepf (2000) highlighted the effect of temperature of the inflowing river water. The temperature of the inflowing water affects the intrusion level into the lake water (Hart & Hart 2006) because of density differences. If the water is warmer, it will enter the epilimnion feeding available nutrients directly into the area where it is directly available for plant/algal growth. In cases where the water temperature of the inflowing water is lower than the epilimnion temperature, the intrusion of inflowing water occurs deeper in the system affecting the direct availability of nutrients for plant/algal growth. Thus, the temperature of the inflowing water can affect the lake water quality.

During storm events, the intrusion depth of the inflowing water is governed by the difference between current and seasonal meteorological conditions (Andradóttir & Nepf 2000). South Africa with its summer thunderstorms may cause large influxes of nutrients into the Reservoirs. With the present study we need to determine the impact thereof on the development of cyanobacterial blooms, taking into account all the factors mentioned above. The study by Van Ginkel & Conradie (2001) highlighted the potential effect of climatic conditions and the impact it had on a cyanobacterial bloom in the Orange River in South Africa. A study by Ha *et al.* (2002) confirmed the trigger effect of precipitation patterns on the establishment and extent of *Microcystis* blooms. Already in the 19th century, naturalists recognised that different plants and animals are restricted to fairly narrow climate regimes. Charles Darwin has after his travels published his theory of natural selection that highlighted the adaptability of species to climate variability and change for sustainable development and the survival of species (from Burroughs 2003).

When oxygen is most abundant during spring periods, the most important factor controlling bacterial activity is temperature (Sander & Kalff 1993) because when temperatures exceed the 10°C threshold up to 10-fold increases in bacterial activity can be expected (Goedkoop &

Törnblom 1996). The high specific heat of water gives rise to the fact that large volumes of water change temperature relatively slowly (Wetzel & Likens 1979). Thus, large volume water bodies tend to moderate local climates, which provide longer growing seasons for aquatic life, and serve as integrated recorders of recent climatic phenomena. It must be kept in mind that temperature is related to the heat content of a water body, but only measures the intensity of the heat and not the amount of heat stored.

2.3 Impacts of Climate and Weather on Freshwater Systems

This section of the literature survey was done to determine important, external, climatic variables for the development of harmful algal blooms and what historical reservoir modelling were done in South Africa and internationally.

2.3.1 *Climate, Weather and Changes*

The word climate originates from the word *'klima'*, which means inclination and refers to the angle of the sun's rays above the horizon. The early Greeks knew that the climate was cooler if the inclination of the sun was low. That is why the climate is hotter on the equator and colder at high latitudes (Burroughs, 2003). Burroughs (2003) highlighted the difference in meaning between climate and weather. Climate is the general expected weather situation for a specific month, season or year. This information is founded on statistics built up from observations over many years. Weather is what is happening to the atmosphere at any given time. However, the climate of a specific location on earth is much more complicated than only the inclination of the sun's rays. Other factors, such as altitude, distance from the sea, overall circulation of the atmosphere and geological formation of the substrata is also important.

Another unclear boundary is the difference between climatic variability and climatic change. This follows from the definitions for climate and weather. Climate change is about shifts in meteorological conditions that lasts for years. These changes involve a single parameter, such as temperature or rainfall, but they are usually associated with more overall variation in the weather patterns. According to Burroughs (2003), climate changes are linked to an overall warming or cooling of the globe. Thus, the difference between climate variability and climatic change has a certain degree of arbitrariness and depends on the timescale considered.

The increased ability to monitor climatic and weather changes by satellites for example, led to climate data and information to form increasingly part of the decision making process. It also improved our methodologies in managing water resources, because of the recognition of cyclical and cascading effects of human activities on the water quality and quantity along the hydrological pathways (Peters and Meybeck 2000, Schulze 2000). Evidence of global warming has extended the limnological interest in the climate, as climate is intrinsic to everything important to society (Gleick *et al.* 2001). The quality of water resources can be as important as or even more important than water quantity. Global climate changes will have a wide range of effects on the quality of freshwater systems by changing temperatures (Van Ginkel & Silberbauer 2007), flows, runoff rates and timing, and the ability of freshwater systems to assimilate waste and pollutants. Human activities within catchments also have a major effect on the water quality of a freshwater system.

There are two main reasons why we need to determine the effects from the historical natural variability of the climate. Firstly to put recent changes due to human impacts into context and secondly to get a better feel if the climate as we know it can undergo sudden and dramatic shifts. These will enable us to better understand the effects that may be expected under future conditions.

2.3.2 *Climate and Weather Impacts on Phytoplankton*

The phytoplankton community is an ideal indicator of changes in climatic conditions because of the dominance by short-lived organisms that respond rapidly to changes in the weather (Catalan & Fee (1994) as quoted by Weyhenmeyer *et al.* (2002)). Blenckner *et al.* (2002) also established a clear connection between the large-scale climate and the local weather conditions in winter and the effect it has on the phytoplankton biovolume within a system. The cyanobacterial biovolume in 17 European lakes was found to increase considerably after a mild winter (Weyhenmeyer *et al.* 2002). These studies highlighted the potential effect of increased winter temperatures on the freshwater system due to global climatic changes as predicted by IPCC (1992) and Burroughs (2003). Different cyanobacterial species have different optimum growth temperatures (Tang and Vincent 1999, Falconer 2004), and climate variability may lead to different algal/cyanobacterial species becoming dominant in the enriched systems. The potential toxic effect that cyanobacterial blooms have, indicated that the changes within freshwater system can harbour potential serious health impacts in future, due to climate change.

It may be important to investigate extreme climatic events when considering global change scenarios in forecasting possible long-term and short-term effects within the freshwater system. Wetzel (1989) found that in a small second-order stream in Michigan between 60 and 80 percent of the annual PO_4-P, that is available for phytoplankton growth, from this stream to the receiving lake occurred during three major precipitation events. It is therefore apparent that the climatic and weather conditions play a major role in freshwater systems by providing the essential nutrients in unpredicted ways.

2.3.3 *Modelling Climate and Biological Processes*

Models are both idealizations and simplifications of real systems. The idealization component is where complex mathematical expressions are replaced by equations, which describes ideal conditions. Simplifications are achieved by limiting the number of components in a model to those of interest (Steel 1997, Rossouw 2000). This philosophy of Occam's razor when applied to models means that one should eliminate all unnecessary information relating to the analysed problem (Starfield *et al.* 1990). There is also a certain amount of scepticism as to achievability of predicting and explaining biological processes within a reservoir or lake (Steel 1997) and the fact that a modeller should not promise more than what is feasible.

Numerous simulation models have been tested to predict different components of eutrophication in South Africa (Grobler 1985, Rossouw 1990, 2000, Venter & Herold 2001, Gal *et al.* 2003) and elsewhere (Dillon & Rigler 1974; Rast & Lee 1978; Barth & Gakstatter 1978; Robson & Hamilton 2004, Guven & Howard 2006). None of the models used in South Africa to date have attempted to predict algal/cyanobacterial species *per se*. Deterministic models (Jorgensen 1976; Bendorf & Recknagel 1982, Guven & Howard 2006), time series analysis models (Whitehead & Hornberger 1984), and fuzzy-logic models (Recknagel *et al.* 1994) are the major classes of models for the same purpose (Karul & Soyupak 2001). The use of hybrid ecological algorithms, with genetic programming, to optimise the deterministic models are also used (Muttil and Lee 2005; Cao *et al.* 2006; Chan *et al.* 2006).

Venter and Herold (2001) tested the MINLAKE model on Roodeplaat Dam to determine the applicability in predicting future trophic status of a reservoir. Results showed that the hydrodynamic and phosphate concentration behaviour were simulated well, however, the

prediction of the degree of change in Chl *a* concentrations, ammonia and nitrate concentration was less successful. The authors suggest that the model is not applicable for trophic status prediction but do have a number of other practical applications.

The REMDSS model developed by Rossouw (1990) also has the limitation that it does not simulate the in-lake situation in a reservoir when it is regulated by other variables than phosphorus. Rossouw (2000) addressed these limitations with a further project, where other sub-models were investigated to determine the applicability to the previous shortcomings of REMDSS. These shortcomings included the predictability in reservoirs where algal growth is limited not by phosphorus, but either by light (because of the turbidity in a system) or by nitrogen.

Most of these models are aimed at predicting algal growth and not specific phytoplankton or cyanobacterial species. Since the empirical models uses a limited number of variables to predict Chl *a* concentrations, it is natural to draw the conclusion that the models only provide a rough estimate with a low degree of precision, with serious over-simplifications (Karyl & Soyupak 2001). The complex processes that lead to eutrophication cannot be explained by simplistic approaches. The SALMO model used by Recknagel (1997 & 2001) is a process-based deterministic model that uses inflow, PO_4-P and NO_3-N concentrations in the inflow, solar radiation and temperature to determine phytoplankton biovolume, functional algal groups, zooplankton bio-volume and oxygen concentrations. The simulation library of SALMO-OO utilised 25 lake databases from Europe, Asia and Australia to assemble best performing model structures for the following 5 categories of lakes: Mediterranean hypertrophic, Mediterranean mesotrophic, Temperate hypertrophic, Temperate mesotrophic, and Temperate oligotrophic. The study by Recknagel (1997) showed that it was even possible to determine the different driving forces of for e.g. *Microcystis* and *Oscillatoria* species. The characteristics of the artificial neural network and fuzzy logic methodology that allows learning complex systems, and predicting high precision responses, make this model applicable to what will be attempted with this study.

One of the research needs identified by Steel (1997), Lawrence *et al.* (2000) and Harding (2006) are the development of improved simple dynamic models that will enable managers to link reservoir algal responses with catchment runoff, stream flow, nutrient delivery processes and reservoir draw down and discharge levels. The dynamic models to determine catchment

runoff from rainfall patterns (Midgley & Pitman 1978, Schulze 1984, Smithers *et al.* 2002, Smithers & Schulze 2006, Van Heerden 1978) in order to design flood hydrographs have been refined over many years. The development of useful harmful algal bloom predicting models are still lacking.

Evolutionary Algorithms uses mechanisms inspired by biological evolution, e.g. reproduction, mutations etc. Evolutionary algorithms consistently perform well approximating solutions to all types of problems because they do not make any assumption about the underlying landscape (Wikipedia 2007). The use of evolutionary algorithms can enable the modeller to discover predictive rule sets for Chl *a* dynamics and other algal groups or species, which can lead to improved simple dynamic or even complex dynamic models (Cao *et al.*2006).

The literature survey has shown that many years of research improved our knowledge regarding the most important variables that contribute to cyanobacterial bloom-formation and many atempts has been made to develop predictive capabilities within the eutrophication field with variying degrees of success. This study will test a number of different types of modelling techniques to determine the most applicable method to enhance our predictive capability regarding harmful algal and cyanobacterial bloom forming genera in South Africa.

CHAPTER 3

SAMPLING SITES

The variability of southern Africa climate that cause extreme variability in river flow necessitate the development of man-made lakes or reservoirs to supply the growing water demand of South Africa. The existence of the large dams in the primary and secondary catchments of South Africa has produced a lentic environment of considerable size and limnological significance (Allanson 2004). Reservoirs are used extensively as fresh water resources for human consumption, recreational, agricultural and industrial uses. Management there off is essential for the sustainable use of these developed resources.

The five Reservoirs included in this study all lie within the Crocodile West/Marico Water Management Area and will in future be managed by the Crocodile West/Marico Catchment Management Agency. The Catchment area falls within both the Gauteng and North West Provinces of South Africa (Fig. 3.1).

3.1 Location of the Sampling Sites

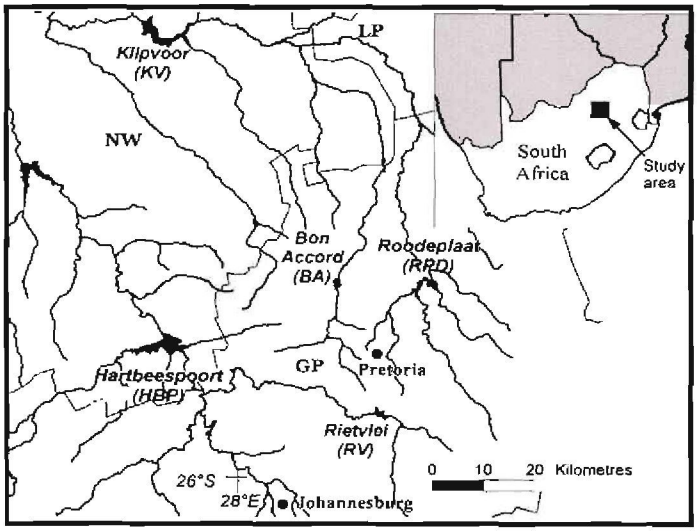


Figure 3.1 The location of the five Reservoirs that were used as sampling sites for the study. NW = North West Province and GP = Gauteng Province.

3.2 *Climatic conditions of the Sampling Sites*

A summary of the hydrological and climatological components are shown in Figures 3.2 to 3.5. These figures give an indication that in general the Reservoirs have similar climatic conditions, with warm wet summers and dry, cool winters. The North West Province is slightly warmer and dryer than the Gauteng region. One should remember that this is a general overview for the two provinces. The relative humidity never goes below 50 percent. Local effects and differences may be dampened because of the smoothing effect of the simulation of models in determining the general overview (Schulze, 1997).

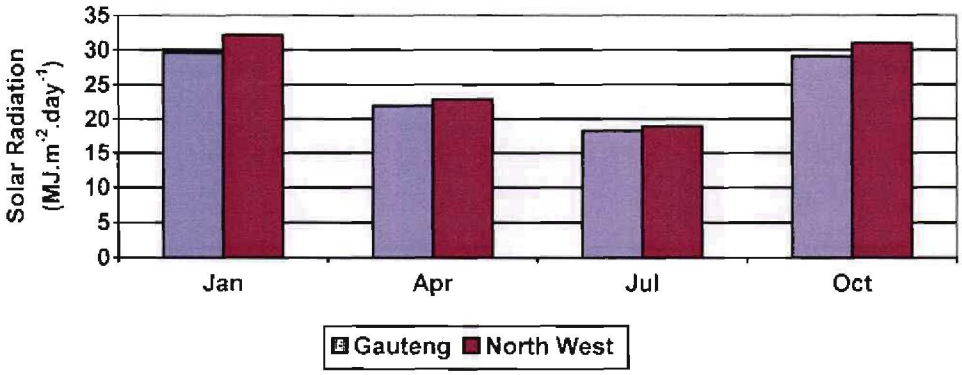


Figure 3.2 The monthly mean solar radiation in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997).

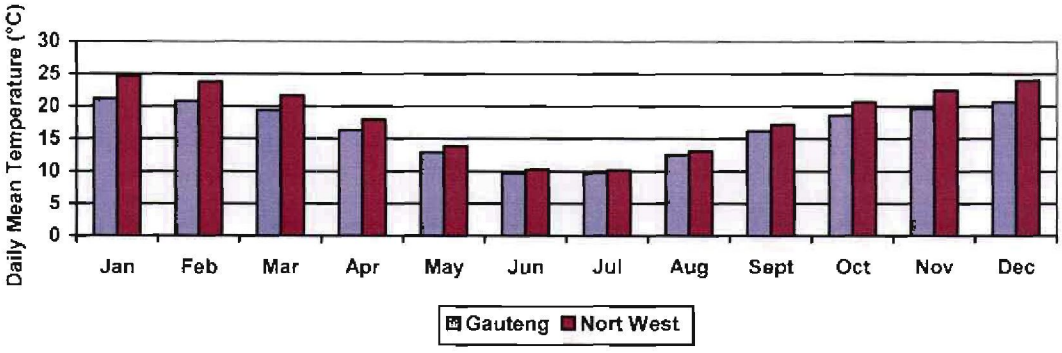


Figure 3.3 The mean daily air temperature in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997).

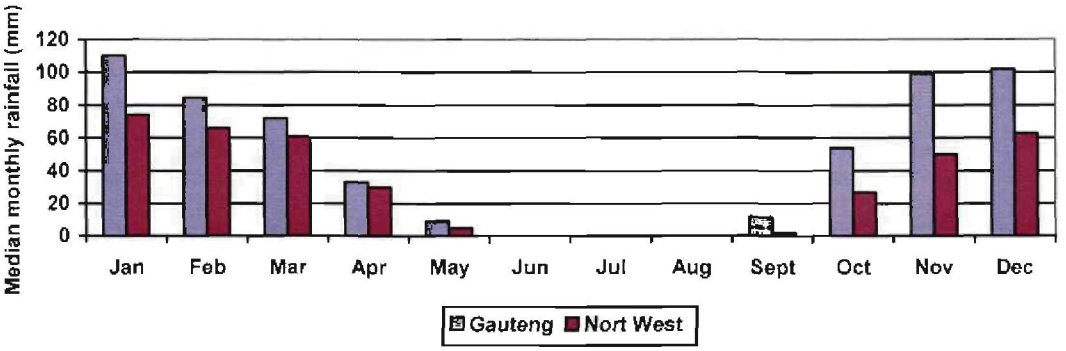


Figure 3.4 The means of the median monthly rainfall in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997).

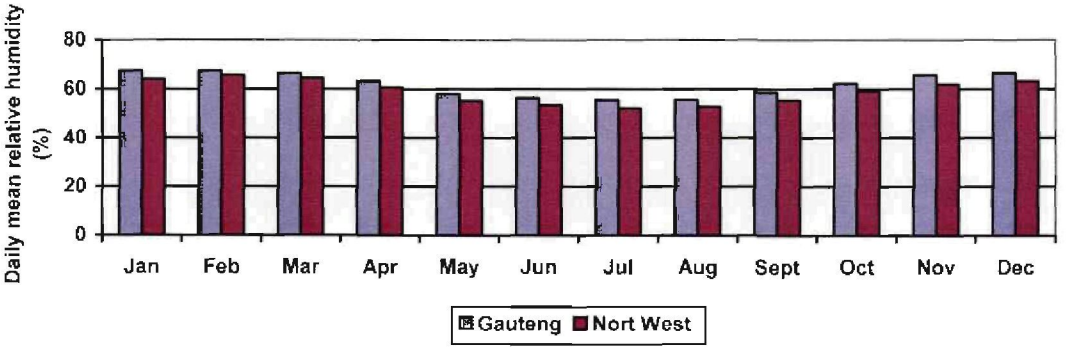


Figure 3.5 The mean daily relative humidity in the Gauteng and North West Provinces of South Africa (compiled from Schulze 1997).

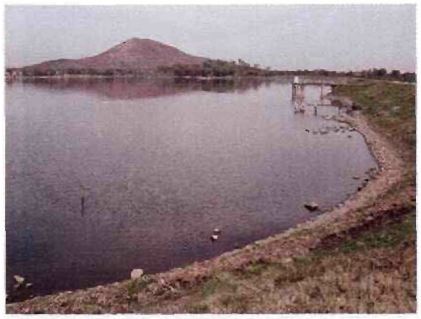
3.3 *Previous Studies and background of the Sampling Sites*

The selected dams which are all within or downstream of the main urban areas of Gauteng, are subject to severe eutrophication and the formation of seasonal cyanobacterial blooms. Of these Reservoirs Rietvlei was identified as the most eutrophied Reservoir (Van Ginkel *et al.* (2001), however, none of these Reservoirs has such pronounced cyanobacterial blooms as is experienced in the Hartbeespoort Dam (Zohary & Robarts 1989, Harding *et al.* 2004). Roodeplaat Dam is the only other Reservoir in South Africa that at this stage shows eutrophication symptoms of similar severity as experienced in the Hartbeespoort Dam (Van Ginkel *et al.*, 2000b).

Table 3.1 Morphological and hydrological characteristics of Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs (Department of Water Affairs 1980, Stofberg 1983, Department of Water Affairs 1986) NA = Not available

Reservoir	Bon Accord	Hartbeespoort	Klipvoor	Rietvlei	Roodeplaat
Latitude	-25.6214	-25.724722	-25.131111	-25.876389	-25.622
Longitude	28.18889	27.85	27.811111	28.265278	28.373
Climate	Temperate	Temperate	Temperate	Temperate	Temperate
Trophic Status	Hypertrophic	Hypertrophic	Hypertrophic	Hypertrophic	Hypertrophic
Volume (fsl) (10^6m^3)	4.293	195	43.8	12.88	41.9
Level (asl) (m)	1203	1162	989	1476	1 314
Maximum depth (m)	7.4	32.5	20.9	19	43
Mean depth (m)	3.6	9.6	5.78	6.2	10.6
Surface area (km^2)	1.7	20	7.58	2.06	3.97
Catchment Size (km^2)	315	4112	6138	492	690
Mean annual residence time (year)	NA	0.51	0.32	NA	0.69
Mean air temperature (Min/Max) (°C) (2000-2005)	0.8/35.0	-3.5/33.8	-1.6/36.5	-3.4/28.7	1.8/36.2
Water surface temperature (Min/Max) (°C)	10.6/29.3	14.4/25.7	12.8/30.2	10.1/30.7	15.2/27.8
Circulation Type	Warm Dimictic	Warm Monomictic	Warm Monomictic	Warm Monomictic	Warm Monomictic
Sedimentation accumulation (% of fsl)	33.3	15.57	3.28	5.23	NA
Catchment Composition (%)					
Urban developments	43.2	18.3	7.1	21.2	29.4
Agricultural	22.7	4.5	24.9	35.0	8.7
Natural veld	34.0	76.7	68.0	43.5	61.5
Industrial	0.1	0.4	0.0	0.3	0.4

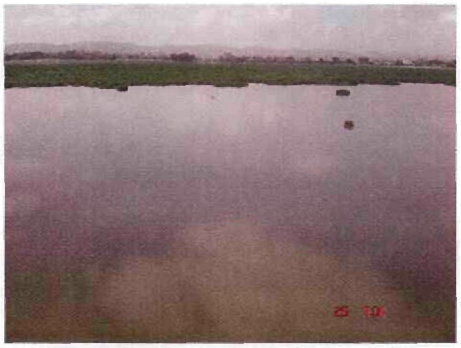
3.3.1 Bon Accord Dam



The Bon Accord Dam is situated on the Apies River 15km north of Central Pretoria. The dam wall was completed in 1925. The dam was build to supply irrigation water to areas for vegetable farming. It consists of a 'soil' wall with the overflow on the side. In 1937, the overflow was lowered by 2.13m (Department of Water Affairs 1980). The natural runoff is supplemented by storm water runoff, of the canalised Apies River, as well as industrial and domestic effluent from Pretoria (Tshwane municipality). The Daspoort Waste Water Treatment Works contributes a major component of the inflow into the Bon Accord Dam, particularly during the dry season.

According to Van Ginkel *et al.* (2000a), Bon Accord is a well-mixed system with no definite thermocline indicating polimictic conditions. However, anaerobic areas do exist in the deeper areas as shown by the oxygen profiles, probably due to organic breakdown near the sediments. During the period 1983 to 1999 there was a noticeable decrease in both phosphorus and nitrogen concentrations since the implementation of the 1 mg/L P effluent standard at the Daspoort Waste Water Treatment Works. However, the Bon Accord Dam is still classified as eutrophic to hypertrophic and was the fifth most eutrophic system in South Africa according to the prioritisation of Van Ginkel *et al.* (2000a).

Bon Accord is occasionally dominated by cyanobacterial blooms. During this study, the Bon Accord was often dominated by dinoflagellate blooms during the spring to summer periods. The system is also under threat from water hyacinth (*Eichhornia crassipes*) that has been managed by using biological control, manual control and chemical herbicide dosing.



The depth to volume relationship of the Bon Accord Dam in Figure 3.6 shows the depth distribution of the water volume in the Bon Accord Dam at full supply level.

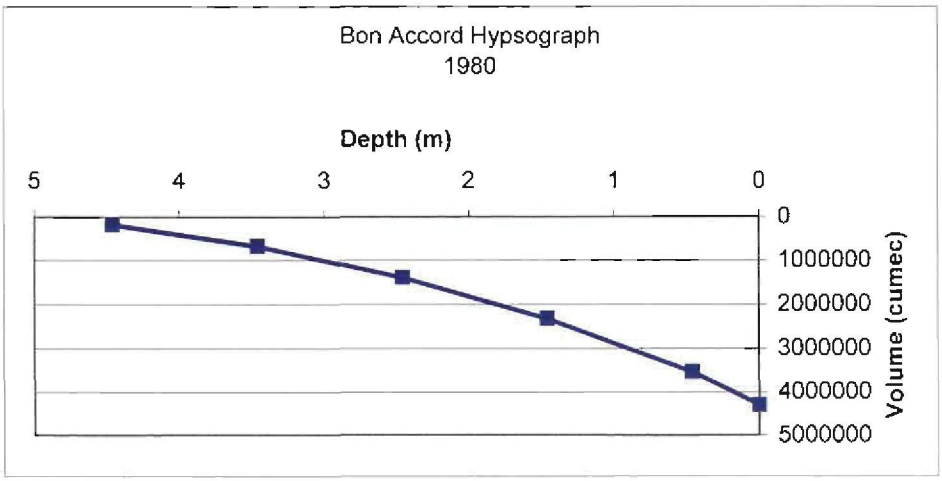


Figure 3.6 The depth to water volume graph of the Bon Accord Dam that is related to the depth to area relationship for the Reservoir.

3.3.2 Hartbeespoort Dam

Hartbeespoort Dam was build and completed in 1925, primarily for irrigation use by downstream farmers. Since then, the uses of the system increased to include domestic and recreational use. Two perennial rivers drain into the Hartbeespoort Dam, namely the Crocodile and the Magalies Rivers. Of these two main rivers, the Crocodile River, which contributes the largest volumes, is draining the highly populated area of Johannesburg and Pretoria with sixteen wastewater treatments works in the catchment. The Magalies River, on the other hand drains almost entirely modestly intense agricultural land. There are also two smaller seasonal tributaries, namely the Swart Spruit and the Leeu Spruit. The morphological and hydrological characteristics of the Hartbeespoort Dam are shown in Table 3.1.

Hartbeespoort Dam is one of the Reservoirs of South Africa that has been studied extensively (Robarts 1984; NIWR 1985; Chutter 1989; Zohary & Robarts 1989; Chutter & Rossouw 1991; Van Ginkel *et al.*2000a; Van Ginkel *et* al. 2002, Harding *et al.* 2004a&b). The Reservoir is a hypertrophic Reservoir and is internationally known for the abundance and extent of its phytoplankton population, especially the mass occurrences of *Microcystis* sp. This cyanobacterial species and its ability to form hyper-scums have been notorious since the middle 1980's (Zohary & Robarts 1989). During the summer of 1988/89 *Microcystis* sp. scums were to be treated or managed with destratification by means of aeration. This was not implemented as *Microcystis* sp. disappeared from Hartbeespoort Dam. There was, however, speculation that *Microcystis* may return in abundance (Chutter & Rossouw 1991); which it did since 2000 (McKay 2000 *Pers Comm.*)

Chutter & Rossouw (1991) found that the total phosphorus concentrations declined from 0.7 mg/l since 1983 and 1984 to 0.2 mg/l in 1988 and to 0.14 mg/l in 1989 and 1990. Van Ginkel *et al.* (2000a) noted that the mean total phosphorus concentration only during 1992 and 1993 was higher than 0.13 mg/l. The study by Chutter and Rossouw (1991) also found that there was a decline in nitrogen concentrations from 1986 to 1990 with TN:TP ratios changing from 5 (1980 – 1985) to 10 (1986 – 1988) to 17 (1989) and 21 (1990). According to the study by Xie *et al.* (2003) the change in TN:TP ratio is rather the effect than the cause for the changes in phytoplankton composition. The Van Ginkel *et al.* (2000a) report showed that the TN:TP ratios were never below 10 during the period 1989 to 1999 and only when nitrogen became limiting did *Microcystis* disappear from Hartbeespoort Dam. Chutter & Rossouw (1991) also found during their study on this dam that the zooplankton species composition shifted from a *Daphnia* and *Ceriodaphnia* dominant system to a system dominated by cyclopoid copepods, which is a predator and would seriously affect the community structure and the eventual impact of the zooplankton on the phytoplankton community.

The study by Van Ginkel *et al.* (2001) highlighted that Cyanobacteria and in particular, *Microcystis*, is a seasonal phenomenon in the Hartbeespoort Dam. The extent of the bloom and the dominance of *Microcystis* just varies over time.

Concerns about the eutrophication situation in the Hartbeespoort Dam were again raised during 1999 when fish kills occurred in the Hartbeespoort Dam and taste and odours became a problem in the Hartbeespoort Water Treatment Works. A study by Resource Quality Services (the then Institute for Water Quality Studies of DWAF) was initiated and highlighted a bloom of *Ceratium hirundinella*, a dinoflagellate that had not previously been found in the system (Van Ginkel *et al.*2001).

The depth to volume relationship of the Hartbeespoort Dam in Fig 3.7 shows the depth distribution of the water volume in the dam and highlights the depth of the Hartbeespoort dam, the second deepest dam in this study.

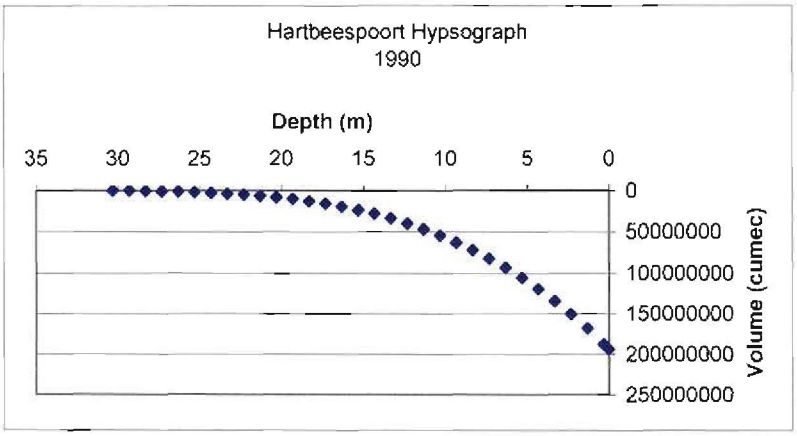
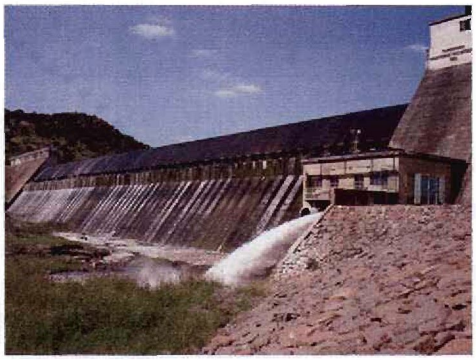


Figure 3.7 The depth to water volume graph of the Hartbeespoort Dam that is related to the depth to area relationship for the Reservoir. (aquired from the Department of Water Affairs and Forestry).

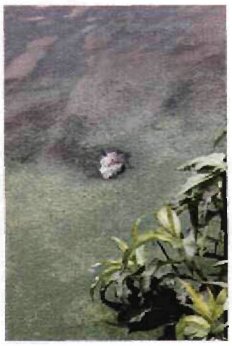
During the study period from 2001 to 2005, the water levels in the Hartbeespoort Dam reached low levels.

3.3.3 Klipvoor Dam



The Klipvoor Dam is situated on the Pienaars River downstream of both the Bon Accord and the Roodeplaat Dams. The dam was build to regulate the flow of the Pienaars River, in order to provide an assured water supply for irrigation downstream along

the Pienaars and Crocodile Rivers. The dam was proposed to provide domestic water to the following towns: Ga-Rankuwa, Uitvalgrond, Temba, Leboneng, and the South African Police Training Centre at Hammanskraal. The construction of the dam wall was completed in 1970.



The catchment comprises mainly extensive rural informal settlements and agricultural land. The Reservoir lies within a nature reserve and is used for recreational fishing. Water from the Reservoir is used downstream for irrigation. Regular fish kills has been reported in the Reservoir. These have been related mainly to low oxygen and high ammonia concentrations during increased temperature and high pH conditions within the Klipvoor system, which is a result of high organic loads.

The right flank at the Klipvoor dam consists of red, coarse-grained poryphyritic Bosveld granite and white quartzite, known as the main quartzite of the Smelterskop stage of the Pretoria series, on the left side.

The depth to volume relationship of the Klipvoor Dam in Fig 3.8 shows the depth distribution of the water volume in the dam and again highlights the depth of the dam.

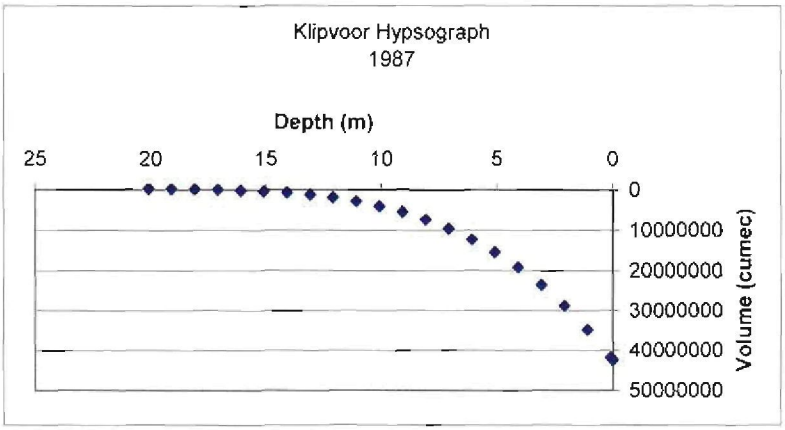
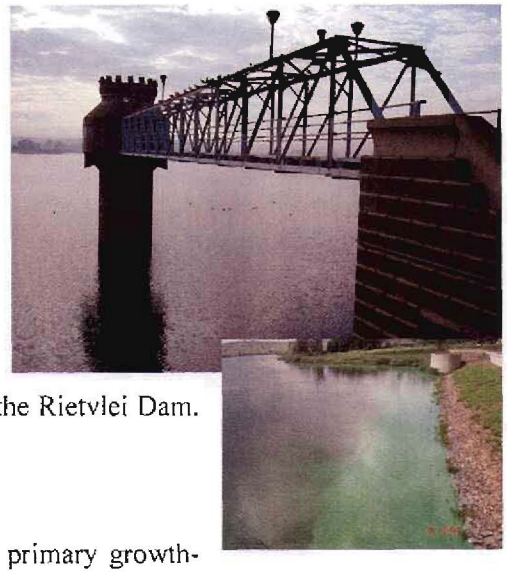


Figure 3.8 The depth to water volume graph of the Klipvoor Dam that is related to the depth to area relationship for the Reservoir.

3.3.4 Rietvlei Dam



The Rietvlei Dam was completed in 1934 on the Hennops River, southeast of Pretoria. The dam is situated in the Tshwane (Pretoria) municipal (Steyn *et al.* 1975). The main purpose of the water was for domestic supply to Pretoria. The dam was already identified by Toerien *et al.* (1975) as an eutrophied system. Treated secondary sewage effluent from the City of Kempton Park is discharged into the Swart Spruit that flow into the Hennops River upstream of the Rietvlei Dam. Massive algal blooms have been reported since 1972.

Steyn *et al.* (1975) has noted that nitrogen was the primary growth-limiting nutrient and suggested that there is a high potential for blooms of nitrogen-fixing cyanobacteria, e.g. *Anabaena circinalis*. However, *Microcystis* also regularly occurs in the system (Van Ginkel, 2000a)

The depth to volume relationship of the Rietvlei Dam in Fig 3.9 shows the depth distribution of the water volume in the dam and again highlights the depth of the dam.

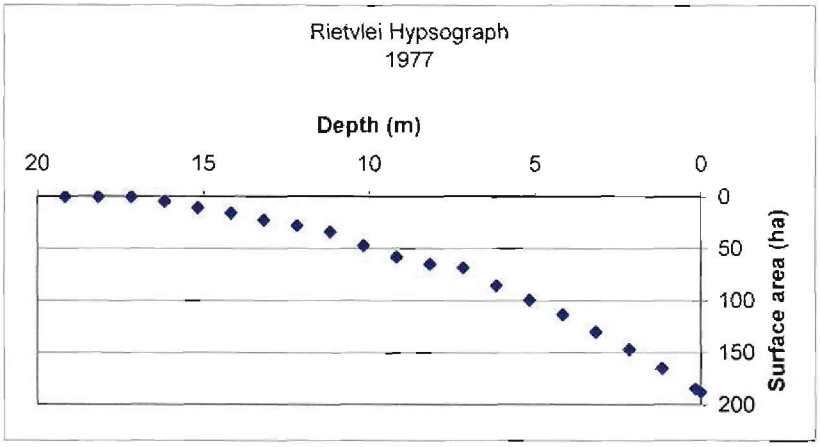
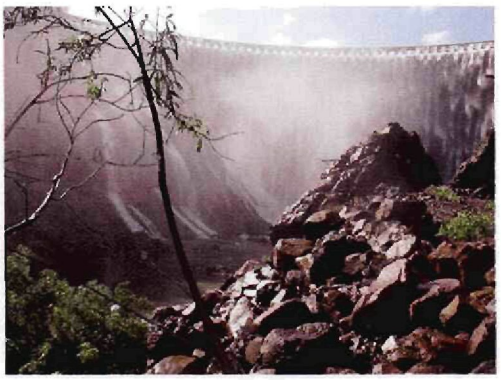


Figure 3.9 The depth to water volume graph of the Rietvlei Dam that is related to the depth to area relationship for the Reservoir.

3.3.5 Roodeplaat Dam



The Roodeplaat Dam is situated at the confluence of the Pienaars River, the Moreleta/Hartbees Spruit, and the Edendale Spruit, 20 km northeast of Pretoria. These rivers drain the highly populated areas of northern and eastern Pretoria. The catchment includes two point sources of pollution, namely, a) the Baviaanspoort Water Care Works (WCW) situated on the Pienaars River, and b) the Zeekoegat WCW that discharges directly into the Roodeplaat Dam at the Rowing/Canoe Club. Diffuse sources include formal and informal settlements, as well as agricultural activities along the banks of the tributaries (Hohls *et al.* 1998).

The Reservoir is used extensively for recreational purposes. Recent studies stated that the Roodeplaat Dam is highly eutrophic (Hohls, *et al.* 1999; Van Ginkel *et al.* 2000a; Van Ginkel *et al.* 2000b). The Tshwane Municipality in association with the Department of Water Affairs and Forestry have completed a new drinking water treatment works that is under management of Magalies Water and is operational since 2005. The Reservoir is an important recreational resource (fishing; local, national & international rowing, canoeing, and other water related uses) for the greater Gauteng area. The water is, therefore used for direct contact - and semi-contact recreation, agricultural use downstream and domestic use in Pretoria.

The investigation by Van Ginkel *et al.* (2000), after notification of a cyanobacterial bloom in the Roodeplaat Dam, has shown that there was a gradual increase in the mean annual total phosphorus and ortho-phosphate concentrations since 1991. Roodeplaat Dam was classified as eutrophic in 1989 but has since been classified as a hyper-trophic system (Van Ginkel *et al.* 2000, Van Ginkel *et al.* 2001)

The depth to volume relationship of the Roodeplaat Dam in Fig 3.10 shows the depth distribution of the water volume in the dam and again highlights the depth of the dam.

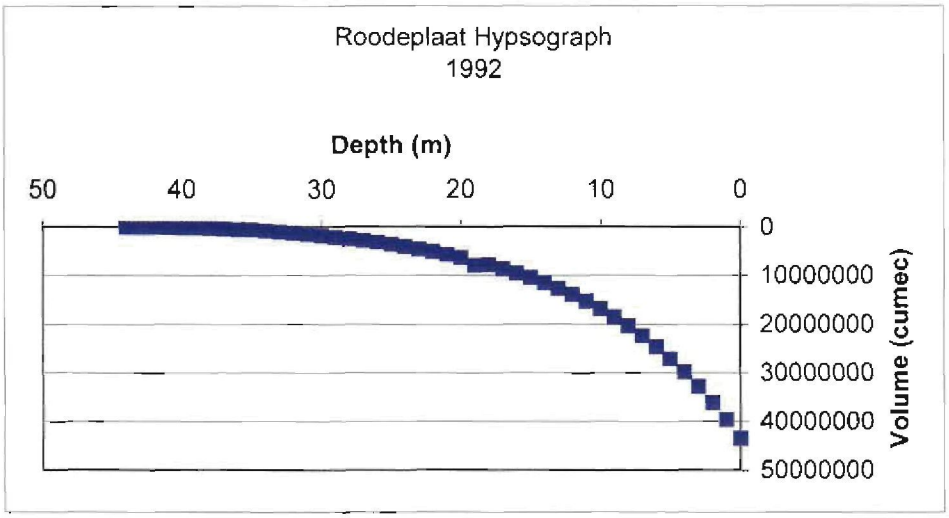


Figure 3.10 The depth to water volume graph of the Roodeplaat Dam that is related to the depth to area relationship for the Reservoir.

CHAPTER 4

DATA COLLECTION AND ANALYSIS

In this chapter, Chapter 4, the data collection and analysis are discussed. The Chapter is divided in two sections, a) the limnological and climatological data collection and analysis and b) the statistical analysis of the data.

4.1 Limnological & Climatological Data Collection Methods

The sampling sites for the five Reservoirs in this study, Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat dams (Fig. 3.1), are taken at a site close to the dam wall and is the sampling site for the National Eutrophication Monitoring Program (NEMP) of South Africa (DWAF 2002). Integrated 0-5m hosepipe samples were taken every two weeks at Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat reservoirs (Gerber *et al.* 2004a). The Bon Accord Dam is too shallow for an integrated hosepipe sample and, therefore subsurface grab samples were taken. Three of the Reservoirs (Hartbeespoort, Rietvlei and Roodeplaat) are used extensively for drinking water abstraction, and all five are used for recreation.

4.1.1 *Limnological Sampling*

The water quality samples from each Reservoir are subdivided into a macro inorganic sample, an algal identification sample and a Chl *a* sample. The macro (major inorganic chemicals) were analysed on automatic analysers at Resource Quality Services of the Department of Water Affairs and Forestry.

The macro chemical variables measured are: calcium, chloride, dissolved major salts (DMS), electrical conductivity (EC), fluoride, potassium, magnesium, ammonium, nitrate & nitrite, sodium, ortho-phosphorus, sulphate, silica, total alkalinity and pH. Additional to the normal macro samples the Kjeldahl nitrogen (KN) and total phosphorus (TP) are measured using digestion methods. (DWAF, 2006a & 2006b). To determine the dissolved inorganic nitrogen (DIN) concentrations the sum of the ammonium, nitrate and nitrite was calculated. To determine the total nitrogen (TN) concentrations the sum of the Kjeldahl nitrogen, nitrate and nitrite concentrations were calculated.

Algal identification samples are preserved with Lugol's solution. For identification, 10 ml sample are poured into a counting chamber and left to settle overnight or for a minimum of 6 hours. Algal identification is done with an inverted microscope and results are expressed as a percentage of the total algal population. The method is described in detail in the methods manual of the RQS (Department of Water Affairs and Forestry, 1997).

For Chl *a* determination, at least 500 ml of sample water are filtered through a 45µm Whatman filter paper. The Chl *a*, is then extracted from the filter paper into 10 ml ethanol. The Chl *a,* content is then measured on the extract using a spectrophotometer (Department of Water Affairs and Forestry, 2003a).

On each sampling occasion, *in situ* temperature and oxygen measurements were taken at 1m depth intervals with an YSI meter. Secchi disc readings were taken regularly at the sampling site in each Reservoir by lowering the 20cm black and white disc into the water and measuring the depth at which the disc is not visible anymore. The Secchi disc reading was used to determine the extinction coefficient as follows:

Extinction coefficient = 2.7/SD

Lawrence *et al.* (2000)

The rationale when sampling for total microcystin was that a whole Reservoir was considered one site. Wind action and currents cause cyanobacteria to accumulate in patches, therefore, during each sampling event (once every two weeks), five different sites where cyanobacteria accumulated around the edges of a Reservoir were found and sampled. This means that the sites for the toxins were not necessarily the same sites during the study period. This method of sampling was decided on as a regular sampling site does not always portray toxicity of cyanobacteria because of the phenomenon of shifting masses of cyanobacteria. One litre, subsurface grab samples (Gerber *et al.* 2004b) were taken to the laboratory for analysis. The samples were then frozen and thawed three times to lyse the algal cells and release the intra-cellular toxins. Total microcystin concentration was determined for the shifting five sites with the ELISA method (Department of Water Affairs and Forestry 2003b) to determine if toxins are present and to determine maximum, minimum and mean concentration in a Reservoir. This was done to determine correlations between the regular sampling results and the presence of

cyanobacterial toxins. This will increase the knowledge on the hazard when cyanobacteria are present in a reservoir, which is used for recreational and drinking water purposes.

4.1.2 Determination of the depth distribution of cyanobacterial toxin and algal species dominance

In two of the Reservoirs, Hartbeespoort and Roodeplaat Dams, the depth distribution of total microcystin was determined once a month for the period October 2003 to September 2006. Samples were taken with a Van Doorn sampler at 0m, 1m, 2m, 5m, 10m, 15m and 20m depths. These results were used to determine the depth profile of total microcystin in the two Reservoirs.

Algal identification and Chl *a* samples, were taken at the same depths as the total microcystin samples. Phytoplankton samples were preserved with Lugol's solution and were identified under an inverted microscope. An estimation of the dominance (%) of the phytoplankton genera as described in section 4.1.1 was used to determine the frequency occurrence of the cyanobacterial genera (*Microcystis*, *Anabaena*, *Oscillatoria*, *Pseudoanabaena*, *Cylindrospermopsis*, etc.) in the five reservoirs. Chl *a* concentrations, were determined by filtering a known volume of water through a 45um Whatman filter paper and the concentration of Chl *a* was determined spectro-photometrically between wavelengths 665.5 and 750 (Department of Water Affairs and Forestry 2003a).

The data are presented with Microsoft Excel Surface graphs that could show the temporal and spatial distribution of the different variables.

4.1.3 Climatological data collection

The South African Weather bureau and the Institute for Soil, Climate and Water provided the meteorological information for the period 2000 – 2005. This data was used as input for the primary five-year dataset. The meteorological sites used for the Bon Accord Reservoir is primarily the Pretoria NBI site (19991) with filling in of data with the Roodeplaat-AGR site (17254) and the Pretoria, Roodeplaat Site (30093). The meteorological site used for the Hartbeespoort Reservoir is the Brits Agricultural Site (17085). The meteorological sites used for the Klipvoor Reservoir are the Borakalalo Site (05495196) and the Brits Agricultural Site (17085). The meteorological site used for the Rietvlei Reservoir is the Johannesburg International

Airport Site. The meteorological information for the Roodeplaat Reservoir was extracted from four sites (Pretoria, Roodeplaat (30093); Pretoria National Botanical Inst. (30687); Roodeplaat–AGR (17254) and Irene (0153385)).

The variables provided were daily rainfall, daily maximum and minimum temperatures, solar radiation, sunshine hours and relative humidity. Calculated variables included total rainfall, mean average maximum temperature and mean average minimum temperature for the 14 days prior to sampling because of the potential impact on the development of specific phytoplankton genera.

4.1.4 Multivariate Statistical Analysis

When investigating variations in plant communities across a wide variety of environmental conditions, one often finds large differences in the plant communities but there is also a certain degree of consistency or predictability in the variation (Lepš & Šmilauer 2003). This consistency and variability are not always apparent as a visible gradient. To enable one to understand the variability and predictability of a community, as well as to determine the most important environmental and community components, ordination methods are available tools (Lepš and Šmilauer 2003). Ecologists have used these tools since the 1950's and during their evolution these methods have developed into a mixture of various methods that are widely used.

For this study, the CANOCO methods of ordination and constrained ordination were used to summarize community patterns and compare the suggested gradients within the researchers independent knowledge of environmental conditions. These methods also enable one to statistically test the predictive power of such knowledge.

The primary data set was obtained, as described above. This data set contains a collection of observations and analytical results. This primary data set included the cyanobacterial species composition and the different algal groups' composition in the five Reservoirs, Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat. For the purpose of the multivariate analyses this primary data, or response variables, are associated with explanatory variables. This data is divided into two groups. The first are the environmental variables and refers to the variables that are important in the role of predictors of community response (Lepš and Šmilauer 2003, Ter Braak). The other group represents the co-variables (or covariates) that are explanatory variables

with a known influence on the response variables (Lepš and Šmilauer 2003). In this instance, the algal genera are the response variables.

The statistical models used on the multivariate data set from the five Reservoirs included a) indirect gradient analysis using the most prominent or the principle component analysis (PCA) and b) the direct gradient analysis using the redundancy analysis (RDA). The results include Monte Carlo tests that provides eigenvalues (EV) for a n umber of axis and the total cumulative EV, cumulative percentage variance (CPV) of the different variable sets and the significance of the results as the F-ratio and the P-value.

4.2 Results and Discussion

4.2.1 Limnological Characteristics and Climatic Conditions of the Sampling Sites

The limnological characteristics of the five reservoirs are very similar. All five Reservoirs are hypertrophic systems with water temperatures varying between 10.4°C and 32.4°C. Hartbeespoort and Klipvoor Reservoirs are slightly warmer than the other three Reservoirs. The Reservoirs are clear systems with Secchi disc readings of up to 7.2m. The much shallower Bon Accord is more turbid than the other Reservoirs and has a maximum Secchi disc reading of 1m. This may be due to the effect of wind action that sitrs up sediments in the shallow reservoir..

Bon Accord Dam has large amounts of total nitrogen in the system with a mean total nitrogen concentration of 23.9mg/L and a maximum of 167.5mg/L. The DIN concentrations in the Bon Accord varied between 0.9mg/L and 163.1mg/L. The Reservoir is just downstream of the Daspoort Sewerage Treatment Works and the extremely high nitrogen concentrations indicate that there is occasionally or regularly a breakdown in the treatment works. The median total nitrogen of all five Reservoirs varied between 0.97mg/L and 3.0mg/L. These concentrations are above the nitrogen limited threshold value of 0.1mg/L (Oliver and Ganf 2000) and none of these systems is, therefore, nitrogen limited.

When the phosphorus concentrations are considered, the ortho-phosphorus varied between 3µg/L and 4084µg/L and the total phosphorus varied between 26µg/L and 5582µg/L. These concentrations are far above the threshold values of 10µg/L and 20µg/L where ortho-phosphorus and total phosphorus respectively becomes limiting (Hesketh and Brookes 2000, Oliver and Ganf

2000). All these systems have extremely high nutrient concentrations which are thus not limiting, even if the TN:TP ratios may go below the limiting values.

Hartbeespoort Dam has the lowest concentrations of both DIP and TP, yet the extent of the algal and cyanobacterial blooms is much more intensive than what is found in the other Reservoirs. The higher temperatures at the Hartbeespoort Reservoir could be a driving force of this phenomenon. The lower DIP concentrations may be the result of uptake by cyanobacteria or other bloom forming algae.

The electrical conductivity of the Reservoirs varied between a minimum of 28.8mS/m and a maximum of 100.6mS/m. The mean electrical conductivity readings of the five Reservoirs varied between 44.9mS/m and 61.7mS/m, and show that there is little variability in the salinity of these five Reservoirs (Table 4.1). The major ion compositions were not investigated any further as the focus of the study was more on the nutrient conditions that play a significant role in algal and cyanobacterial bloom development.

When the dominant phytoplankton species are considered, the results show that in the five Reservoirs there are basically two phytoplankton groups that are dominant, namely the Cyanobacteria (in the summer to autumn periods) and the Dinoflagellates (in the winter to spring periods), except for in the Bon Accord Dam where the Dinoflagellate species (primarily *Ceratium hirundinella)* was dominant throughout the year for most of the study period. The other four Reservoirs are primarily dominated by the Cyanobacteria with occasional Dinoflagellate blooms.

It is worth mentioning that three of the study sites, namely Bon Accord, Hartbeespoort and Roodeplaat Dam were also infested with *Eichhornia crassipes* during the study period. This could have an impact on the water quality of the Reservoirs. In Bon Accord and Hartbeespoort the *Eichhornia crassipes* is well under control, but during 2005 the infestations in the Roodeplaat Dam increased to large areas due to the inefficiency of the chemical control methods at this specific site.

Table 4.1 Limnological properties of the Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs during 2000 to 2005

Reservoir	Bon Accord				Hartbeespoort			
Measured Variables	Mean ±STD	Min	Median	Max	Mean ±STD	Min	Median	Max
Water temperature (°C)	20.1±4.8	10.6	21.4	29.3	20.9±4.5	11.3	21.6	32.4
Secchi depth (m)	0.4±0.2	0.2	0.4	1.0	1.5±1.0	0.2	1.3	6.6
pH	8.3±0.5	7.1	8.3	9.6	8.5±0.4	7.2	8.5	9.7
Dissolved inorganic nitrogen (mg. L^{-1}) as N	21.6±47.8	0	1.0	163.1	1.4±0.8	0.02	1.4	3.6
Total nitrogen (mg. L^{-1}) as N	23.9±48.0	0.9	3.0	167.5	2.7±0.7	0.02	2.7	5.85
Dissolved inorganic phosphorous (ug/L) as P	163±155	12	114	1190	54.5±53.2	3.0	35.0	493
Total phosphorous (ug/L) as P	383±322	96	285	1978	112.7±59.3	26	93	493
Electrical Conductivity (mS/m)	48.6±9.3	28.8	48.7	100.6	50.1±3.7	42.1	50.3	58.1
Chl a (µg. L^{-1})	304±334	7.0	196.3	2281	46.3±59.6	0.58	26.1	617.4
Microcystis biovolume (cm^3/m^3)	19.2±51.7	0	0	278.0	14.2±24.3	0	3.3	126.6
Anabaena biovolume (cm^3/m^3)	0.3±1.7	0	0	15.7	0.2±1.1	0	0	12.0
Oscillatoria biovolume(cm^3/m^3)	0.7±3.0	0	0	25.1	0.1±0.7	0	0	4.0
Dinoflagellate biovolume (cm^3/m^3)	91±126	0.0	49.2	820.2	6.5 ±22.5	0.00	0.01	176.1
Reservoir	Klipvoor				Rietvlei			
Measured Variables	Mean ±STD	Min	Median	Max	Mean ±STD	Min	Median	Max
Water temperature (°C)	22.6±4.55	12.8	23.8	30.2	19.6±4.4	9.3	20.6	30.7
Secchi depth (m)	1.2±0.5	0.02	1.2	3.0	1.9±0.9	0.1	1.9	4.5
pH	8.6±0.5	5.7	8.7	9.5	8.5±0.5	7.3	8.4	9.9
Dissolved inorganic nitrogen (mg/L) as N	0.27±0.52	0.04	0.09	4.35	1.1±1.0	0.0	0.9	4.7
Total nitrogen (mg/L) as N	1.18±0.81	0.33	0.97	4.82	2.5±1.5	0.0	2.2	8.8
Dissolved inorganic phosphorous (ug/L) as P	482±192	34	490	1451	578±705	6.0	219.6	4084.0
Total phosphorous (ug/L) as P	699±291	56	699	2447	920±986	0.0	370.1	5582.0
Electrical Conductivity (mS/m)	61.7±10.1	29.7	61.1	81.9	52.3±7.1	37.7	53.5	64.5
Chl a (ug/L)	110.6±131.9	2.8	66.0	966.9	54.8±105.9	1.0	26.2	1290.0
Microcystis biovolume (cm^3/m^3)	4.6±13.7	0	0	89.7	10.1±23.8	0	0.2	152.4
Anabaena biovolume (cm^3/m^3)	4.2±34.2	0	0	386.8	0.2±0.9	0	0	5.6
Oscillatoria biovolume(cm^3/m^3)	0.3±1.1	0	0	9.1	0.0±0.2	0	0	2.2
Dinoflagellate biovolume (cm^3/m^3)	24.8±49.9	0.0	6.9	298.1	7.6±28	0.0	0.01	205.9
Reservoir	Roodeplaat							
Measured Variables	Mean ±STD	Min	Median	Max				
Water temperature (°C)	21.0±4.4	10.4	21.9	30.5				
Secchi depth (m)	1.9±1.0	0.1	1.7	7.2				
pH	8.8±0.5	7.0	8.8	10.1				
Dissolved inorganic nitrogen (mg/L) as N	0.8±0.7	0.0	0.6	4.5				
Total nitrogen (mg/L) as N	1.8±0.6	0.6	1.7	4.5				
Dissolved inorganic phosphorous (ug/L) as P	156.1±120.9	5.0	132.8	620.0				
Total phosphorous (µg/L) as P	238.3±139.6	12.0	208.0	712.0				
Electrical Conductivity (mS/m)	44.9 ± 3.6	37.0	45.2	57.0				
Chl a (ug/L)	37.3±27.8	1.7	32.9	197.2				
Microcystis biovolume (cm^3/m^3)	7.6±11.8	0	2.7	61.3				
Anabaena biovolume (cm^3/m^3)	0.8±3.8	0	0	25.1				
Oscillatoria biovolume(cm^3/m^3)	0.1±0.4	0	0	3.7				
Dinoflagellate biovolume (cm^3/m^3)	2.6±16.1	0.0	0.0	143.5				

4.2.2 *Phytoplankton Composition of the Sampling Sites*

To determine the dominance and the phytoplankton composition of the five Reservoirs the extinction coefficient is shown with the Chl *a* and the algal groups in each Reservoir (Fig. 4.1 to Fig. 4.5) as light penetration into the water column is essential for phytoplankton growth.

The extinction coefficient, the Chl *a* concentration and the phytoplankton composition of the Bon Accord Dam are shown in Fig. 4.1. The correlation coefficient (r^2) between the extinction coefficient and the Chl *a* concentration in Bon Accord is 0.54, showing the positive correlation between these variables. In the Bon Accord Dam, the extinction coefficient varied between $2m^{-1}$ and $9m^{-1}$. This is indicative of waters with high turbidity or excess algal matter (Wetzel 2003). Extremely high Chl *a* concentrations, are present in the Bon Accord Dam for long periods and a maximum above 2281ug/L was found during October 2003. This extremely high Chl *a* peak was associated with a Dinoflagellate bloom. From October 2003 to January 2003, dinoflagellate blooms that exceeded 1000ug/L primarily dominated Bon Accord. According to Taylor and Pollingher (1987), *Ceratium*, are widely tolerant and ubiquitous, however, most dinoflagellate species have restricted ranges regarding calcium, pH, dissolved organic matter and temperature according to Wetzel (2001).

From March 2003, large Cyanophyta blooms periodically dominated the phytoplankton. The most important Cyanophyta taxon during these blooms was *Microcystis*, but *Oscillatoria* was also occasionally found in the Bon Accord Reservoir (Fig. 4.6).

During the study period, extreme Dinoflagellate blooms dominated the phytoplankton community of the Bon Accord Dam. During the period of October 2000 to 2002, the occurrence of the Dinoflagellate *Ceratium hirundinella* was dominant, and very little Cyanobacteria were found in the phytoplankton community during this period. The *Microcystis* dominance and blooms are present from March to May in 2003, November 2003 to June in 2004 and from December 2004 to June 2005 in the Bon Accord Dam.

The extinction coefficient in the Hartbeespoort Dam seldom exceeded $4m^{-1}$, but did reach $8m^{-1}$ during the summer of 2002-2003. A peak of $6m^{-1}$ was reached during the subsequent years. The high extinction coefficient is associated with the high concentrations of Dinoflagellate blooms and suggests that the turbidity and high concentrations of suspended matter in Hartbeespoort Dam

were due to algal growth. The correlation coefficient (r^2) between the extinction coefficient and the Chl *a* concentration in Hartbeespoort is 0.54, showing the positive correlation between these two variables (Fig. 4.2).

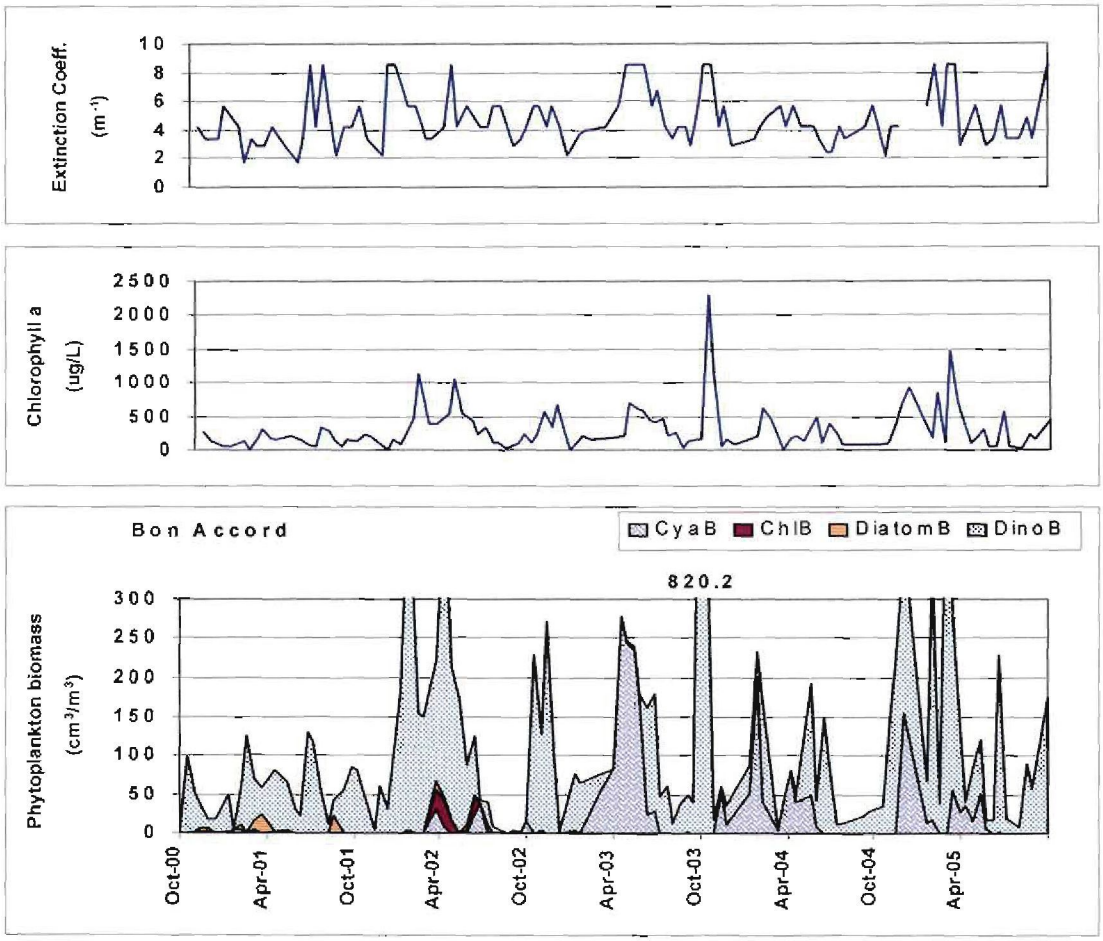


Figure 4.1 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Bon Accord Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB.

The Chl *a* concentrations only once exceeded 500ug/L during the study period. During 2000 to 2001 Chl *a* concentrations was low but started increasing during the summer of 2001-2002. For Hartbeespoort Dam the period from June 2002 to June 2003 was marked with very high Chl *a* concentrations with initially a Dinoflagellate bloom during the winter to spring period followed

by an extensive cyanobacterial bloom. After 2003 there was a slight decrease in the extent of the summer Cyanobacterial bloom (*Microcystis*), although it never reached the same levels that persisted during 2000 to 2001. The summer peaks of Cyanobacteria that consist primarily of *Microcystis* are a regular phenomenon in the Hartbeespoort Reservoir and reached crises conditions during 2003, when the DWAF regional Office initiated the physical removal of cyanobacterial matter from the system. This elevated the noxious odours around the Reservoir. Occasionally *Anabaena* and *Oscillatoria* species are also found in the system but at negligible concentrations (Fig. 4.7).

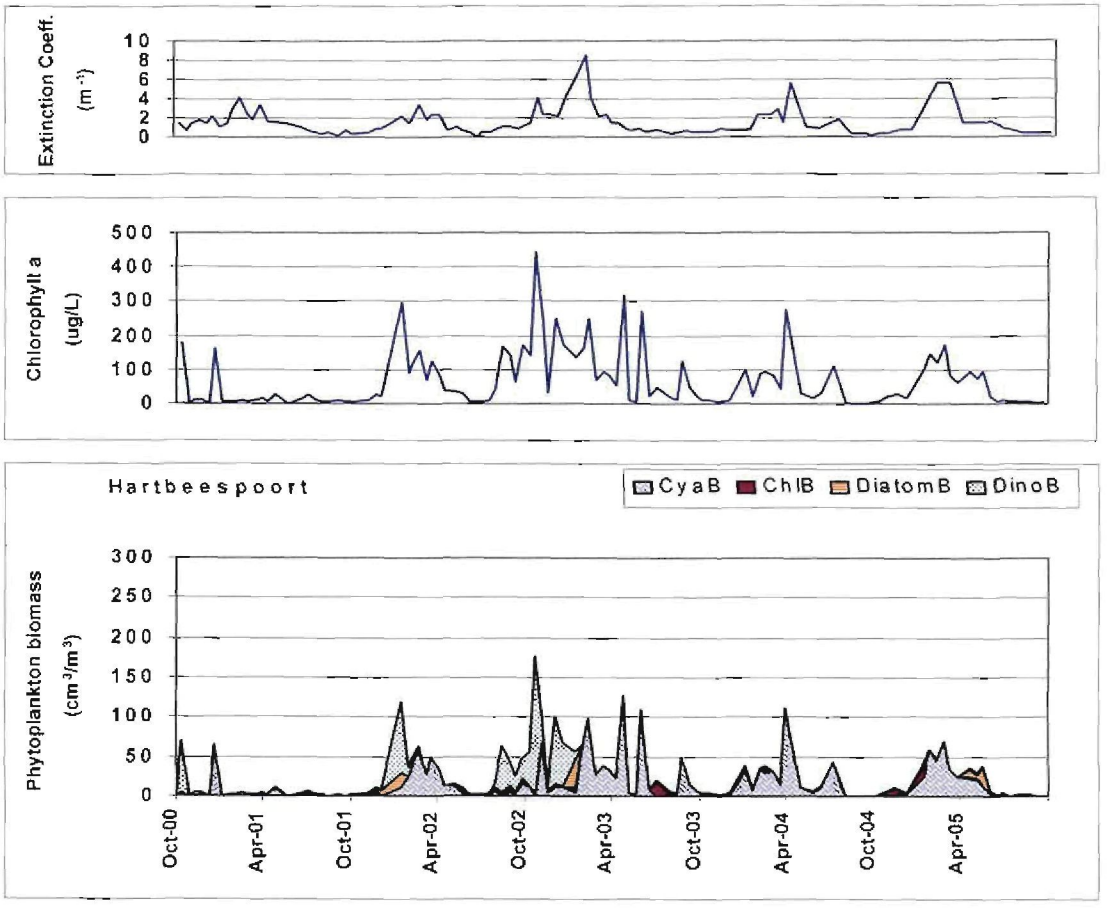


Figure 4.2 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Hartbeespoort Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB.

In the Klipvoor Dam the extinction coefficient is primarily below $5m^{-1}$ but it did reach peaks in excess of $15m^{-1}$ that were associated with the *Microcystis* and dinoflagellate blooms (Fig. 4.3). The correlation coefficient (r^2) between the extinction coefficient and the Chl *a* concentration in Klipvoor is 0.45, showing the weak positive correlation between these two variables. The Chl *a* concentrations, were regularly above 250ug/L and reached a maximum of 966.9ug/L during the spring of 2002. This was associated with a bloom of *Anabaena*. *Microcystis* is the dominant cyanobacteria genera in bloom forming conditions but also *Anabaena* and *Oscillatoria* occur occasionally in dense concentrations in Klipvoor Dam (Fig. 4.8).

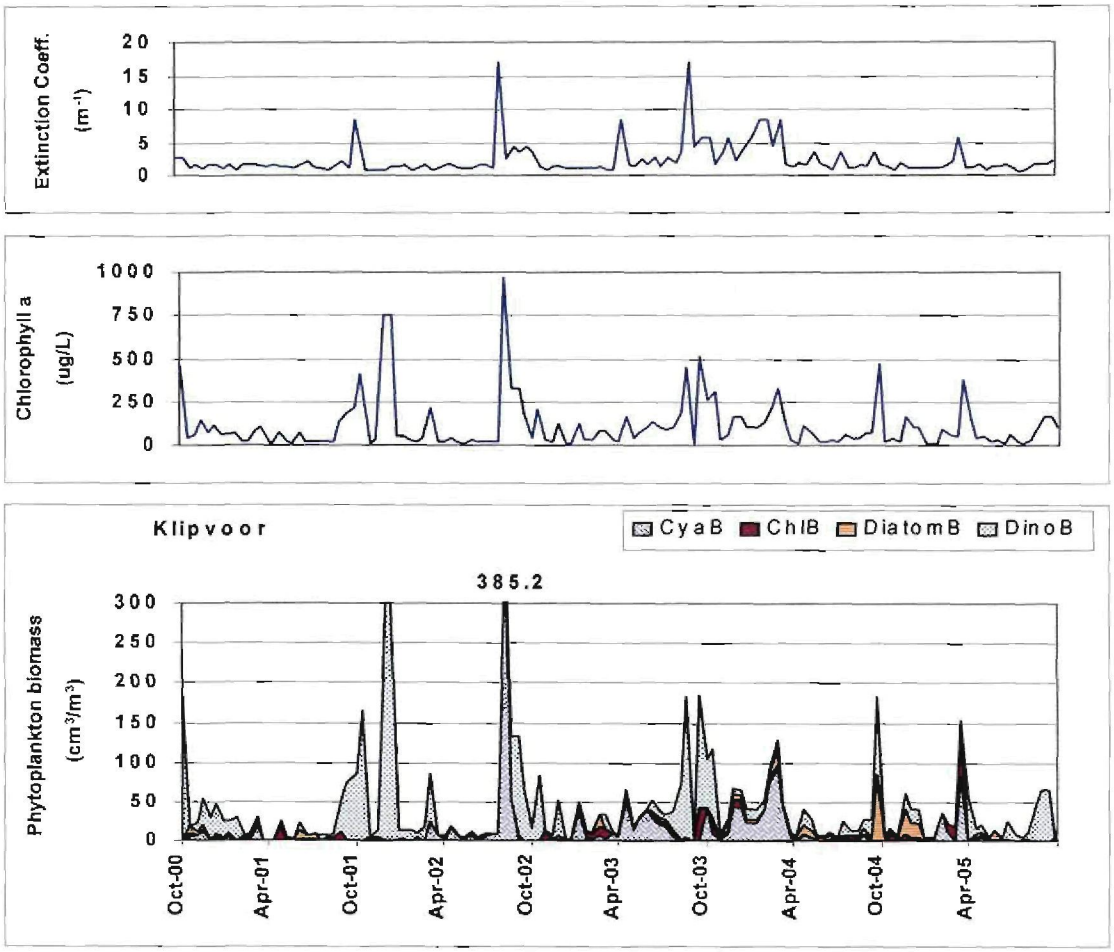


Figure 4.3 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Klipvoor Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB.

The Dinoflagellate blooms in the Klipvoor Reservoir consisted primarily of *Ceratium hirundinella*. Klipvoor is thus a mixed system regarding the dominance of phytoplankton taxa. *Ceratium hirundinella* was dominant during 2000 to 2002, while *Microcystis* became dominant from January 2003 to October 2005 and dominated during the summer periods.

The extinction coefficient in the Rietvlei Dam was generally low but occasionally during the autumns of 2002, 2003 and 2004 exceeded 5 m^{-1}. All the peaks in the extinction coefficient in the Rietvlei Dam were associated with cyanobacterial blooms. One can, therefore, speculate that this indicates that in the Rietvlei Dam light penetration into the Reservoir may be more important for cyanobacterial species as compared to the situation in the Reservoir during *Ceratium* blooms when compared to the other Reservoirs. The correlation coefficient (r^2) between the extinction coefficient and the Chl *a* concentration in Rietvlei is 0.45, showing the weak positive correlation between these two variables.

These results indicate that the Rietvlei Dam is a clear Reservoir. The autumn to spring periods reflects very low Chl *a* concentrations as compared to the summer to autumn periods and it regularly exceeded 100ug/L during the midsummer periods, reaching peaks of up to 500ug/L. This may be due to the colder winters, as the Rietvlei Dam is situated in the Highveld area of South Africa as shown by the colder minimum temperature (Table 4.1). *Microcystis* and *Anabaena* dominated the Cyanophyta blooms and *Oscillatoria* was occasionally present in noticeable concentrations in Rietvlei Dam (Fig. 4.9). At the beginning and the end of the study period the phytoplankton was dominated by dinoflagellates, primarily consisting of *Ceratium hirundinella*. The highest Chl *a* concentration was once again associated with the Dinoflagellate blooms, although the summer period of October 2002 to May 2003 had high Chl *a* concentrations associated with *Microcystis*.

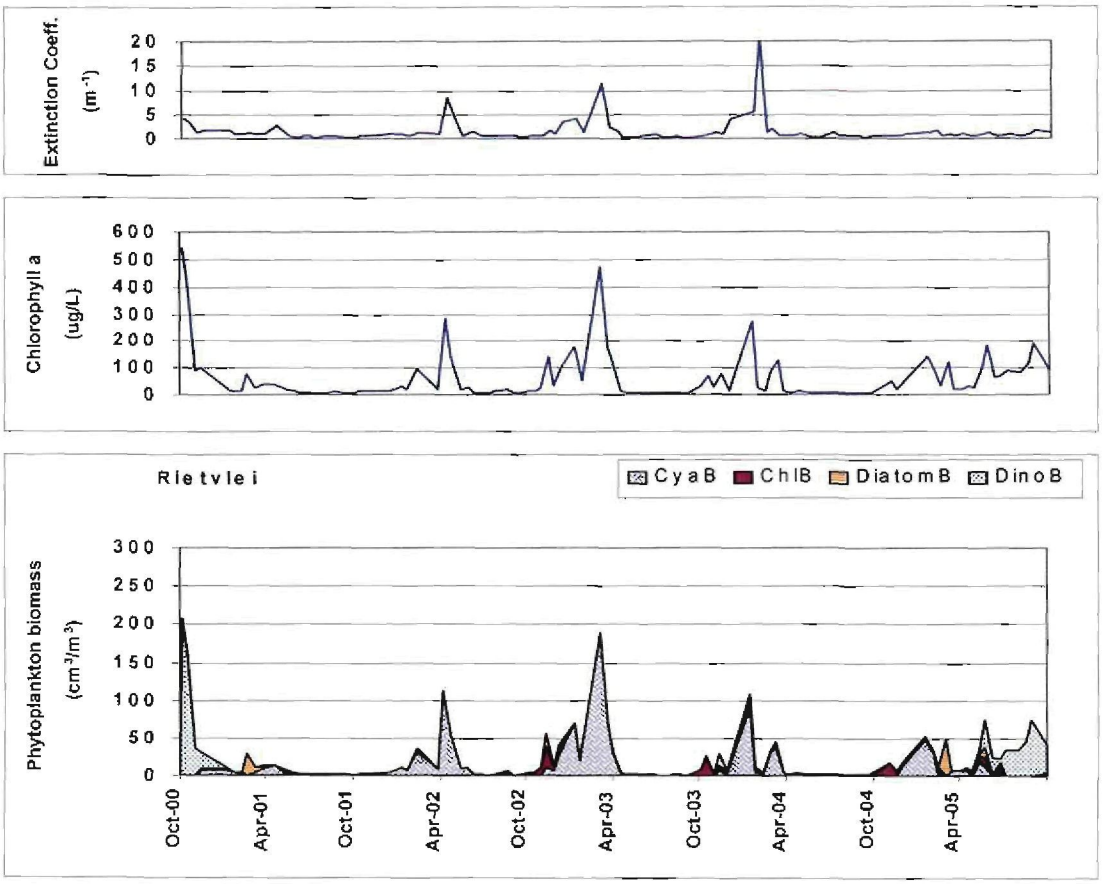


Figure 4.4 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Rietvlei Dam from October 2000 to September 2005. Cyanophyta biovolume (CyaB), Chlorophyta biovolume (ChlB), Diatom biovolume (DiatomB) and Dinoflagellate biovolume (DinoB).

The extinction coefficient in the Roodeplaat Dam was consistently below $2m^{-1}$ and only exceeded $4m^{-1}$ once during the summer of 2002 when the phytoplankton population was dominated by the Chlorophyta taxa. The correlation coefficient (r^2) between the extinction coefficient and the Chl *a* concentration in Roodeplaat is 0.64, showing the positive correlation between these two variables. The Roodeplaat Dam is thus the clearest system of all the sampling sites studied. The Chl *a* concentrations seldom exceeded 100ug/L with the highest peak of >350ug/L found in 2005, suggesting that Roodeplaat Dam was the Reservoir with the lowest primary production of the five sites. Roodeplaat Dam regularly still has a large contribution of Chlorophyta species to the

phytoplankton population, however, summer dominance by the Cyanophyta occurred consistently.

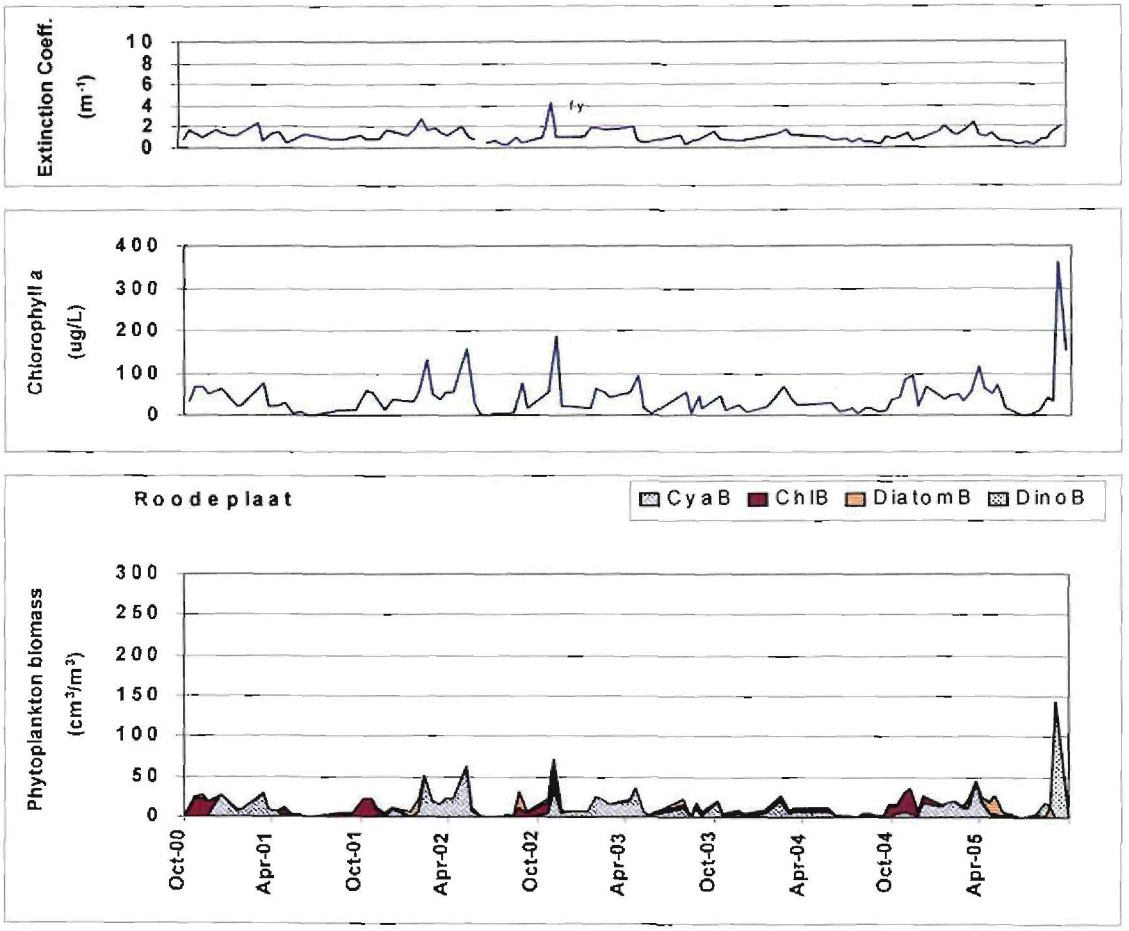


Figure 4.5 The extinction coefficient (m^{-1}), Chl *a* concentrations (ug/L) and the phytoplankton biovolume composition (cm^3/m^3) in the Roodeplaat Dam from October 2000 to September 2005. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB and dinoflagellate biovolume: DinoB.

The highest Chl *a* concentration in Roodeplaar Dam was associated with a Dinoflagellate bloom at the end of the study period. *Ceratium*, although regularly found, was found for the first time since monitoring started in Roodeplaat Dam in the early eighties, in such high concentrations. *Microcystis* is the dominant cyanobacterial taxon in the Roodeplaat Dam, but also *Anabaena* and *Oscillatoria* are occasionally present (Fig. 4.10).

4.2.3 Total Microcystin Occurrence in the Reservoirs

Total microcystin (TM) concentrations were found to be orders of a magnitude higher than in other northern hemisphere countries such as Finland and the USA (Van Ginkel *et al.* 2006). This deems it necessary to investigate the seasonality, extent and associated cyanobacterial species in the hypertrophic water resources of South Africa. The health implications of these toxins present in the recreational water resources and the implications for the drinking water industry necessitates a better understanding of the extent and occurrence of the TM toxins in South African conditions.

The total microcystin (TM) analysis was done from October 2003 to Spetember 2005. There are, thus, only two years of quantitative data available from each Reservoir during the study period. This gives a good indication of the extent of TM in the hypertrophic systems of South Africa and the cyanobacteria species that are associated with the occurrence of the total microcystin toxins present in each of the Reservoirs studied.

In the Bon Accord Dam the presence of TM, seem to be inconsistently delayed after high cyanobacterial biovolume in the Reservoir (Fig. 4.6). The TM concentrations are regularly below 2000ug/L but extremely high concentrations exceeding 7000ug/L were found in March 2005. The cyanobacterial species associated with the occurrence of TM is *Microcystis*. This association is not always direct, as TM concentrations seem to have a delayed effect when compared to the peaks in *Microcystis* biovolume (Fig. 4.6). This may be caused by the fact that the sampling sites are not at the regular monitoring site and the toxins are released at sites where excessive accumulations of the cyanobacterial biovolume are found in down wind areas of the system. During March to May 2004 *Oscillatoria* dominated the cyanobacteria. This occurrence of *Oscillatoria* was not associated with TM presence.

In the Hartbeespoort Dam (Fig. 4.7) the TM concentration were in excess of 29000ug/L in the spring of 2003 during a bloom of *Microcystis* that persisted throughout the winter of 2003. TM concentrations as high as this should be considered a serious health risk to recreational users and may cause breakthrough of the toxins in water treatment works that extract their raw water from such a hypertrophic water source. The presence of TM in a system like the Hartbeespoort Reservoir is not seasonally constant but is primarily associated with *Microcystis*. The extent of TM is not in direct correlation to the extent of the cyanobacterial biovolume (Fig 4.7). The TM concentration cannot be correlated with the Chl *a* concentrations in the resource (Van Ginkel *et*

al. 2006), as other phytoplankton species can be present and dominant at extremely high Chl *a* concentrations, as was found in a number of the studied sites.

The unpredictability of the extent of TM, necessitate one to assume that when *Microcystis* is present in the Hartbeespoort Dam, there are toxins present. For safety reasons it is recommended that the necessary management actions should be in place to ensure the safety of all users.

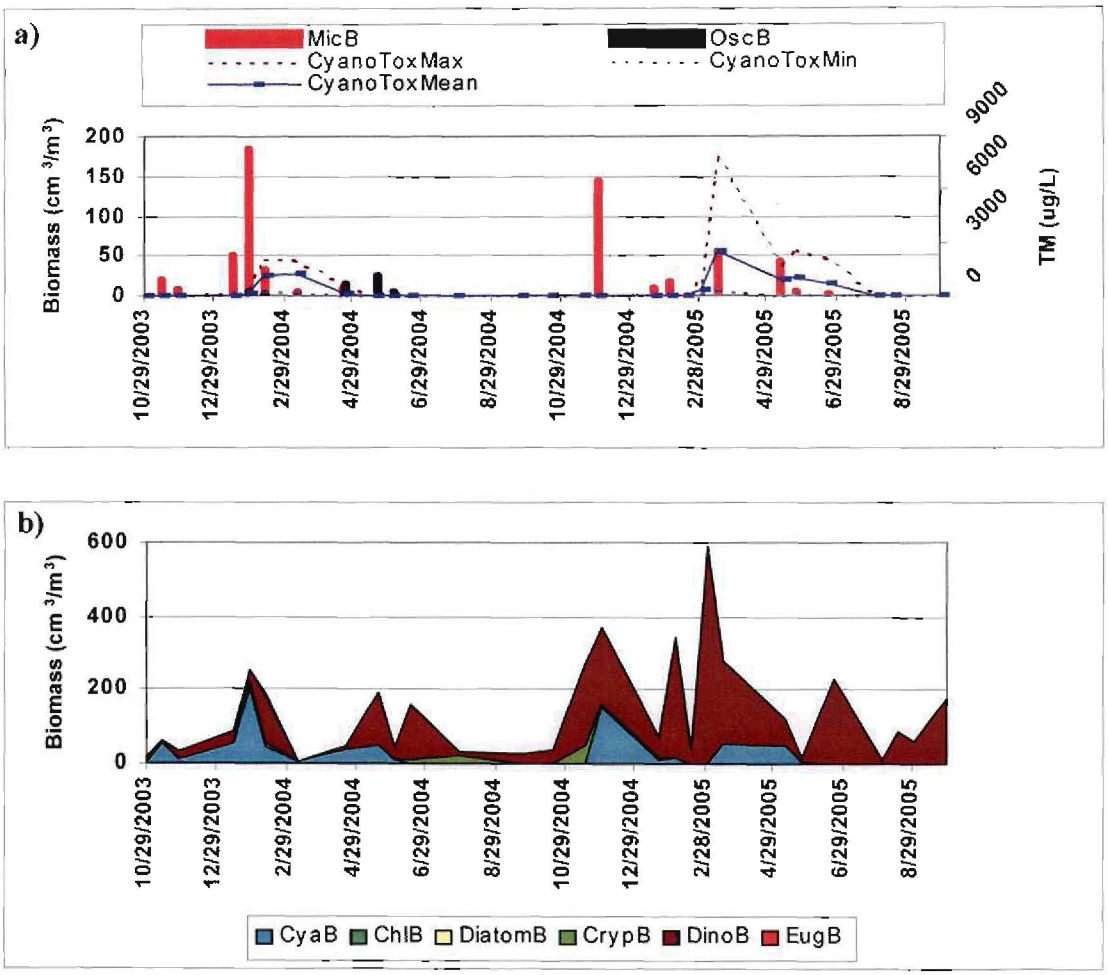


Figure 4.6 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Bon Accord Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB and *Oscillatoria* biovolume: OscB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB.

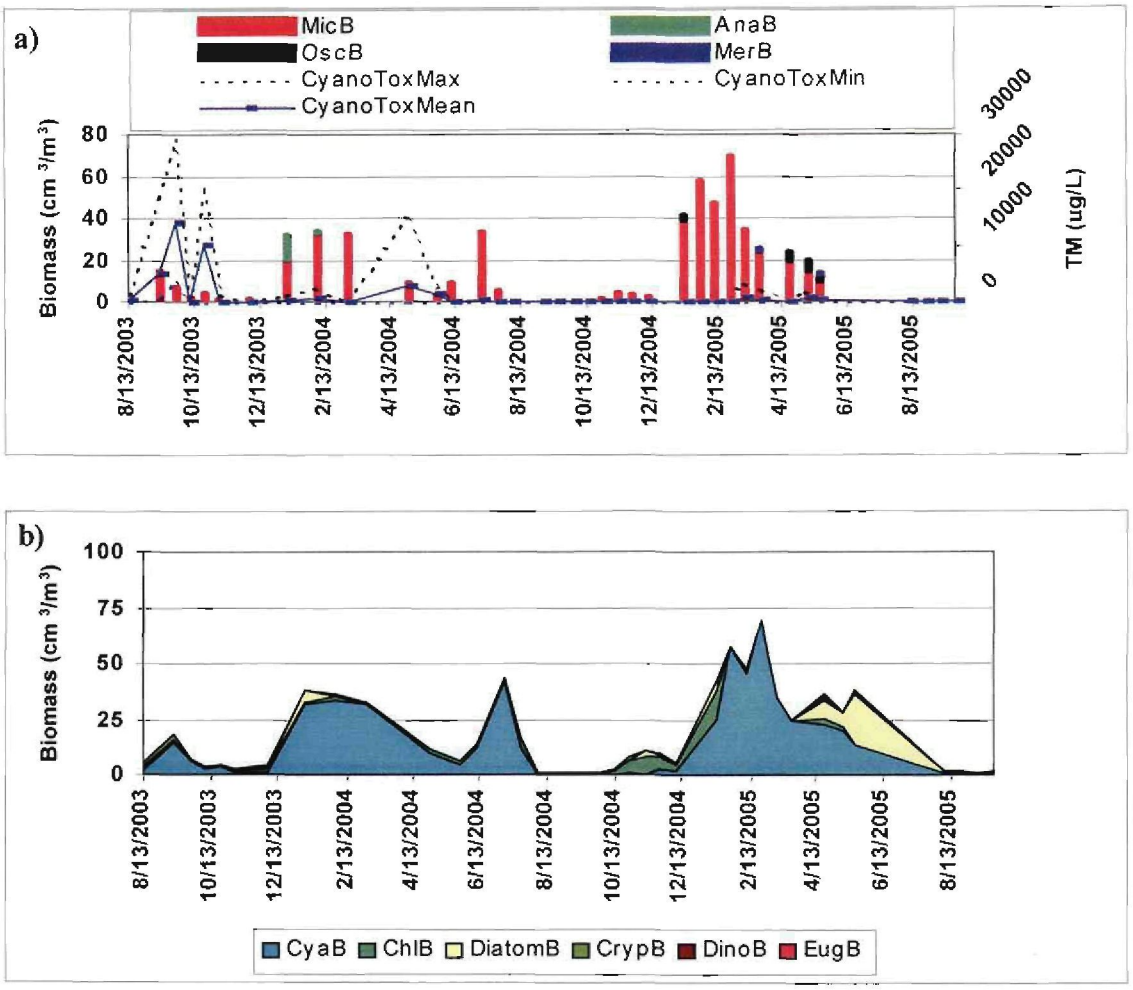


Figure 4.7 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Hartbeespoort Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB and *Merismopedia* biovolume: MerB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB.

In the Klipvoor Dam (Fig. 4.8) the highest TM concentrations was associated with a *Microcystis* bloom, with *Oscillatoria* also present, during January 2003 to March 2003 when TM concentrations in excess of 6000ug/L were found. No cyanobacterial bloom was established in the Klipvoor Dam during the summer of 2004, with the expected lack of TM in the system. During the summer to autumn period of 2005, TM concentrations of less than 1000ug/L were

found in the Klipvoor Dam. These TM concentrations were found in association with *Microcystis*, *Anabaena* and *Oscillatoria*.

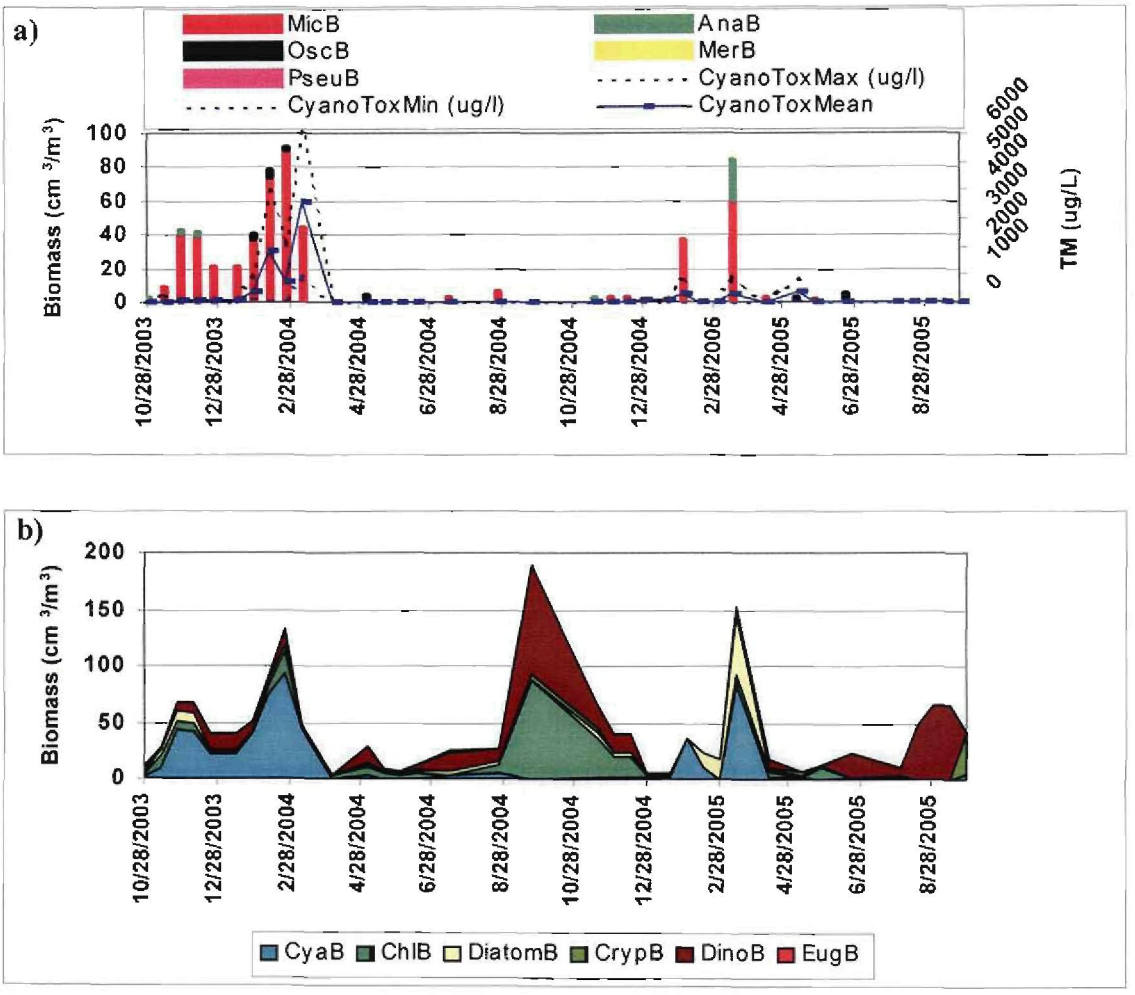


Figure 4.8 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Klipvoor Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, *Merismopedia* biovolume: MerB and *Pseudoanabaena* biovolume: PseuB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB.

In the Rietvlei Dam (Fig. 4.9) maximum TM concentrations in excess of 16000ug/L were found in the summer of 2004. During the summer to autumn of 2005, high TM concentrations

exceeding 4000ug/L were found in the system. As found in the Bon Accord Dam the TM peaks are found subsequent after blooms of primarily *Microcystis*, but *Anabaena* and *Oscillatoria* are also present in noticeable volumes during the summer to autumn of 2005.

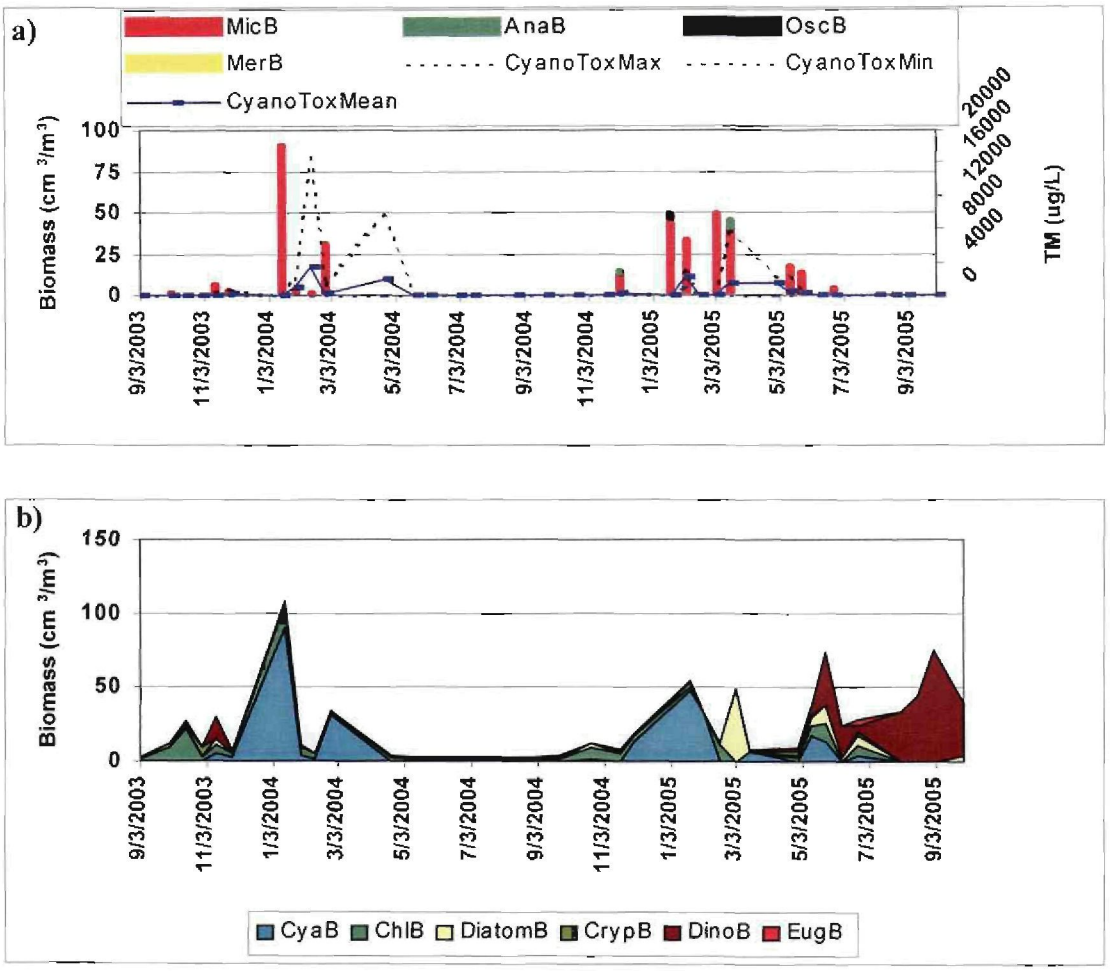


Figure 4.9 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Rietvlei Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, and *Merismopedia* biovolume: MerB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB.

In the Roodeplaat Dam the TM concentrations are much lower than in the other Reservoirs (Fig. 4.10). The TM concentrations seldom exceeded 500ug/L. This is still much higher than the

recreational guideline of 20 ug/L for microcystin-LR, as recommended by the WHO (1999). *Microcystis* is again the species most dominant when TM concentrations are high. During 2005 when *Anabaena* is quite evident, very low or negligibly small concentrations of TM are found in the Roodeplaat Dam. Roodeplaat Dam is the first Reservoir that has manifested growth of *Pseudoanabaena* in association with *Microcystis* and on one occasion with the production of TM in the system.

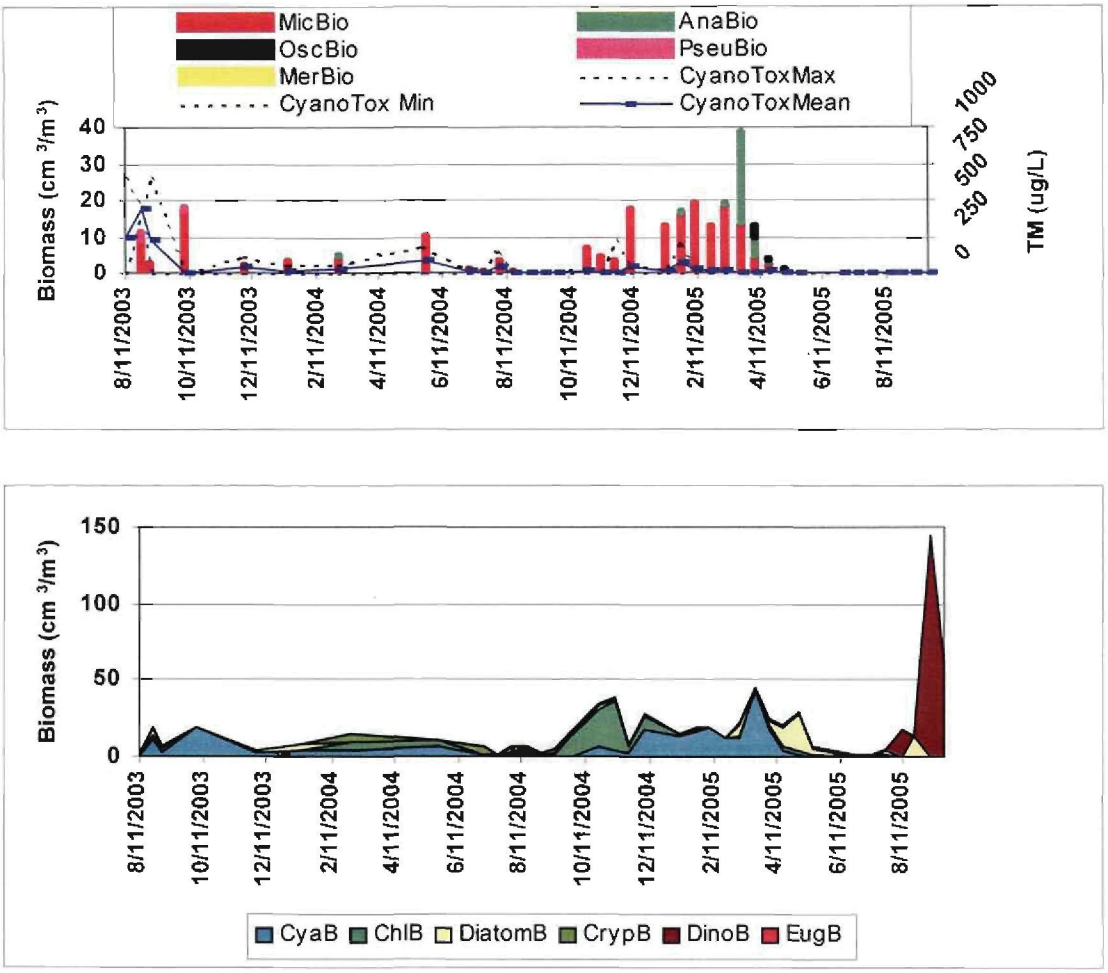


Figure 4.10 a) Total microcystin (TM) concentration (ug/L) and the dominant cyanobacterial species biovolume (cm^3/m^3) present in the Roodeplaat Dam from October 2003 to September 2005. *Microcystis* biovolume: MicB, *Anabaena* biovolume: AnaB, *Oscillatoria* biovolume: OscB, *Merismopedia* biovolume: MerB and *Pseudoanabaena* biovolume: PseuB. b) Phytoplankton community composition. Cyanophyta biovolume: CyaB, chlorophyta biovolume: ChlB, diatom biovolume: DiatomB, cryptophyta biovolume: CrypB, dinoflagellate biovolume: DinoB and Euglenophyta biovolume: EugB.

The system changed to dominance by diatoms and dinoflagellates during 2005 after which no TM was found in the Roodeplaat Dam.

In all five Reservoirs the TM concentrations are primarily found from January and it can last until June, depending on the extent of the cyanobacterial biovolume and bloom period. Only in the Roodeplaat did the TM concentrations persist throughout the year (August 2003 to October 2004).

4.2.4 Temporal and Spatial Depth Distribution of Total Microcystins

At a number of reservoirs in South Africa, water abstraction at different depths is available, as is the case in Roodeplaat. This enables the users to abstract water of best quality for use or for treatment. This practice of water extraction from different depths necessitate information on the depth distribution of cyanobacterial species and total microcystins in the event of a cyanobacterial bloom that do produces cyanobacterial toxins. The dominance of *Microcystis* in Hartbeespoort and Roodeplaat led to the implementation of a monitoring programme at the regular sampling site to determine the occurrence and incidence of TM concentrations over depth in the water column.

The temporal and spatial depth distribution of Chl *a* (Fig. 4.11b), and cyanobacterial biovolume (Fig. 4.11c&d), are compared to the TM concentrations (Fig. 4.11a) in the Hartbeespoort Dam. In Fig. 4.11a, it is shown that TM is produced annually in the Hartbeespoort Dam and that it can reach concentrations in excess of 400ug/L in the water column. TM are mixed through the water column and even on occasion down to a depth of 20m, although the main biovolume of *Microcystis* is, when present, only in the upper 10m. This indicate that TM are mixed into the whole system. This phenomenon may also be explained by the possibility that the sampling technique does not include all the cyanobacteria present.

Anabaena is present for a very short period in January 2004 and may well contribute to the TM concentrations in the Hartbeespoort Dam. The cyanobacterial genera that show dominance during periods of high TM present, are *Microcystis*.

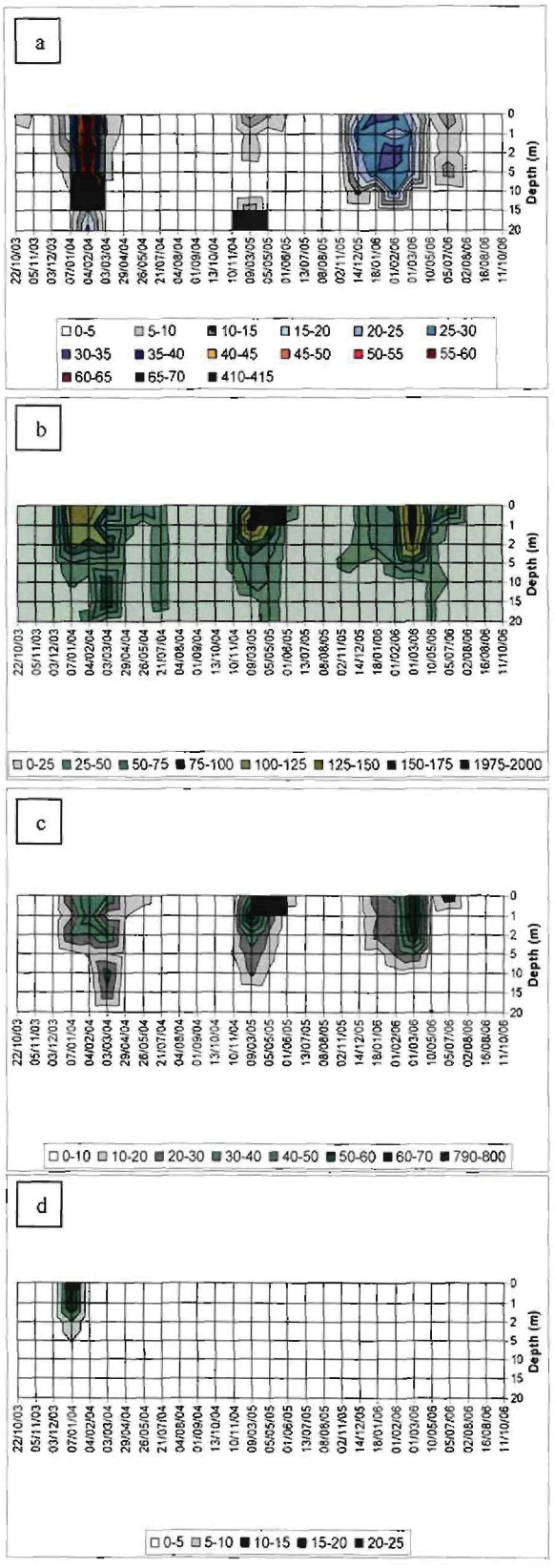


Figure 4.11 The temporal and spatial distribution of a) Total Microcystin concentration (ug/L), b) Chl *a* concentration (ug/L), c) *Microcystis* biovolume (cm^3/m^3) and d) *Anabaena* biovolume (cm^3/m^3) in the Hartbeespoort Dam. Each legend indicates the ranges of concentrations/biovolume of each variable.

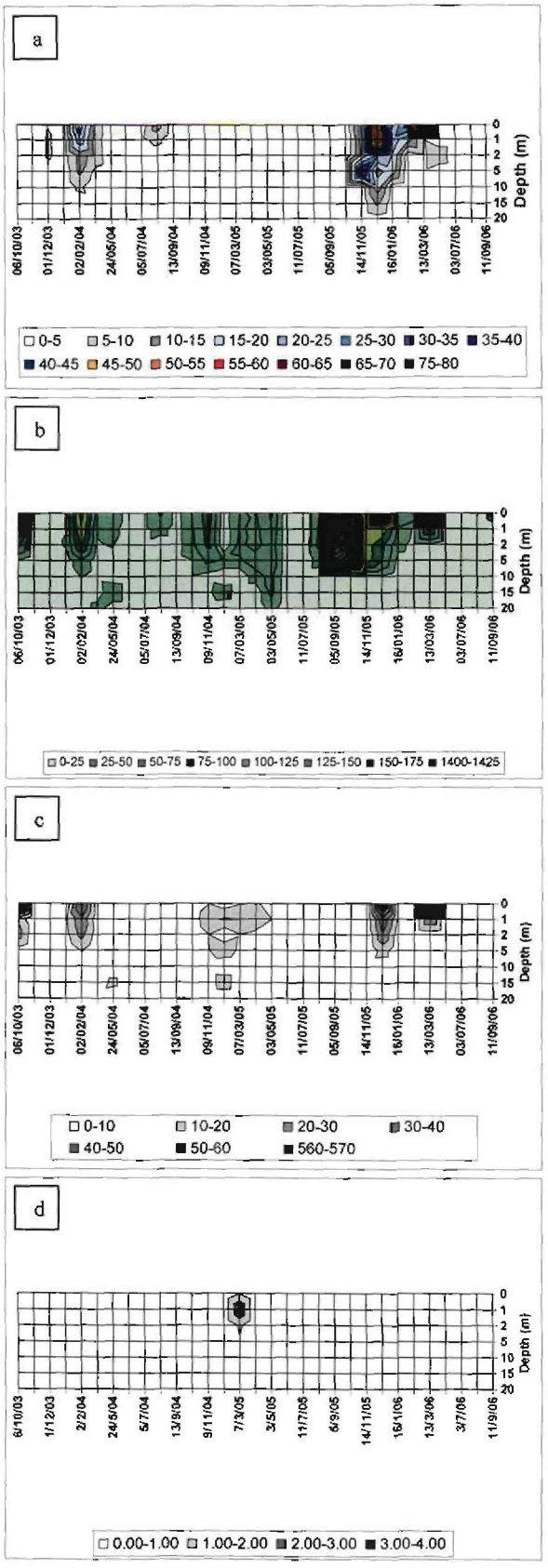


Figure 4.12 The temporal and spatial distribution of a) Total Microcystin concentration (ug/L), b) Chl *a* concentration (ug/L), c) *Microcystis* biovolume (cm^3/m^3) and d) *Anabaena* biovolume (cm^3/m^3) in the Roodeplaat Dam. Each legend indicates the ranges of concentrations/biovolume of each variable.

Similarly, in the Roodeplaat Dam, the temporal and spatial distribution of Chl *a* (Fig. 4.12b), and cyanobacterial genera information, (Fig. 4.12c&d) are compared to the TM concentrations (Fig. 4.12a) in the water column of the Roodeplaat Dam. In the Roodeplaat Dam the TM concentrations in the water column do not exceed 80 ug/L. This can be explained by the much smaller surface area of the Roodeplaat Dam as compared to the Hartbeespoort Dam that will lead to smaller cyanobacterial biovolume with subsequent smaller amount of Tm formation in the Roodeplaat Dam. These events are again associated with the presence and dominance of primarily *Microcystis* as shown in Fig. 4.12.

Both Fig. 4.11 and Fig.4.12 show that the toxins are present on an annual basis and are related to the extent of the *Microcystis* bloom. Users of the raw water for drinking, recreational or agricultural uses can accept that TM is present in high concentrations during extensive *Microcystis* blooms. The necessary precautions need to be implemented as soon as *Microcystis* become the dominant genera in large numbers during the summer periods. An increase in the extraction depth, at least to below 10m, will improve the amount of TM in the water that need to be treated for potable purposes. If a treatment works are not properly equipped to dose with activated carbon, it may be sufficient to extract the water for the Water Treatment Works at a deeper depth.

4.2.5 Multivariate Analysis

To investigate variation of communities across a range of different environmental conditions, huge differences in species composition of the studied communities can be found. However, there is also a certain consistency or predictability in the variations. The gradual change between sites and their communities are often related to differing, but partially overlapping demands of the individual species for environmental factors, such as temperature ranges and other nutrient conditions. The axes along which the communities change in nature are not always visible. A wide range of statistical methods can summarize the variation in the phytoplankton communities, but if we want to stress the continuity of community composition changes, the ordination tools are very useful (Lepš and Šmilauer 2003). CANOCO as one of these tools was used for assessing the collected data.

The ordination diagram approximates the matrix of distance and angles between different species and environmental variables. It is a conclusion from correlations and dissimilarities between

response variables or species and the environmental variables. It is a visual tool to determine the relationships between the species and the environmental variables, the correlations between different environmental variables, etc. (Lepš and Šmilauer 2003).

To determine if there are any differences between the five sampling sites all the results were run through principal component analysis (PCA) in CANOCO (Fig. 4.13). The cluster of samples around the centre of the ordination diagram show that there are in general no distinct differences between the five sites and that the combined results of these five sites can later potentially be used in further analyses assessments or for predictive purposes as one set of data. The low eigenvalues (EV) support this (Fig. 4.13).

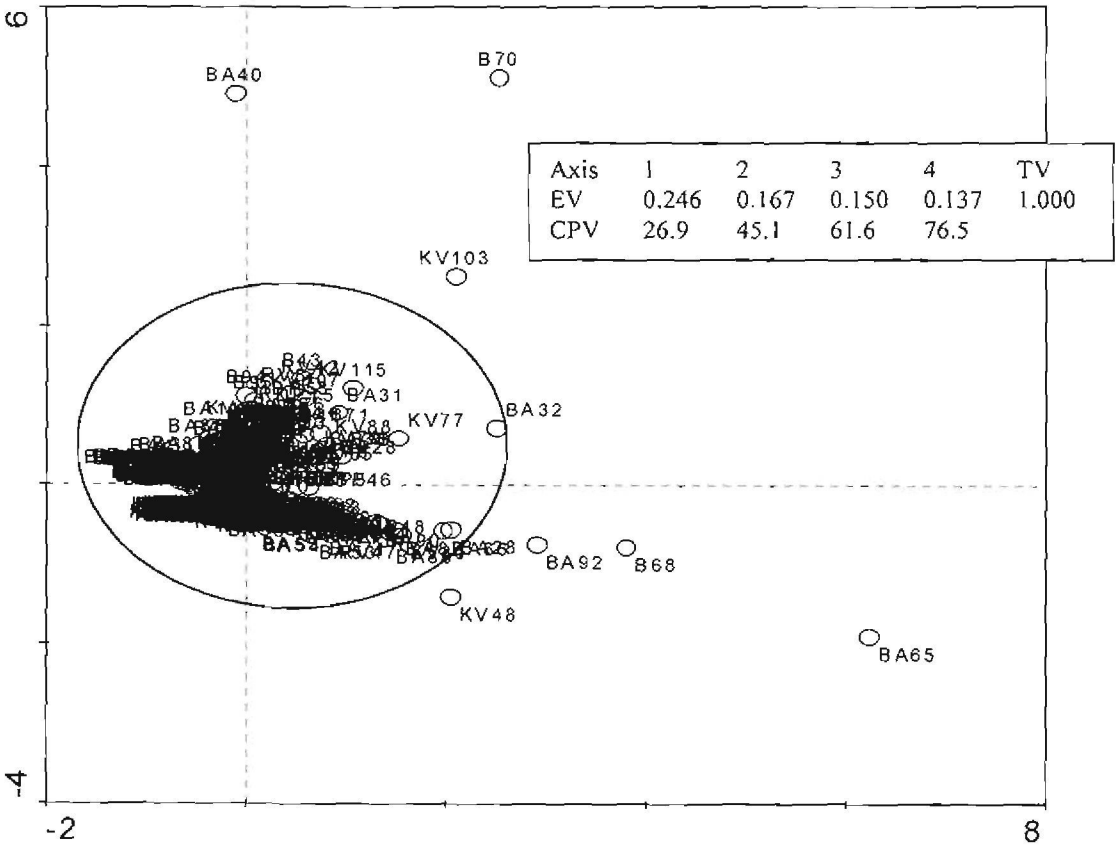


Figure 4.13 The PCA ordination diagram, showing the ordination of all the data from the five reservoirs. The sum of all the EV = 0.916. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat).

Then the data from all the dams were processed using CANOCO to determine if there were any differences between the dams in terms of the algal community composition using the PCA in

CANOCO (Fig. 4.14). A PCA was also performed to determine differences between the various dams with regard to the physico-chemical data collected (Fig. 4.15). With regard to both the community and physico-chemical data, a PCA was unable to group the dams as distinct groupings indicating that the dams in essence behave the same. This simplified the steps for later use of the data in both the training of the neural network modelling as well as the training of the hybrid evolutionary algorithm modelling for this project and in essence enlarged the input data pool for the modelling steps.

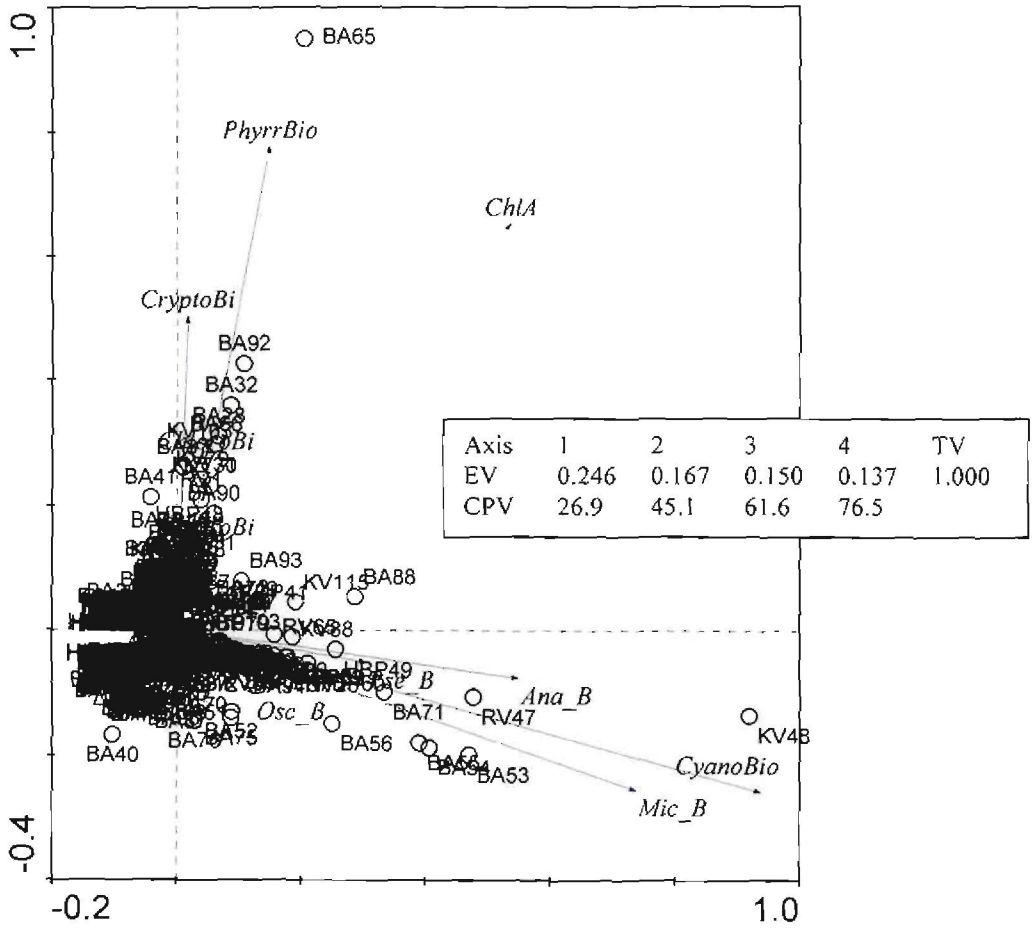


Figure 4.14 The PCA ordination diagram of algal community data in relation to the ordination of all the dam's' data. The sum of all the EV = 0.916. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat & *Ana_B* = *Anabaena* biovolume; *ChlA* = Chl *a*; *CryptoBi* = cryptophyte biovolume; *CyanoBio* = cyanophyte biovolume; *Mic_B* = *Microcystis* biovolume; *PhyrrBio* = Phyrrophyta biovolume (dinoflagellates); *Pse_B* = *Pseudoanabaena* biovolume)

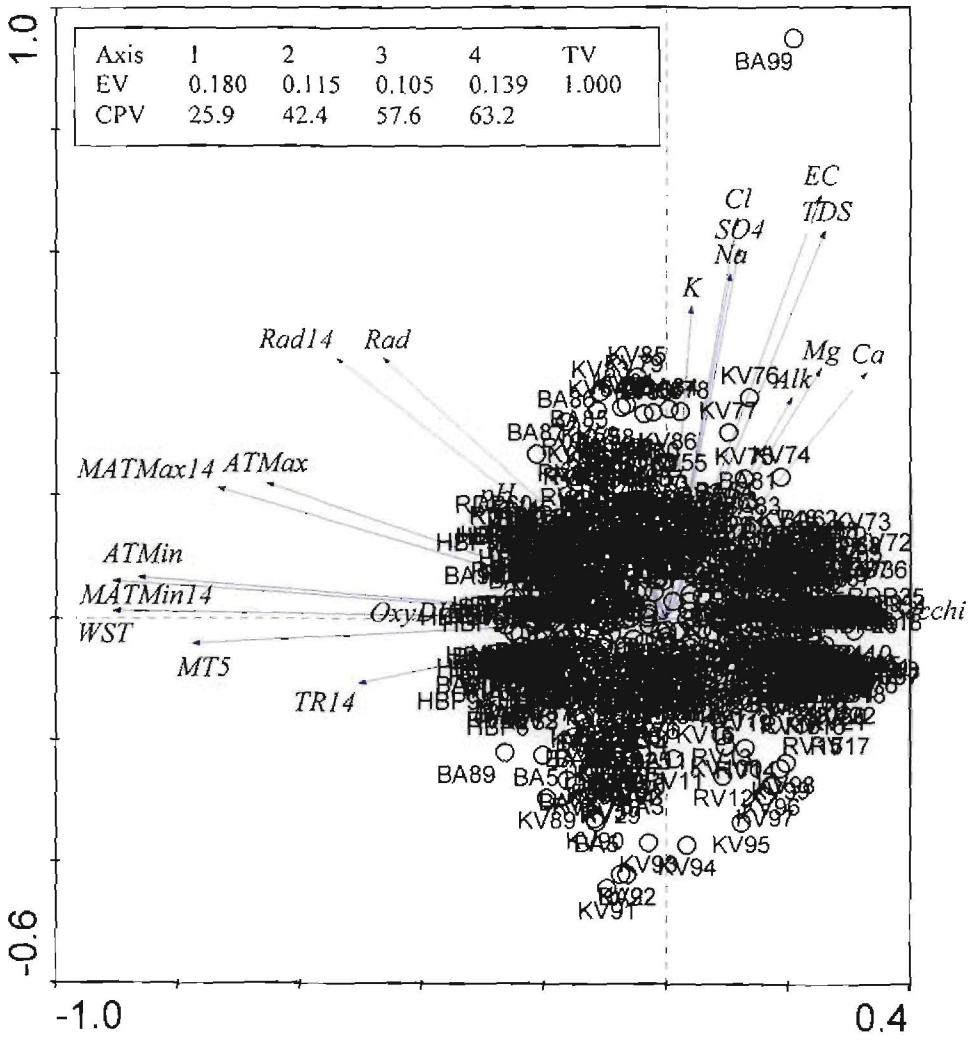


Figure 4.15 The PCA ordination diagram of environmental data (physico-chemical data) in relation to the ordination of all the dam's' data. The sum of all the EV = 0.695. (BA = Bon Accord; HBP = Hartbeespoort; KV = Klipvoor; RV = Rietvlei; RDP = Roodeplaat).

It was now established that there is no differences between the different dams and that the data can be combined in the modelling exercises that followed. There was, however, a need to determine which variables are redundant so that only the most important environmental data variables are used in the further modelling exercises. This was done by using the Redundancy Analysis (RDA) component of CANOCO.

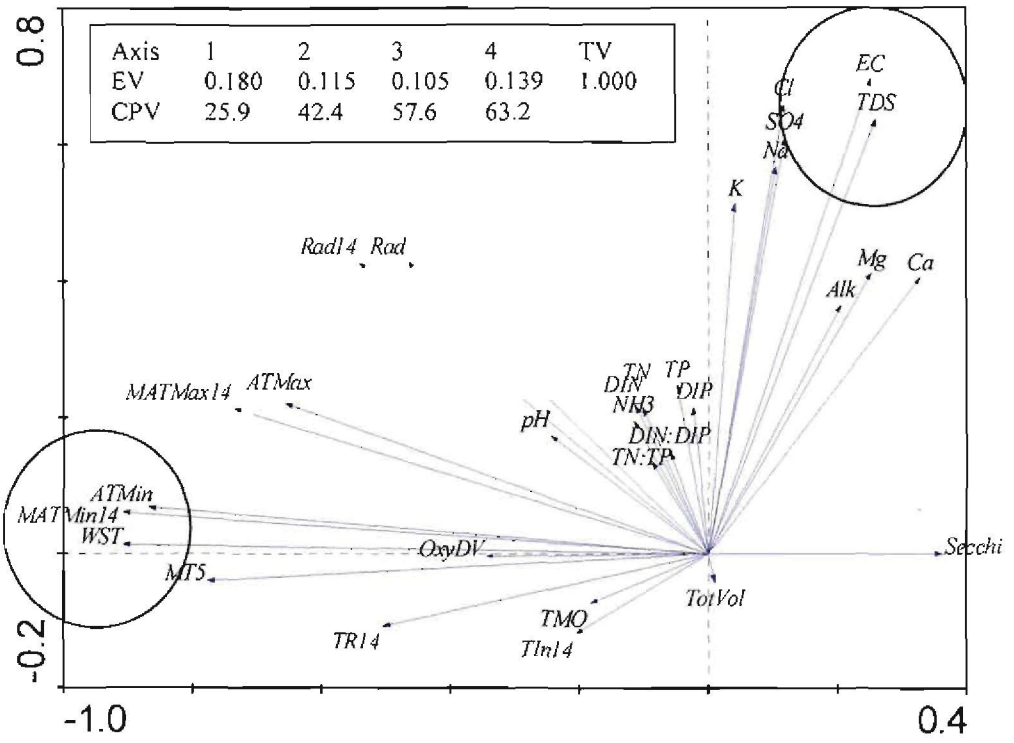


Figure 4.16 The RDA ordination diagram, showing environmental variables and their relationships to each other. The sum of all the EV = 0.695.

All the variables measured were first plotted on an RDA ordination diagram to determine which environmental variables are redundant and can be excluded from further analysis. All the salinity-associated variables could be replaced by one or two environmental variables and it was decided to keep only the EC and discard TDS. This may exclude the relationships of specific algal species to specific anions or cations, however all the cations and anions where relatively closely related to the EC and the TDS. But, if there is a relationship between a specific algal species and more saline conditions, it should be shown by correlations to both EC or TDS (Fig. 4.16) as the main anions and cations are closely related to EC and TDS. In future further research regarding specific anion or cations can then be undertaken.

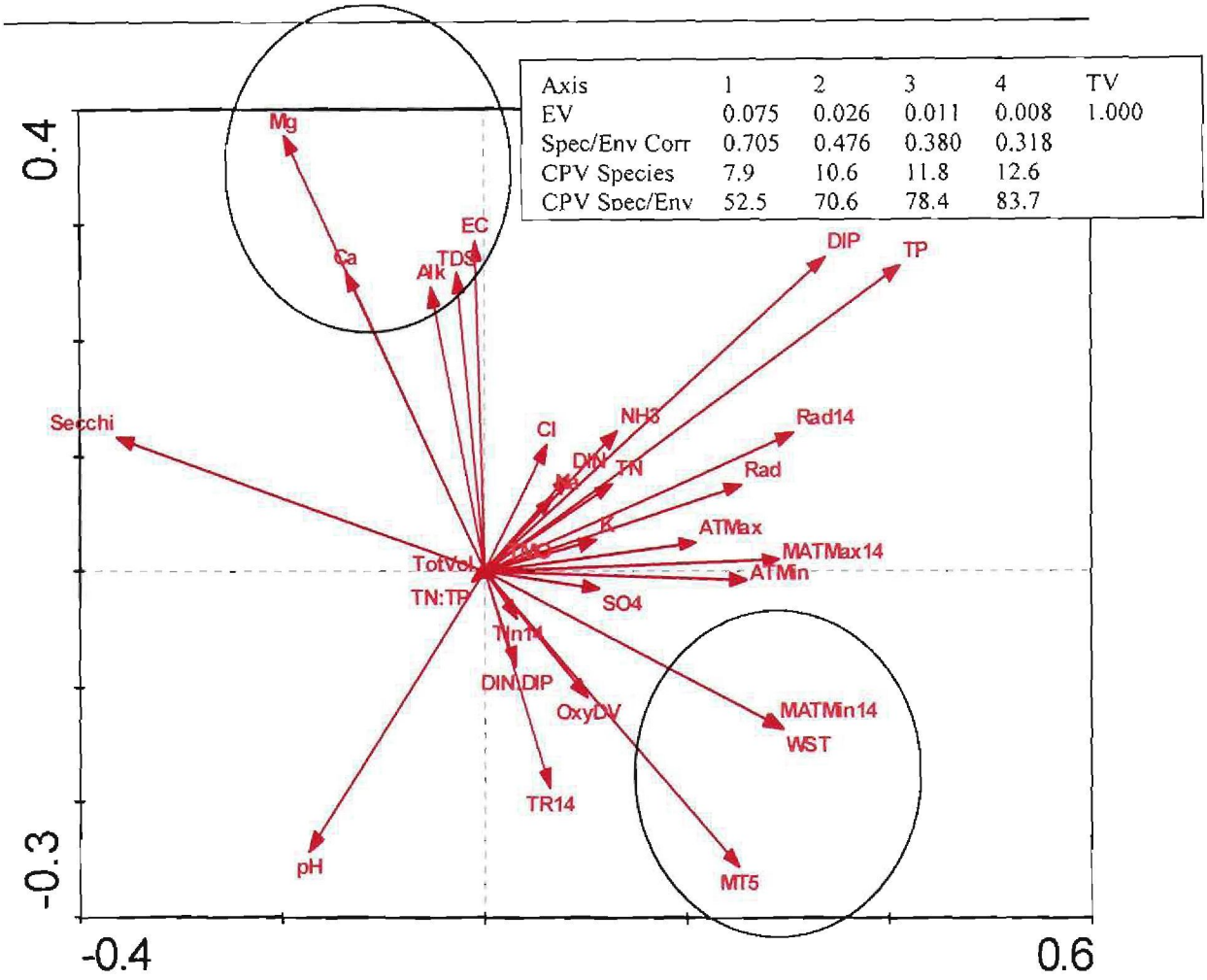


Figure 4.17 The RDA ordination diagram, showing environmental variables and their relationships to one another. Test of significance for 1st axis: EV = 0.075; F-ratio = 43.479; P-value = 0.0020. Test of significance of all canonical axis: Trace = 0.142; F-ratio = 2.897; P-value = 0.0020.

Similarly, there are many similarities between the different measures of temperature (Mean min 14 days preceding sampling – MATMin14; water surface temperature – WST; mean temperature of the 0-5m depth) (Fig. 4.17). The environmental variable with the longest arrowed line, namely the mean minimum temperatures of the 14 days MATMin14) or the water surface temperature (WST) that preceded sampling had the greatest effect on the response variables and was so close together on the diagram that anyone could be used to represent all the temperature related variables. By doing a number of RDA and PCA analyses the redundant environmental variables and principal components are highlighted. This enable one to determine the most important variables that can be used in forecasting or prediction of the response variables.

The CANOCO redundancy analysis (RDA), including the algal groups and the species abundance as the response variables, and the environmental conditions as the environmental variables, showed that the best correlations in the development of the Cyanobacteria biovolume (*CyanoBio*), the *Microcystis* biovolume (*Mic_B*), and Chl *a* concentrations (*ChlA*), were with water surface temperature (WST) and the minimum temperatures 14-days preceding sampling (MATMin14) (Fig. 4.18).

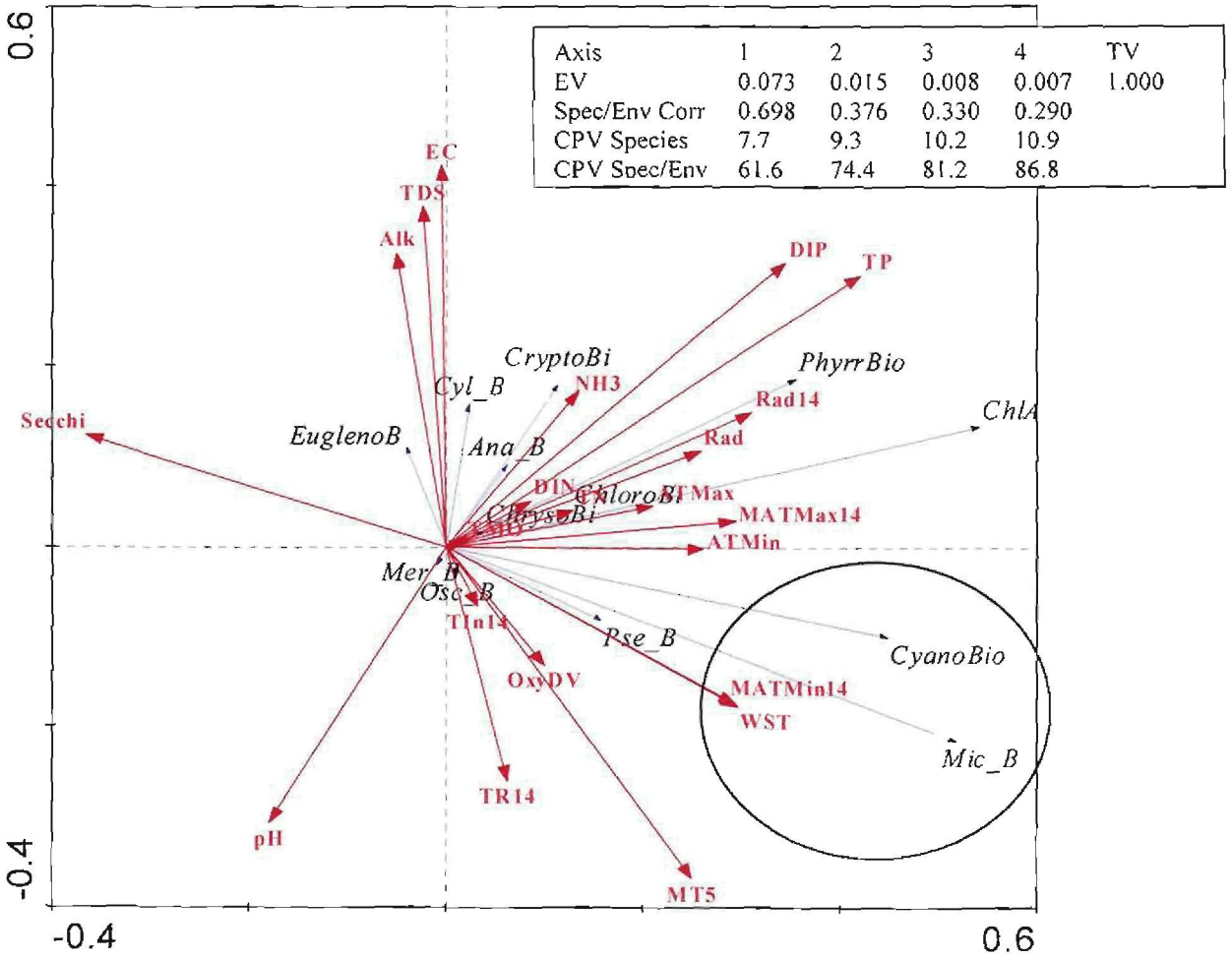


Figure 4.18 The RDA ordination diagram, with standardized data and with the algal groups and species biovolume, as response variables set, and the environmental conditions as environmental variables. Test of significance for 1st canonical axis: EV = 0.073; F-ratio = 43.226; P-value = 0.0020. Test of significance of all canonical axxes: Trace = 0.119; F-ratio = 3.363; P-value = 0.0020.

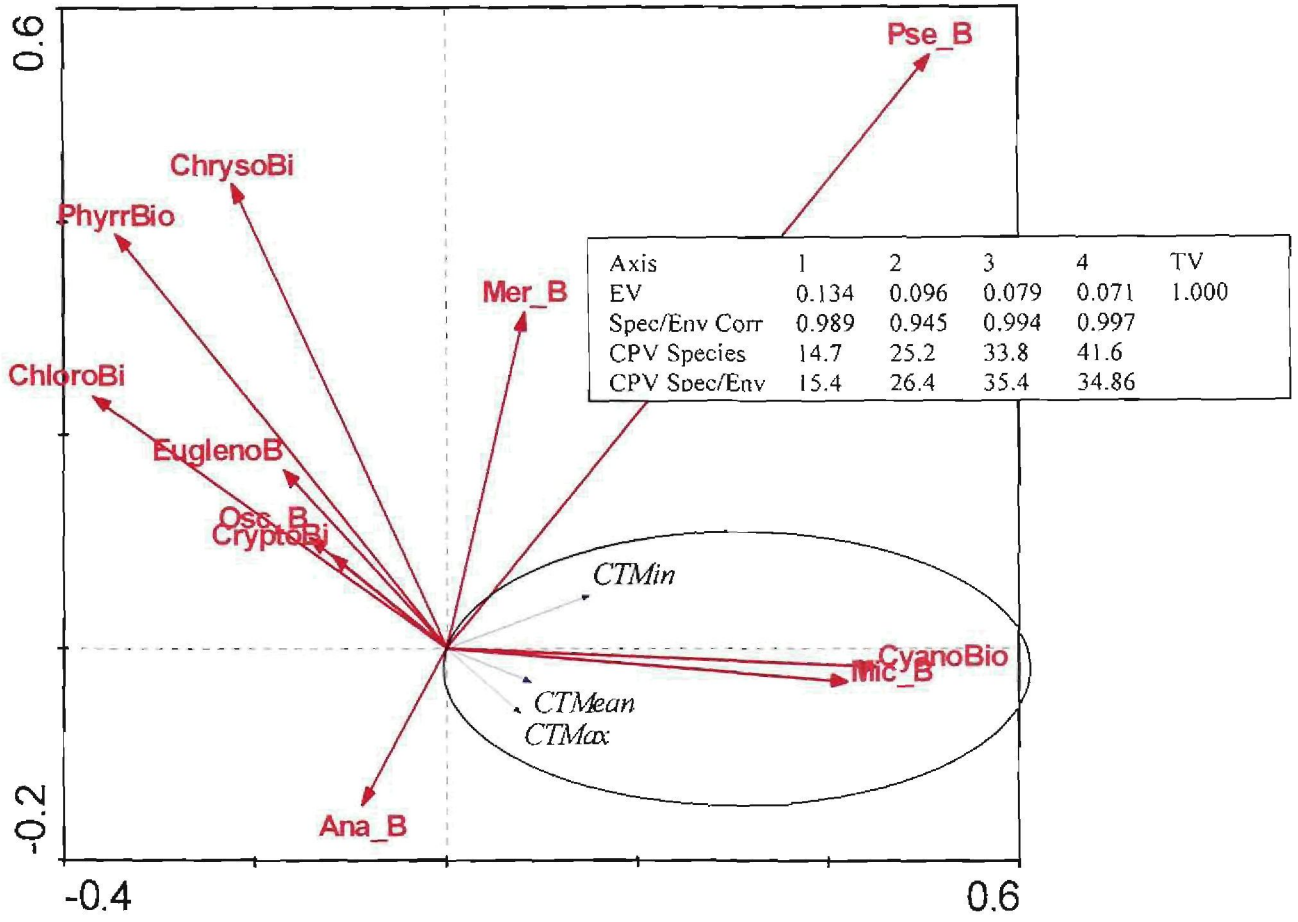


Figure 4.19 The ordination diagram of PCA with non-standardized data with the algal groups and species biovolume as environmental variables and the cyanotoxin concentrations as the species. Test of significance for 1st canonical axis: EV = 0.004; F-ratio = 0.551; P-value = 0.7800. Test of significance of all canonical axxes: Trace = 0.004; F-ratio = 0.192; P-value = 0.8200.

The phytoplankton species that show the best correlation with the TM's are the cyanobacterial biovolume and more specifically *Microcystis* biovolume (Fig. 4.19). When the environmental variables are considered (Fig. 4.20) the best positive correlations with the presence of TM's (CTMax; CTMin & CTMean) as the response variable is dissolved inorganic phosphorus (DIP) (Fig. 4.20) and the best negative correlation with the TM's are the DIN:DIP ratio. These variables can now be used to apply to forecasting of the response variables.

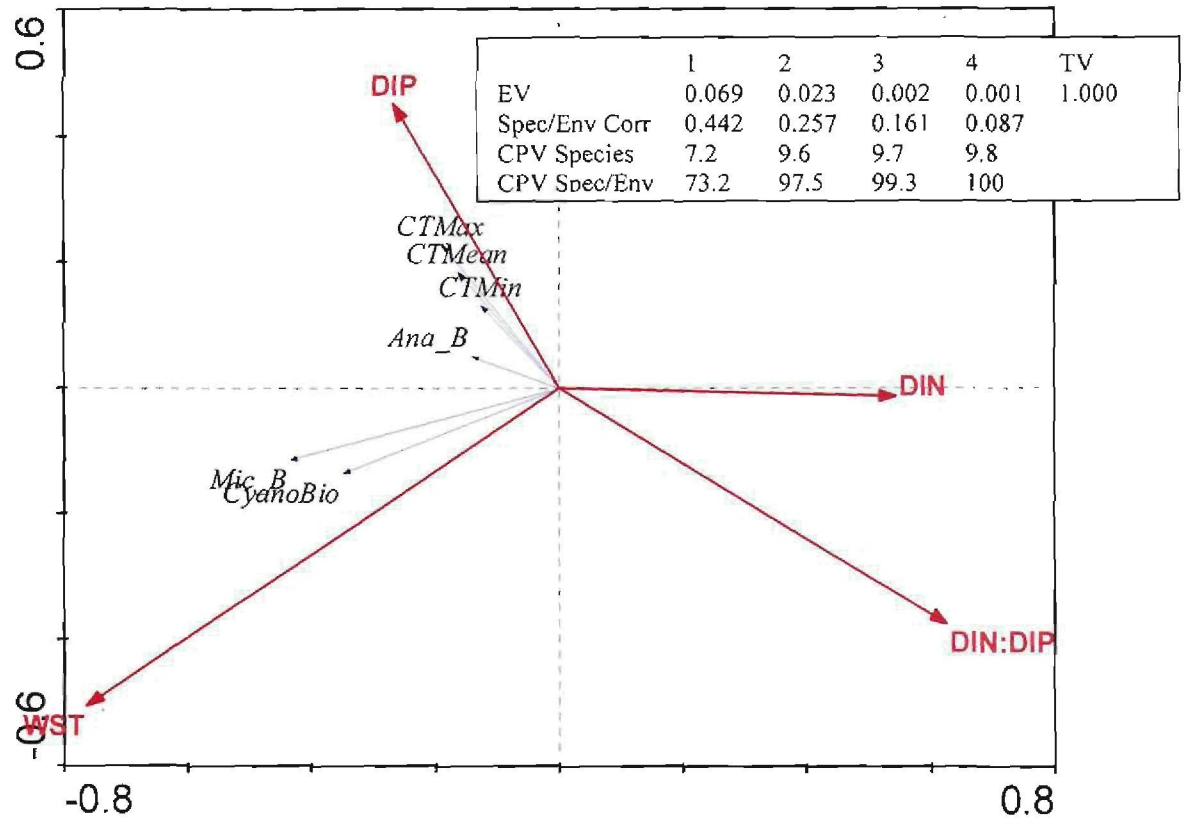


Figure 4.20 The RDA ordination diagram showing the reduced environmental variables and the cyanotoxin concentrations, cyanobacterial biovolume and species biovolume as the response variables. Test of significance for 1st canonical axis: EV = 0.069; F-ratio = 10.432; P-value = 0.0180. Test of significance of all canonical axxes: Trace = 0.094; F-ratio = 3.667; P-value = 0.0140.

The results shown in Fig. 4.20 indicate that the most important environmental variable for forecasting of the cyanobacteria biovolume and specifically the *Microcystis* biovolume is the water surface temperature (WST). The close correlations of the WST, the minimum air temperature (ATMin) and the minimum temperatures 14 days preceding sampling (MATMin14) (Fig. 4.16) show that if any of these variables are available, the other can be determined by a simple linear equation.

A Redundancy Analysis (RDA) was performed using all the community data as species data and the physico-chemical data as environmental data for CANOCO. The RDA assisted in elucidating

relationships between the entities. The RDA was also used to determine the relative importance of the various gradients. Using the RDA triplots the list of contributing variables was reduced in a stepwise manner. The RDA was also unable to discern distinct groupings in the various sample sites (Figure 4.21).

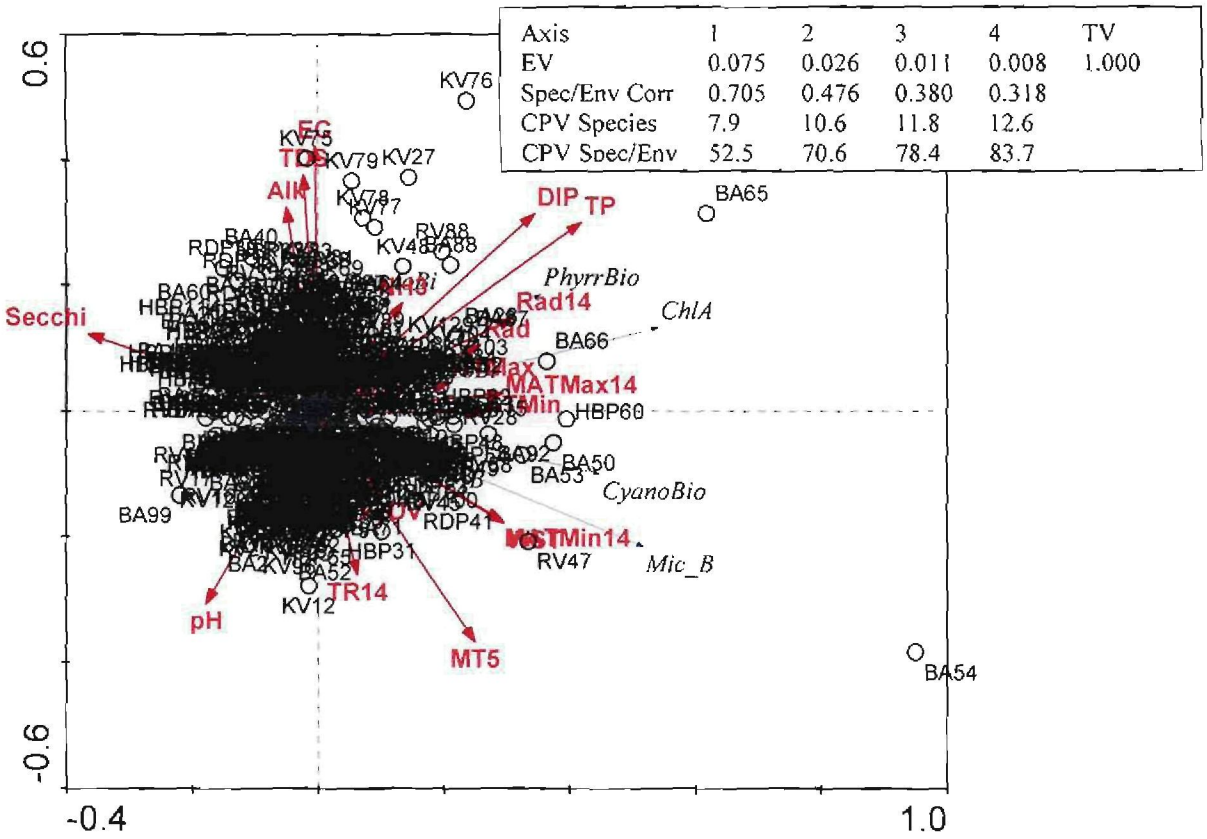


Figure 4.21 The RDA ordination diagram of species and environmental data. All environmental data collected from the five reservoirs during the study period, was used. Test of significance for 1st canonical axis: EV = 0.075; F-ratio = 43.479; P-value = 0.0020. Test of significance of all canonical axxes: Trace = 0.142; F-ratio = 2.897; P-value = 0.0020.

By eliminating relationships between environmental variables and using literature as guidelines the final set of environmental variables selected for further modelling with the artificial neural network was: water surface temperature (WST), dissolved inorganic phosphorous (DIP), dissolved inorganic nitrogen (DIN), the DIN:DIP ratio; and Chl *a* concentration.

For the hybrid evolutionary algorithm modelling TP, DIP, Secchi Depth, pH, TN, DIN, WST, Chl *a* and algae biovolume were used for rule set development and understanding the most important variables in determining the different algae genera biovolume.

4.3 Conclusions

All five study sites were hypertrophic systems that have excess nutrients available. The five sites are situated in the summer rainfall area of South Africa. All the sites are exposed to regular cyanobacterial and increasingly dinoflagellate blooms. The results show that the cyanobacterial blooms in the Crocodile West/Marico Water Management Area are dominated by *Microcystis*, and although *Anabaena* and *Oscillatoria* are found occasionally in large numbers, they do form a negligibly small component of the annual phytoplankton community.

The dinoflagellate blooms are exclusively *Ceratium hirundinella* and these dinoflagellate blooms are associated with extremely high Chl *a* concentrations in the five Reservoirs. *Ceratium* dominated Bon Accord and Klipvoor Dams during the 5-year study period, while Hartbeespoort and Rietvlei experienced *Ceratium* blooms during 2000 and again in 2005. For the first time Roodeplaat Dam experienced a *Ceratium* bloom in 2005. The cause for the changing nature of the phytoplankton composition from previously summer dominance by *Microcystis* and winter dominance by diatoms to *Ceratium* in these five reservoirs, need to be investigated.

The presence of TM's in all the reservoirs was primarily associated with the dominance and blooms of *Microcystis*. The other cyanobacterial species occasionally found in association with the TM's are *Anabaena* and *Oscillatoria*. They are both known TM producers and could contribute to the toxin production in all the reservoirs (WHO 1999). Extremely high toxin concentrations are found in these hypertrophic reservoirs of South Africa. The Department of Water Affairs and Forestry with the assistance of the future CMA's need to develop strategies to manage both the eutrophication levels in the Crocodile-West/Marico Water Management Area, as well as the risk imposed by the toxins, produced by the cyanobacteria in the water resources, on drinking water quality and recreational users health.

The PCA of CANOCO indicated that the reservoirs are similar in both algal community and physico-chemical variables. This establishes that there are no major differences between the reservoirs and that the data can be combined in the modelling exercises that followed. The

multivariate analysis shows that of the environmental factors, temperature (both WST & MATMin14) is the most important factor in the development of cyanobacteria biovolume in these hypertrophic systems. This is, therefore, one of the main driving forces and can be used as a indictor of climate and weather impacts on a system. According to the multivariate analysis, the main driving force that correlates with the TM is the DIP concentrations and DIN:DIP ratio in these systems. This variable will thus play an important role in the prediction of the cyanobacterial toxin production (TM).

Based on the results discussed water surface temperature (WST), dissolved inorganic phosphorous (DIP), dissolved inorganic nitrogen (DIN), the DIN:DIP ratio; total phosphorous (TP), total nitrogen (TN) and Chl *a* concentration were the variables used in the further modelling approaches.

CHAPTER 5

MODELLING TECHNIQUES

In addition to establishing the correlations between algal species, TM's and other environmental variables to establish the most important variables in algal and cyanobacterial bloom formation, it is necessary to investigate modelling techniques that can be used in forecasting and predictions of algal and cyanobacterial blooms for management purposes. This Chapter discusses and evaluates the different modelling techniques used, for the understanding and prediction of eutrophication, and phytoplankton succession of the reservoirs under study, are discussed and evaluated.

5.1 Modelling Methods

5.1.1 Vollenweider Model

The Vollenweider Model is a simple empirical model based on the relationship between the total phosphorus and Chl *a* measured in a system, rather than understanding the biological processes of the system. The Vollenweider model was developed as an eutrophication management tool. It does not provide information on the algal or cyanobacterial species that may pose a problem in a system (Vollenweider 1976).

$$\textbf{Chl } a = 0.28 * TP^{0.98} \qquad (5.1)$$

Where Chl *a* is the predicted mean Chl *a* concentration in mg/m^3 which is equal to ug/L and TP is the mean total phosphorus concentration in the early summer measured in mg/m^3.

$$TP = (L * t) / z * (1 / (1 + \text{sqrt}(z))) \qquad (5.2)$$

Where:

- L is the specific phosphorus load (mg/m^2/year).
- t is the water residence time determined as volume/flow.
- z is the mean water depth measured in meter and determined as volume/surface area.

Case studies were done for the hypertrophic Hartbeespoort, Klipvoor and Roodeplaat Reservoirs looking at different management scenarios as inflow data that is essential for applying the

Vollenweider Model, was available for these three Reservoirs. Bon Accord and Rietvlei Reservoirs were excluded because of a lack of inflow data.

The information needed to do the Vollenweider model is shown in Table 5.1.

Table 5.1: Method to determine the constant parameters included for the Vollenweider Model.

Parameter	Definition	Unit	Calculation
TAI	Total Annual Inflow	m^3/year	Sum of 365 consecutive daily flows
MIP	Mean In stream PO^4-P	mg/m^3	Annual Average PO^4-P concentrations
MAPL	Mean Annual PO^4-P load	mg/year	TAI x MIP
SA	Surface Area	m^2	Known surface area at full supply level
PAL	Phosphorus Area Load	mg/m^2/year	MAPL/SA
V	Volume	m^3	Volume at full supply level
RT	Residence Time	year	V/TAI
MD	Mean depth	m	V/SA
TP	Mean Annual Total phosphorus	ug/l	Annual Mean TP concentrations to determine the effectiveness and applicability of the model
Chl *a*	Mean Annual Chl *a*	ug/l	Annual Mean Chl *a* concentrations to determine the effectiveness and applicability of the model

Three management scenarios were investigated, namely a) TP reduction by 50%, b) TP elimination by 90%, and c) Doubling of the residence time.

5.1.2 SALMO Model

The SALMO model developed by Bendorf and Recknagel (1982), is a process-based deterministic model that uses inflow, PO_4-P and NO_3-N concentrations in the inflow, solar radiation and water temperature to determine phytoplankton biovolume, functional algal groups, zooplankton biovolume and oxygen concentrations in the lake. Although only a few input variables are considered the model considers a great number of internal control mechanisms. The model has been tested on a number of lakes with different trophic states by Bendorf and Recknagel (1982).

Constant parameters determinations are shown in Table 5.2. The number of the first decade of ice cover in South Africa is 1, as no ice cover is found. The first four constant parameters are basically determining the different seasons, while the other constants indicate the start situation in the reservoir regarding biological conditions and nutrient and oxygen concentrations as well as the underwater light conditions.

Table 5.2: Methods to determine the constant parameters included in SALMO.

Parameter	Definition	Unit	Calculation
STW	Number of first decade with ice cover		
VZF	Number of first decade after ice cover		
STS	Number of first decade with thermocline		
VZH	Number of first decade after thermocline		
X1(0), X2(0), X3(0)	Biovolume of cyanobacteria on day 1 Biovolume of diatoms on day 1 Biovolume of green algae on day 1	cm^3/m^3 cm^3/m^3 cm^3/m^3	(Chl *a* x %Cyanobacteria x 0,75) x 3 (Chl *a* x %diatoms x 0,75) x 3 (Chl *a* x %Chlorophyta x 0,75) x 3
Z1(0)	Biovolume of zooplankton on day 1	cm^3/m^3	Not available
D(0)	concentration of detritus on day 1	mg/l	[Suspended Solids]
P(0)	concentration of PO4-P on day 1	mg/l	[PO4-P]
N(0)	concentration of NO3-N on day 1	mg/l	[NO3-N]
O(0)	concentration of diss. oxygen on day 1	mg/l	Average [DO] for 0-5 m depth
LTMAX	Maximum underwater light transmission		$LTMAX = (\varepsilon_{min} - 0.92)/(-0.46)$ $\varepsilon_{min} = 4.6 / (1.7 * z_{secchi})$

SALMO-OO data input start with the winter period and the data for Roodeplaat, Hartbeespoort and Klipvoor dams for the period July 2003 to June 2004 was prepared as 36decade periods as required by SALMO-OO. All the input variables required by SALMO-OO, the units and the method of calculation are shown in Table 5.3. The daily inflows of the Hartbeespoort, Klipvoor and Roodeplaat dams were available and it is essential in the model. It was decided to exclude Bon Accord and Rietvlei from this section as no flow data, which is essential for the SALMO-OO Model, was available.

Tab. 5.3: Methods to determine the input variables included in SALMO.

Variable	Definition	Unit	Calculation
V	Total water volume of the whole lake	million m^3	Calculate V for total water depth of day 1 of each decade by means of the depth to volume relationship
VE	Epilimnion water volume	million m^3	Calculate difference of V-VH
VH	Hypolimnion water volume	million m^3	Calculate VH for hypolimnion depth of day 1 of each decade in summer by means of the depth to volume relationship
ZMIXREAL	Maximum water depth of lake during spring, autumn and winter or Depth of thermocline during summer	m	 To be determined by depth at which maximum temperature difference per meter occurs
ZMIX	Mean mixing depth of total volume or epilimnion	m	Ratio of total volume V to total area during spring, autumn and winter or Ratio of epilimnion volume VE to total area during summer
ZHM	Mean mixing depth of hypolimnion	m	Ratio of hypolimnion volume VH to hypolimnion area during summer
QIN	Mean daily flow of streams entering the lake surface	m^3/day	
QHIN	Portion of the mean flow of streams entering the hypolimnion	m^3/day	
QOUT	Mean daily outflow from the lake surface	m3/day	$QOUT = (\Sigma QIN - (V_{i+1} - V_i))/10$
QHOUT	Mean daily outflow from the hypolimnion	m3/day	
SRF	Factor reflecting strong rain events with implications for limiting underwater light	-	Absolute maximum ratio of flow measurements of consecutive days for each decade: $abs(max (Q_i/Q_{i-1}))$
I	Mean global solar radiation	J/cm^2 /day	
T	Mean water temperature over the lake depth or the epilimnion depth	°C	
TH	Mean water temperature over the hypolimnion depth	°C	
PIN	Weighted mean PO4-P concentration of streams entering the lake	mg/m^3	
NIN	Weighted mean NO3-N concentration of streams entering the lake	g/m^3	
POMIN	Weighted mean detritus concentration of streams entering the lake	g/m^3	In-lake suspended solids concentration

A study by Cetin *et al.* (2005) investigated more process models for algal growth and grazing to be implemented in SALMO-OO to determine the applicability of such libraries of different growth and grazing equations on the modelling of lake ecosystems of different trophic status and

climate. The original SALMO-OO model was used in this study. The growth and grazing equations of Arhonditsis and Brett (2005), a model developed for the eutrophic Lake Washington (USA), was used additionally to improve the results. The main differences in the two models are the maximum photosynthesis rate per day, the half-saturation constant for P uptake by algae, phytoplankton settling velocity, optimum temperature for phytoplankton growth and the preference factor for each algal functional group.

5.1.3 Artificial Neural Networks (ANN)

Neural networks are simple computational tools for examining data and developing models that help to identify interesting patterns or structures in the data. The data used to develop these models is known as training data. Once a neural network has been exposed to the training data, and has learnt the patterns that exist in the data, it can be applied to new data (the testing data) thereby achieving a variety of outcomes (Smith 2002).

The method of ANN has been inspired by the biological nervous system, hence the name of the modelling technique. One of the most significant advantages of ANN is their ability to learn from a limited set of data. A well-trained and verified ANN for the specific problem at hand recognises the data and make predictions with often desirable accuracy (Karul and Soyupak 2006). However, one has to be wary of the pitfalls of neural networks, e.g. overtraining of a limited data set. Where too much training occurs, the network only memorizes the training set and loses its ability to generalize to new data (Mendelshon 1993).

Two known types of ANN, a) the multilayered feed forward neural network (MFNN) and b) the self-organising map (SOM) were tested and are discussed as part of the study.

5.1.3.1 Multilayered Feed Forward Neural Network (MFNN)

The multilayered feed forward neural network (MFNN), is an example of a neural network trained with supervised learning (Rumelhart & McClelland 1986). With the models that have supervised learning, the data used for training contains the complete information about the characteristics of the data and the observable outcome. The study aims at developing a model that can learn the relationship between the inputs (the environmental variables) and the outputs (the algal groups). The MFNN is trained repeatedly with numerical data of the inputs and the outputs

in order to develop a model by learning to represent the relationships correctly (Figure 5.1). The method includes a three layer feed forward neural network.

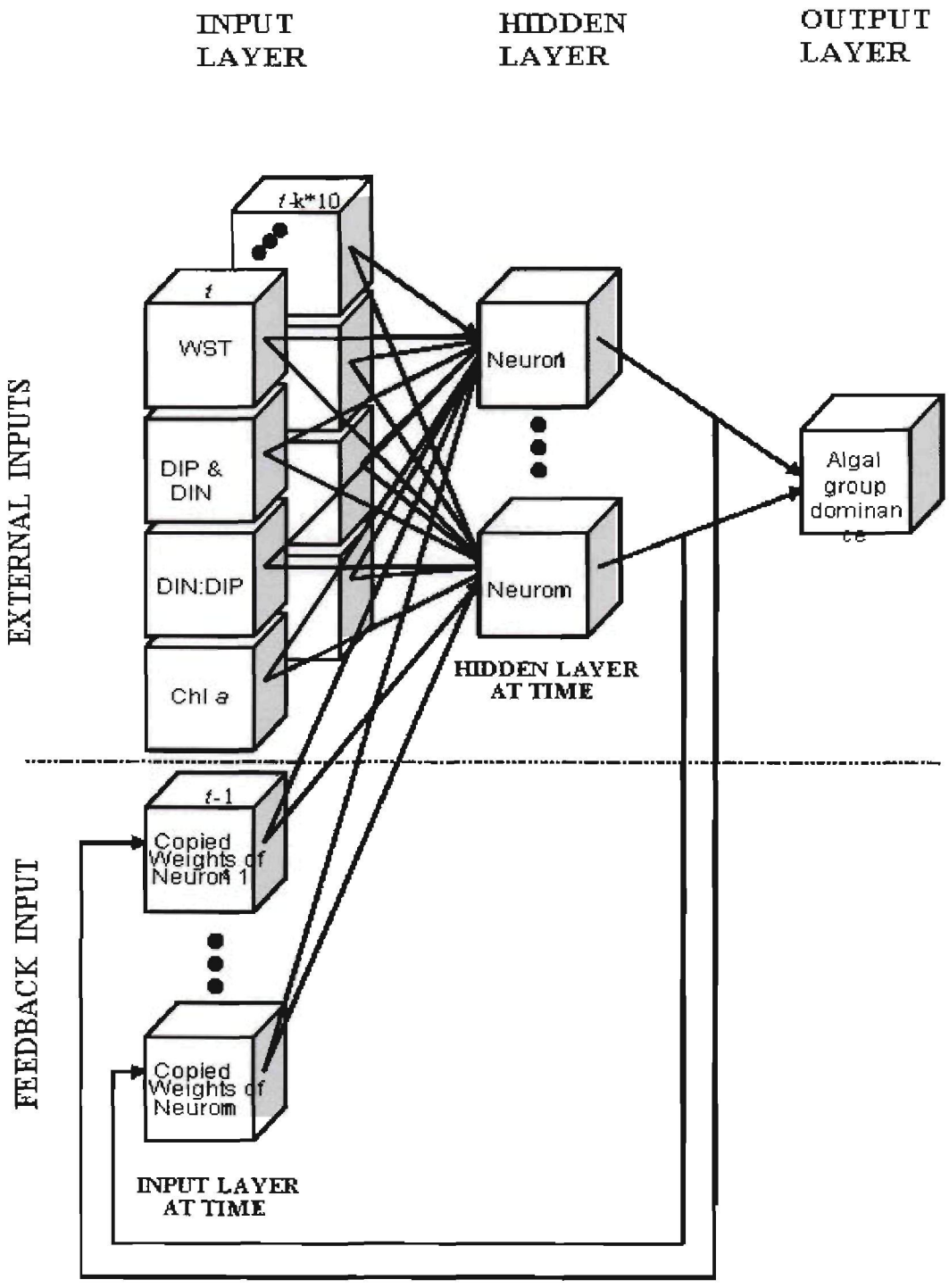


Figure 5.1 The structure of the supervised MFNN as used in this study (redrawn from Recknagel *et al.* 2006).

The input layer in this study, includes the variables WST, DIP, DIN, DIN:DIP and Chl *a*. This then feeds into the hidden layer (Neuron 1-n), which is fed into the feedback inputs. The final layer is the output, in this case the algae group as determined by the input data.

The software package initially selected for the study was NeuroSolutions 5.04, by NeuroDimensions Inc. Due to certain limitations encountered during the study, the package Forecaster XL, was subsequently selected. Both NeuroSolutions 5.04 and Forecaster XL are subsequently discussed.

5.1.3.1.1 NeuroSolutions 5.04

As stated in section 4.2.5 on the multivariate analysis the final set of environmental variables used for the training of the artificial neural networks (ANN) included the following measured parameters: water surface temperature (WST), dissolved inorganic phosphorous (DIP), dissolved inorganic nitrogen (DIN), the DIN:DIP ratio; and Chl *a* concentration. Furthermore, the community data was grouped into cyanotoxin (CT) producers and non-cyanotoxin producers (Non-CT) to simplify the modelling process.

5.1.3.1.2 Forecaster XL

The same data used with NeuroSolutions 5.04 was re-analysed with the Forecaster XL package, another MFNN. Firstly, the single dam (Hartbeespoort Dam) data was analysed (Experiment 1-4). Secondly, the analysis was done on the five reservoirs data (Experiment 5-12). In both these investigations, three approaches were followed to attempt the best fit and to determine the usefulness of the modelling techniques to South African Reservoirs and the importance of the main variables on the outcome of the model.

a) Firstly, prediction of the algae blooms with only the environmental variables included and without the measured algae dominance, and
b) Secondly, including the environmental variables as well as the algae dominance data.
c) Thirdly, including the environmental variables, algal dominance and Chl *a*.

The cyanobacteria were divided into two different groups, namely the Non-cyanotoxin group (Non-CT) and the Cyanotoxin producing group (CT). The Non-CT group included species like

Pseudoanabaena and *Merismopedia*. The CT group included all the known cyanobacterial toxin producers, e.g. *Microcystis*, *Anabaena*, *Oscillatoria* and *Cylindrospermopsis*.

5.1.3.2 Self-Organising Map (SOM)

The second type of neural network investigated is the self-organizing map (SOM). SOM is the most common example of a neural network, trained with un-supervised learning, and was developed by Kohonen (1982; 1988). The SOM requires that the data contain inputs that describe the characteristics of the variables or fields. The SOM then learns to ordinate and cluster or segment the data based on the similarities and differences of the input variables only (Smith 2002, Recknagel *et al.* 2006).

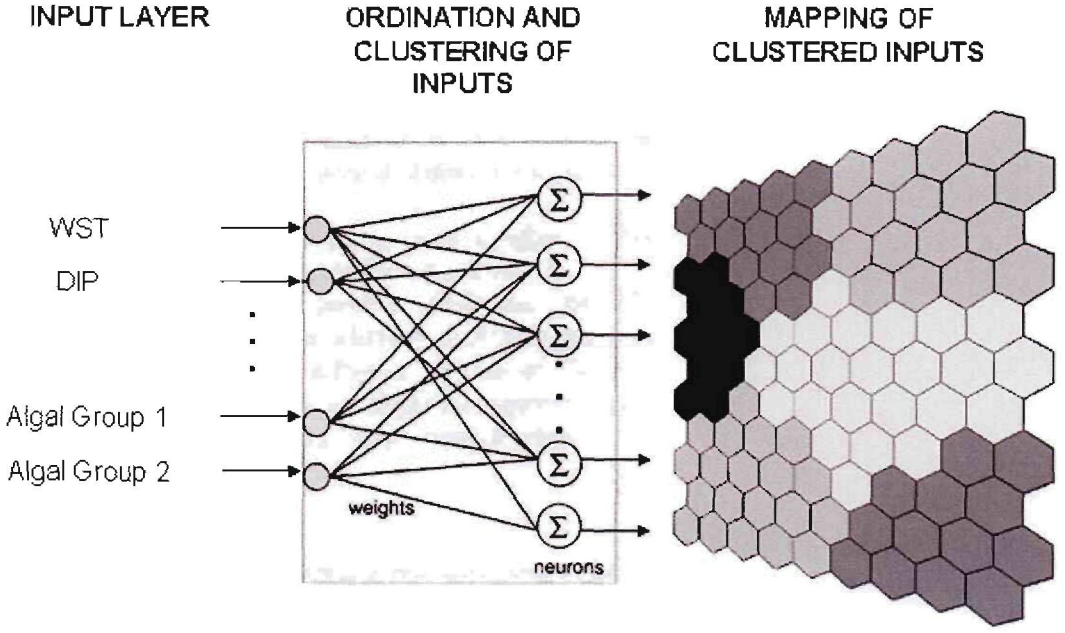


Figure 5.2 The structure of the non-supervised SOM for ordination and clustering of inputs (redrawn from Recknagel *et al.* 2006)

The SOM also has an input layer with the known variables and a hidden layer where the ordination and clustering are created with the mapping of the clustered output as the output layer (Figure 5.2).

The learning process, in the hidden layer, is roughly as follows:

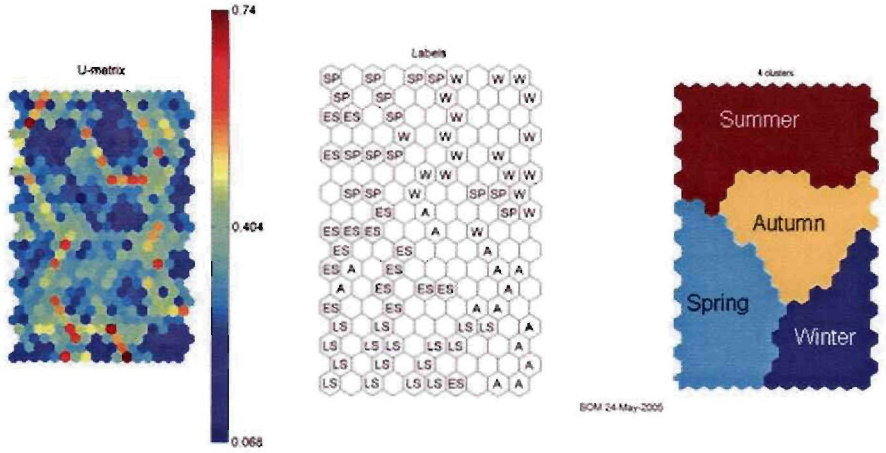
a) It initialises the weights for each output unit;

b) It cycles until weight changes are negligible for each input pattern (by presenting the input pattern, finding the winning output unit, finding all units in the neighbourhood of the winner and update the weight vectors for all those units; and

c) It reduces the size of neighbourhoods if required (Kohonen 2007).

The non-supervised SOM as introduced by Kohonen (1982, 1988) was applied to the before and after impact data sets to ordinate, cluster and map the nutrients, Chl *a* and algal groups with respect to seasons to determine if significant changes took place since the establishment of the Zeekoegat WCW that discharge directly into the Roodeplaat Reservoir.

The result of the training of the non-supervised SOM, by means of the normalised input data the Euclidean distance between the inputs, is calculated and then visualised as a distance matrix (the U matrix – Fig. 5.3a) and a partition map (K-means) (See Fig. 5.3b).



a) Distance matrix map (U-matrix)

b) Partitioned map (K-means)

Figure 5.3 The non-supervised SOMs ordination and clustering maps shown as a) a distance matrix map and b) a partitioned map (Redrawn from Recknagel *et al.* 2006)

The criteria for the ordination of the seasons is shown in Table 5.4 and was selected due to the air temperatures as shown in Table 3.3.

Table 5.4 The classification of seasons for the SOM modelling

Classification Criteria	Period
Summer	November - March
Autumn	April - May
Winter	June - July
Spring	August - October

5.1.4 Rule development by Hybrid Evolutionary Algorithms (HEA)

The bi-weekly data from three reservoirs, the Hartbeespoort, Rietvlei and Roodeplaat reservoirs for the period 1991 to 2004 were used. Simple linear interpolations were done with each data variable from each reservoir to create daily data sets. These interpolated data sets for the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs were merged because of the similarity in hypertrophic and climatic conditions and because the CANOCO analysis showed that, there was no distinct difference between the studied hypertrophic systems. For this study the data training were done with data from 1991 to 2003, and 1993 and 2004 were used for testing the rule sets, giving a total of 36 years for training and 6 years for testing. These two years were chosen because the extent of *Microcystis* dominance was low in 1993 and high in 2004 in the Hartbeespoort Reservoir. This enables the modeller to determine if the extent of high values and low values can be correctly forecasted (Recknagel 2006, *Pers. Comm.*).

A number of experiments were done to determine the potential to use the developed rule set for prediction of real time forecasting of *Microcystis*, 7-days, 14-days, 21-days and 28-days ahead forecasting including Chl *a* as one of the environmental variables. One experiment was done to determine the rule set when Chl *a*, was not included as environmental variable. Similarly, the method was also used to develop a rule set for the real time forecasting of the dinoflagellate, *Ceratium*, a species that is becoming increasingly a nuisance within the studied reservoirs.

The evolutionary algorithm (EA) methods mimics processes of biological evolution, natural selection and genetic variation. The method uses genetic operators and the 'survival of the fittest' principle to search for suitable representations of a problem solution (Cao *et al.* 2006). The ability of the method to apply self-organization, self-learning, intrinsic parallelism and generality,

enables EA's to be applied to recognize patterns, predict outcomes, optimize control and do parallel processing (Goldberg 1989; Bäck *et al.* 1997).

The basic framework of the rule discovery for *Microcystis* biovolume in hypertrophic South African reservoirs is represented in Fig 5.4. The detailed algorithm for the rule discovery and parameter optimization by HEA is the same as used by Cao *et al.* (2006) (Fig 5.5).

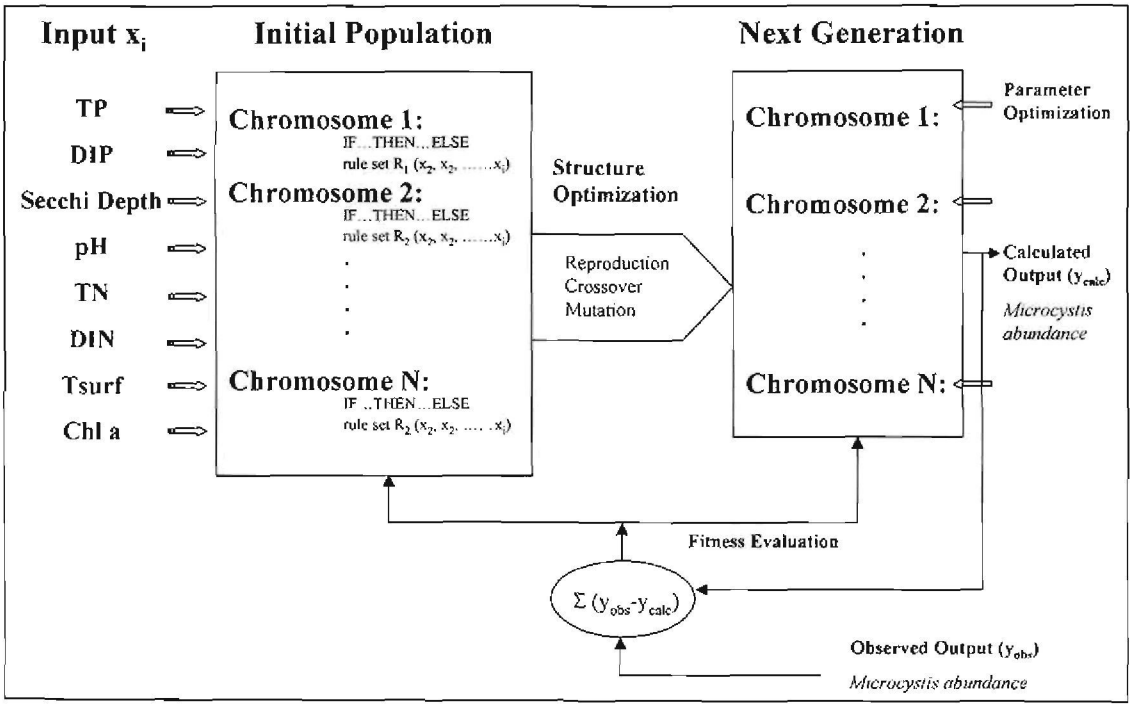


Figure 5.4 Conceptual diagram of HEA for the discovery of a predictive rule set for *Microcystis* biovolume in three hypertrophic South African reservoirs. The same model was used for the real time prediction of the dinoflagellate biovolume predictive rule set.

Genetic programming (GP) is used in the HEA to generate and optimize the structure of rule sets. The genetic algorithm (GA) is used to optimize the parameters of the rule set. GP computer programs are represented as parse trees, where a branch node represents an element from the functions set (this can be arithmetic operators, logic operators and elementary functions of at least one argument). A leaf node can be an element from a terminal set (this can be variables, constants and functions of no arguments) (Muttil and Lee 2005; Cao *et al.* 2006). With each run of the

program the results are evaluated by means of 'fitness cases'. Fitter results are selected for recombination to create the next generation by using the genetic operators, e.g. crossover and mutation. These steps are repeated for consecutive generations until the criteria for termination of the program are met. A general genetic algorithm is then used to optimize the random parameters in the rule set.

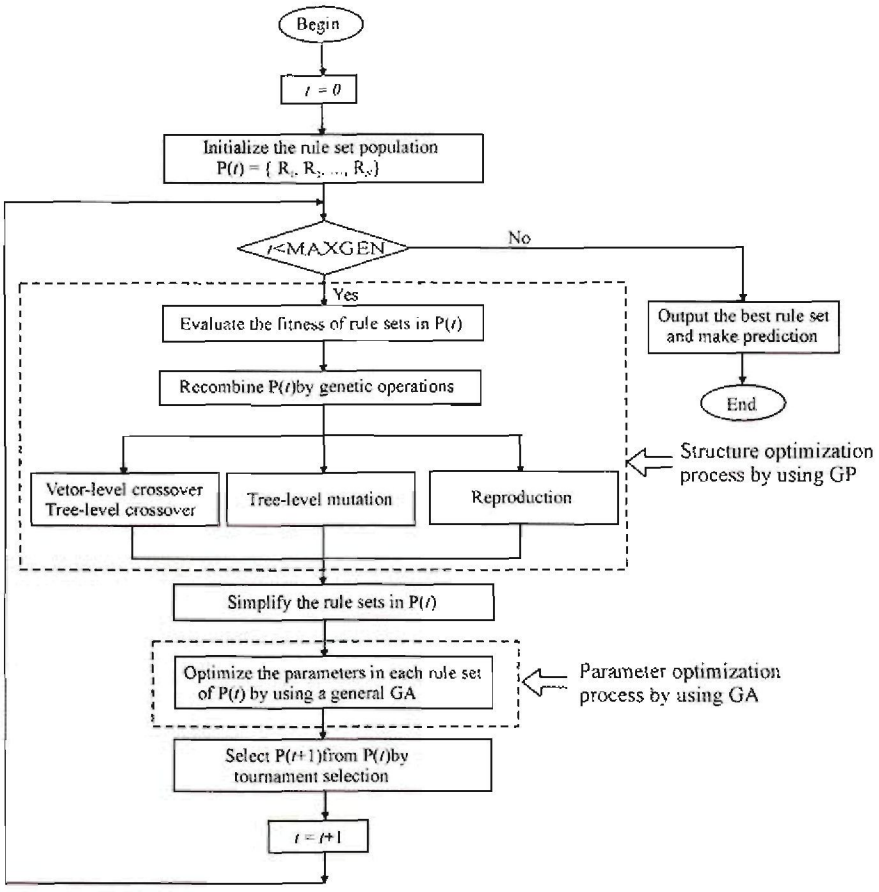


Figure 5.5 Flow chart showing the process of the HEA (redrawn from Cao *et al.* 2006).

The rule sets are expressed in an IF THEN/ELSE tree format and are simplified as discussed in Cao *et al.* 2006. This allows for the development of the rule set to consider and make provision for different conditions within the data set. The model output includes a list of the variables used within each run, as well as the list of outcomes and the effectiveness of each run (Appendix A).

For the experimental programming using HEA, the Hartbeespoort, Rietvlei and Roodeplaat reservoirs data was used to do real time prediction for *Microcystis* and the dinoflagellate,

Ceratium. These two species were selected because they formed the major blooms in the five reservoirs. Experiments 1, 3, 4, 5, 6 were run to model real-time, 7-days forward, 14-days forward, 21-days forward and 28-days forward for the prediction of the seasonal succession of *Microcystis*. The daily input data of the following variables: TP, DIP, Secchi depth, pH, TN, DIN, Tsurf, Chl *a* and *Microcystis* biovolume were used. The data preparation for the forwards experiments were done by shifting the daily algae biovolume data backwards the number of days that are shown in the forward prediction. Training was then done on this data set.

In Experiment 2 Chl *a* was omitted as an environmental variable to see if a reliable rule set could be developed without the Chl *a* as indicator of biovolume. In Experiment 7 a rule-set was developed for the dinoflagellate, *Ceratium*. The daily input data of the following variables: TP, DIP, Secchi depth, pH, SO4, TN, DIN, Tsurf, Chl *a*. and dinoflagellate biovolume were used. For the application of HEA during this study an initial population of 200, a maximum of 100 generations and 20 runs were done on the data set. The experiment was performed on a Hydra supercomputer (IBM eServer 1350 Linux) with a peak speed of 1.2 TFlops using C++ programming language. The parameter settings are listed in Table 5.5.

Table 5.5 Parameter settings of the hybrid evolutionary algorithm rule set discovery for the Hartbeespoort, Rietvlei and Roodeplaat reservoirs

Structure Optimization (GP)	N = 15341 F_L = {AND, OR} F_C = {>,<, ¡Ã, ¡Â} F_A ={+,-,*,/,sin,cos, exp,ln} MAXK = 4 $D_{IF} = D_{THEN/ELSE}$ = 4 MAXGEN = 100
Parameter Optimization (GA)	Popsize = 200 a = -0.5 b = 1.5 M = 8 MAX = 500

To validate the results of the rule sets the correlation coefficient of the measured and fitted data and the Root Mean Square Error (RMSE) of the training error, and the testing error are provided as outcomes of the model. The RMSE are calculated as follows:

$$\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2} \qquad (5.3)$$

Where m is the number, of testing data points, y_i and $\hat{y}_i$ are the *i*th observation and the *i*th predicted value of the output variable *Microcystis*/dinoflagellate biovolume.

More validation was done on the best rule set for every Experiment by determining the Average Actual Error (A AE) and the Mean Square Error (MSE) by conventional methods.

To consider the impurities between the actual measurements and the predicted values in the model outcome the Relative Tolerance was set at 10% for the training data and 30% for the testing set, to determine the percentage good and bad predictions within each best rule set for the training and the testing data sets of each experiment. The formula, used to determine the Relative Tolerance, is shown in formula 5.5 and 5.6 (Greenberg 2003).

$$| v\text{-}V | \leq \tau_r | V | + \tau_a \qquad \textbf{(5.5)}$$

Where V is the measured value, v the predicted value, τ_r is the Relative Tolerance and τ_a is the Absolute Tolerance.

However, when the measured value V = 0, then

$$| v | \leq \tau_0, \qquad \textbf{(5.6)}$$

Where $\tau_0 = 3$. This Tolerance is pre-determined to be an acceptable outcome of the model.

Sensitivity analysis for each environmental variable, used in the RULE set, was done by determining the biovolume outcome if an environmental variable changes stepwise within the range of the data for the specific variable. This was done for each variable used in a rule set. The median of the other variables used in a rule set, was used to determine the impact of a specific environmental variable. The impact on the biovolume outcome of the environmental variables used within each rule set was plotted together to verify the most important variable regarding the effect and greatest impact on the algal biovolume changes. The greater the slope, the greater is the impact on the outcome of the model.

5.2 Modelling Results

5.2.1 Vollenweider Model

The Vollenweider model is an easy to apply empirical model but does not take all the biological processes, into consideration. The model assumes that the system is in a steady state when Thornton and Walmsley (1981) tested the Vollenweider (1976) and Dillen and Rigler (1974) models on South African conditions, it was found that there is a certain degree of potential for predicting the steady state of a system.

This model was developed for eutrophication management purposes and can be used to determine the trophic conditions within a system, and allows the modeller to apply quick management scenarios.

Table 5.6 shows that for the Hartbeespoort Dam, the model over-predicts the TP concentrations by 28.0% and under-predicts the Chl *a* concentration with 68.1%. In the Klipvoor and Roodeplaat dams, the model under estimated the TP concentration with 78.8% and 29.8% respectively. The Chl *a* concentrations was under-estimated in the Klipvoor Dam by only 5.8% and in the Roodeplaat Dam, an over estimation of Chl *a* of 26.5% was found. There is therefore no uniformity in the predictions between these three Reservoirs. However, IETC (2000) maintains that this model can have uncertainties from as low as 30% up to 300%. This may be quite acceptable because of the differences that may exist between these reservoirs. The acceptable discrepancies that are allowed are much higher than the 10% to 30% that are acceptable for the Forecaster XL ANN model that was tested with the data from the hypertrophic South African reservoirs.

The scenarios were randomly selected to test different scenarios and to show what the potential outcome of management actions can be. The 50% P elimination (Table 5.6) showed that there will be a significant reduction in the predicted TP, Chl *a,* and PO_4-P concentrations. This may be enough to improve the hypertrophic status to eutrophic in all three the reservoirs. The 90% elimination of P may change the reservoirs water quality into an oligotrophic state. This may not be achievable due to potential costs involved and it must be kept in mind that there may be uncertainties between 30% and 300%, especially if the under prediction in the case of the Klipvoor Dam is considered.

The Vollenweider Model is, therefore, still quite a usable tool for managers, but the potential discrepancies must be kept in mind when this model is used. It is a useful tool when a system is fairly unknown and little data is available to assist the manager in making decisions regarding such a system.

Table 5.6 Input variables for the Vollenweider Model with the results found by applying three different management scenarios in the hypertrophic Hartbeespoort, Klipvoor and Roodeplaat dams.

Input Variable/Scenarios	**Hartbeespoort**	**Klipvoor**	**Roodeplaat**
TAI (m^3/year)	985270000	113426870	63850000
MIP (mg/m^3)	373	742	369
MAPL (mg/year)	367505710000	84162737540	23554999558
SA (m^2)	20624100	7580000	3970000
PAL (mg/ m^2/year)	17819	11103	5933
V (m^3)	194637710	43800000	41900000
RT (year)	0.20	0.39	0.66
MD (m)	9	6	11
TP (ug/L)	101	751	205
Chl *a* (ug/L)	123	142	33.2
PO_4-P (ug/L)	52.1	50	112
Predicted Reference concentrations	TP: 130 Chl *a*: 39.2 PO_4-P: 43.5	TP: 159 Chl *a*: 47.1 PO_4-P: 52.9	TP: 144 Chl *a*: 42.0 PO_4-P: 48.1
Scenario P Elimination 50%	TP: 65 Chl *a*: 20.9 PO_4-P:21.7	TP: 79 Chl *a*: 24.9 PO_4-P: 26.5	TP: 72 Chl *a*: 22.9 PO_4-P: 24.0
Scenario P Elimination 90%	TP: 13 Chl *a*: 6.3 PO_4-P: 4.3	TP: 16 Chl *a*: 7.1 PO_4-P: 5.3	TP: 15 Chl *a*: 6.7 PO_4-P: 4.8
Scenario Doubling the residence time	TP: 264 Chl *a*: 76.6 PO_4-P: 88.0	TP: 164 Chl *a*: 48.7 PO_4-P: 54.8	TP: 88 Chl *a*: 27.3 PO_4-P: 29.3

5.2.2 Lake simulation library SALMO-OO

The results discussed here will focus on the outputs from the SALMO-OO model and the improvements made by replacing the original SALMO-OO growth model by the growth model of Arhonditsis and Brett (2005) for Lake Washington. The simulation library SALMO-OO developed by the object-oriented implementation of the lake ecosystem model SALMO-OO (Recknagel and Benndorf 1982; Benndorf and Recknagel 1982; Recknagel *et al.* 1995) is the core of a lake simulation library, implemented within JAVA. It provides optional access to alternative causal representations of ecological processes in lakes such as photosynthesis and respiration of diatoms, green and blue-green algae; grazing of diatoms, green and blue-green algae by zooplankton; growth of herbivorous zooplankton, and predation of zooplankton by planktivorous fish. Alternative process representations were adopted from published lake ecosystem models different to SALMO such as Park et al. (1974), Hongping and Jianyi (2002), and Arhonditsis and Brett (2005).

The original SALMO-OO over-estimated the phosphorus and total algal population in Hartbeespoort and Roodeplaat Reservoirs (Fig 5.6 & 5.8). The actual and forecasted results for the Klipvoor Reservoir were much closer, although a certain period during 2003 in Klipvoor was over-estimated (Fig 5.7). The correlation coefficients of all three of the reservoirs between the measured and predicted variables was low and does not seem to reflect the conditions in the reservoirs, with the original SALMO-OO (Fig 5.6 – Fig 5.8).

The application of the Arhonditsis and Brett (2005) growth equation in SALMO-OO did reduce the errors. In the Hartbeespoort Reservoir the fitted PO_4-P concentrations were much closer to the actual measured data, and the RMSE was improved from 79.67 to 37.67 (Fig. 5.6), however, the measured and predicted peaks did not correspond well. With SALMO-OO the total algal biovolume of the Hartbeespoort were slightly over-estimated and there was an improvement of the RMSE from 24.18 to 19.87, after the application of the Arhonditsis and Brett (2005) equation.

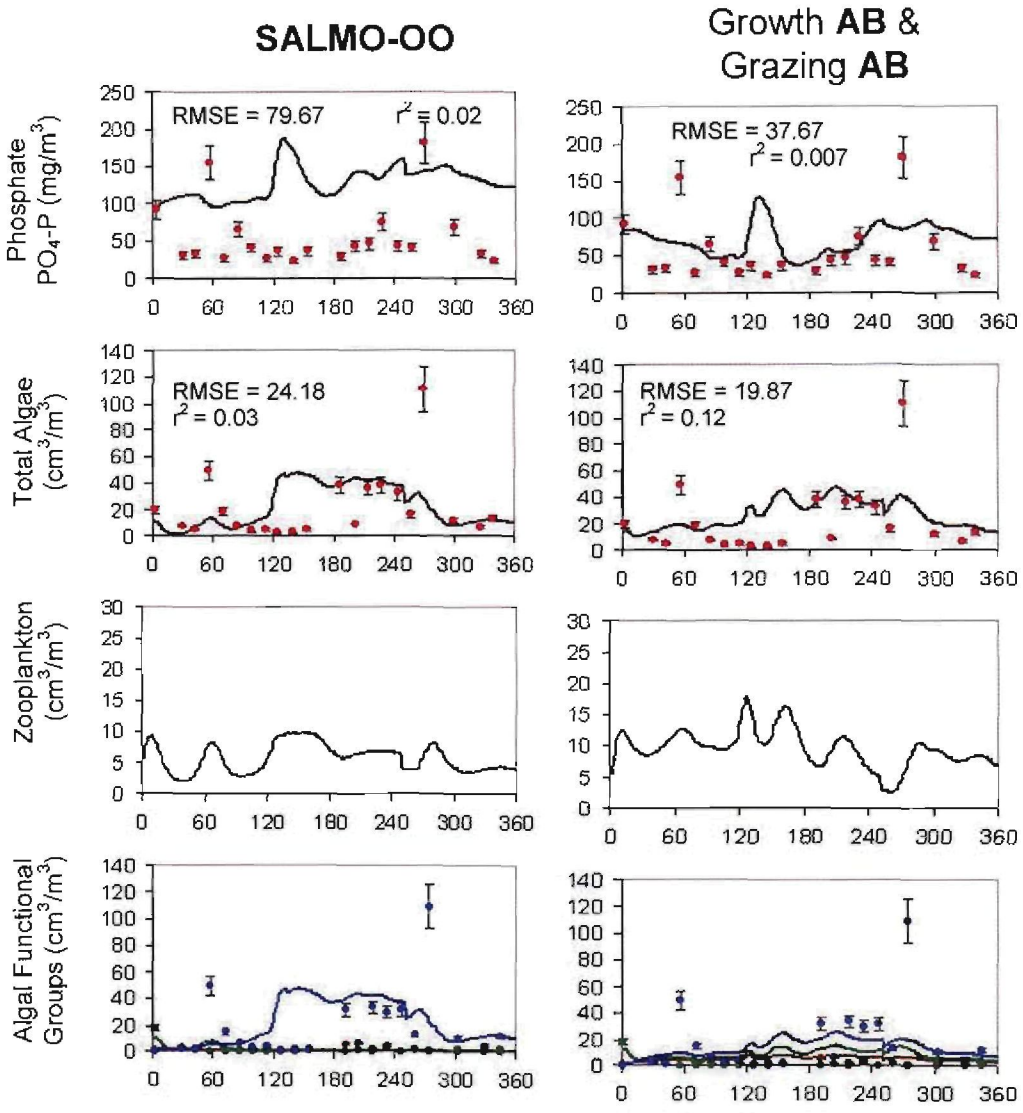


Figure 5.6 Comparison of output results (line) with real measured data (markers) for the Hartbeespoort Reservoir with the original SALMO-OO growth and grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available.

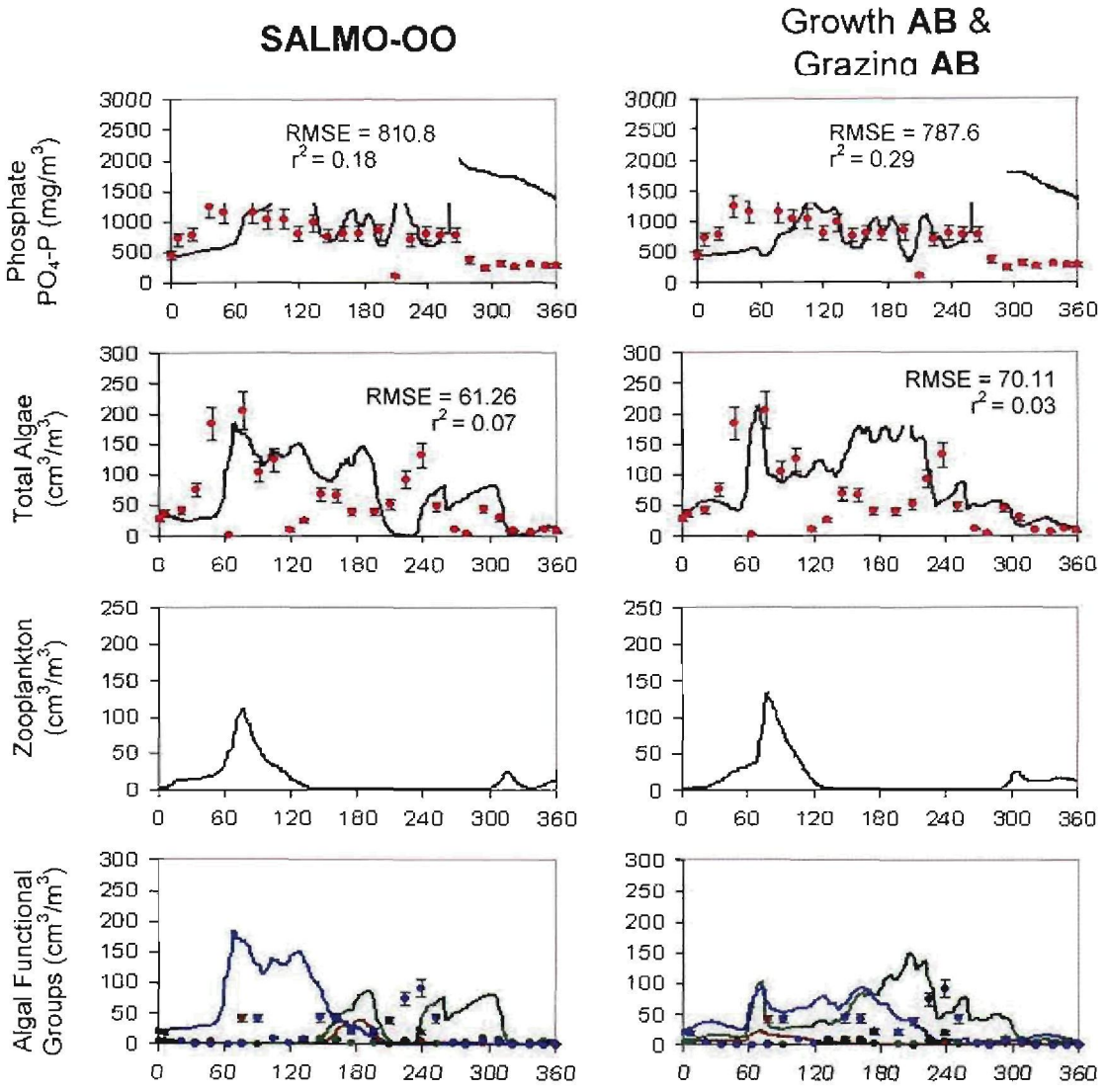


Figure 5.7 Comparison of output results (line) with real measured data (markers) for the Klipvoor Reservoir with the original SALMO-OO growth and grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available.

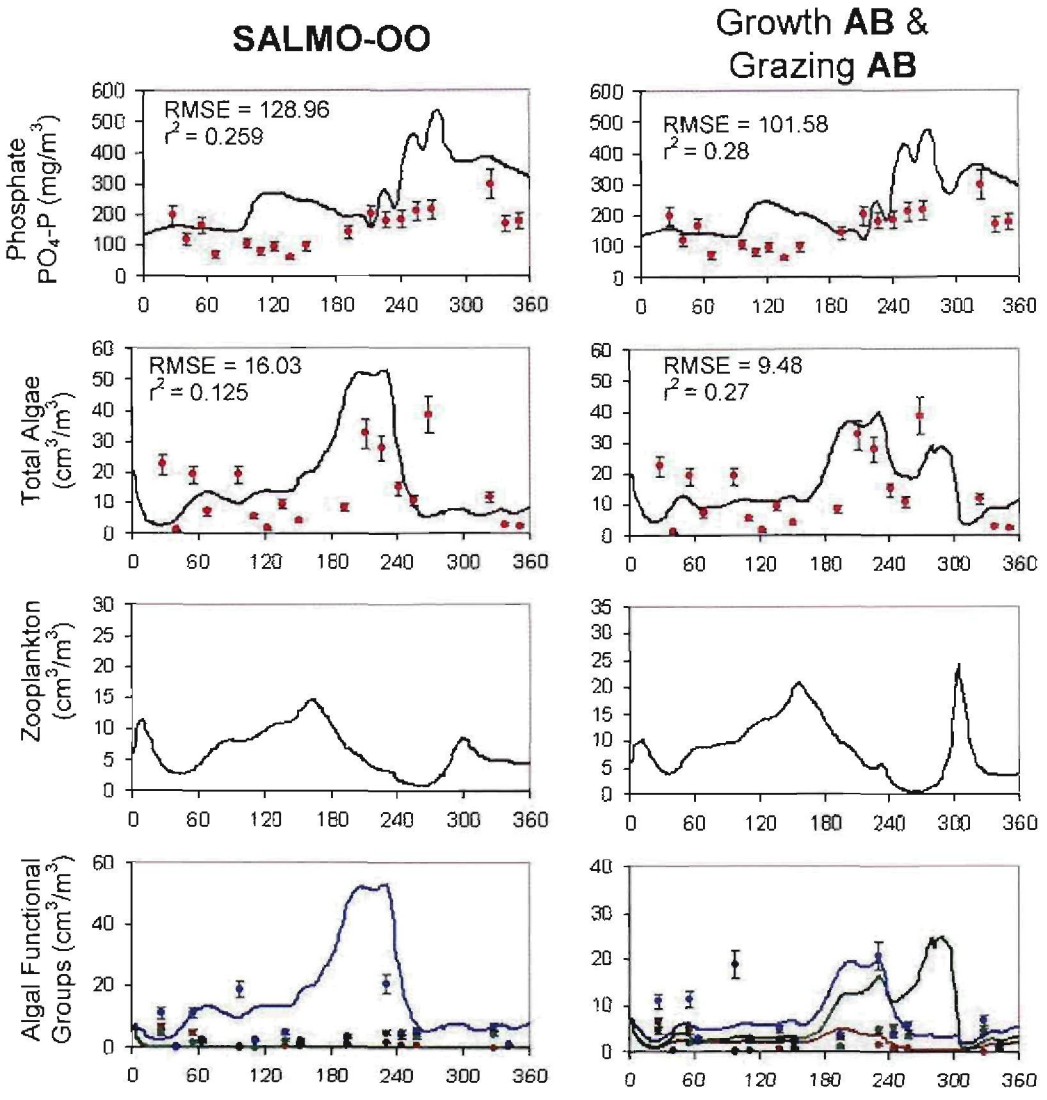


Figure 5.8 Comparison of output results (line) with real measured data (markers) for the Roodeplaat Reservoir with the original SALMO-OO growth and grazing equations; and with the adaptation of the growth and grazing equations from Arhonditsis and Brett (AB) (2005). Zooplankton is part of the model output and is thus included in the graph, although no zooplankton data was available.

5.2.3 Artificial Neural Networks

The artificial neural networks (ANN) are a black box approach where the modeller does not know what is happening within the hidden layer of the model. The modeller does not need to know the complex characteristics within a water system to apply the model. However, the results can be quite accurate regarding the prediction of the outcomes. The outcomes of the model, despite the black box approach, can add insight as weights are allocated to the environmental variables in predictions. This supplies the modeller with the most important environmental variables in predicting algal biomass and blooms.

5.2.3.1 Results with Multilayered Feed forward Neural Network

5.2.3.1.1 NeuroSolutions 5.04

Originally, data from Roodeplaat Dam was modelled with the ANN (NeuroSolutions 5.04). This modelling package achieved very good results. The results using this model are summarized in Table 5.7. The package was able to achieve 93.75% accuracy with the cyanotoxin producing group (CTTot exp) and 90 % accuracy with the non-cyanotoxin (TotNONCT) producing group. A graphical representation of the results is presented in Figure 5.9 and Figure 5.10.

The normalised mean square error (NMSE) and the mean square error (MSE) are the performance measurements used to determine the accuracy of the model. The NeuroSolutions gives both the MSE and the NMSE as well as the mean actual error (MAE). The smaller these values the more accurate the prediction results. According to Velickov (2004), the NMSE is dimensionless and provides a better indicator for the error measurement since it is normalised by the error of the observed data. The r-value as indicated in Table 5.7 shows that there was a very good correlation between the predicted and the measured data. However, the minimum and maximum absolute errors do show that there are potentially, large uncertainties in the outcomes.

Attempts were made to construct the solution equation, using NeuroSolutions 5.04 and other packages. However, the main disadvantage of ANN modelling is that it is a black box approach, and the developed algorithm, that provides the output of the model, could not be determined. However, it must be kept in mind that the aim of the study was to test the applicability of the

models for harmful algal bloom prediction, and not to develop improved knowledge of the models as such.

The final step was to include the data from all the reservoirs in the training data set and leave one reservoir out for testing the model.

Table 5.7: Summary of the initial neural network results using NeuroSolutions for prediction of the cyanotoxin producing group (CTTot exp) and the non-cyanotoxin producing group (TotNONCT).

Performance	*CTTot exp*	*TotNONCT*
MSE	225.9431157	1088.47948
NMSE	0.024515995	0.011171232
MAE	12.91183814	21.47676731
Min Abs Error	0.597656353	0.136831959
Max Abs Error	33.21688555	142.0277588
r	0.987846519	0.994490043
Percent Correct	93.75	90

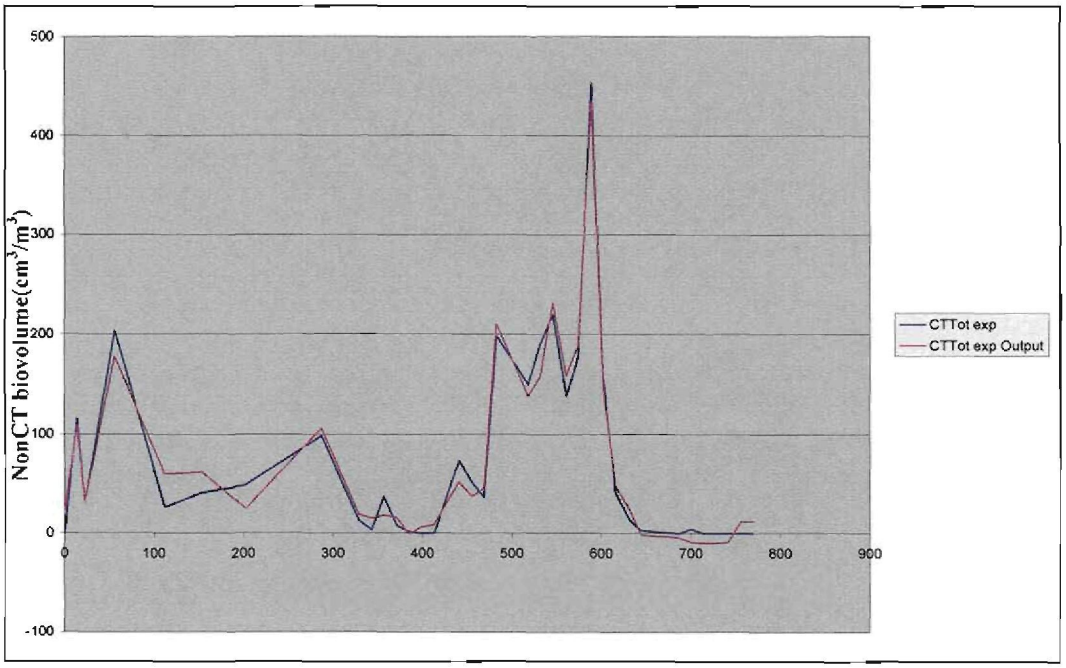


Figure 5.9 Results from the NeuroSolutions 5.04 ANN for the cyanotoxin producing group (CTTot exp), showing real time forecasting on the Roodeplaat Dam data.

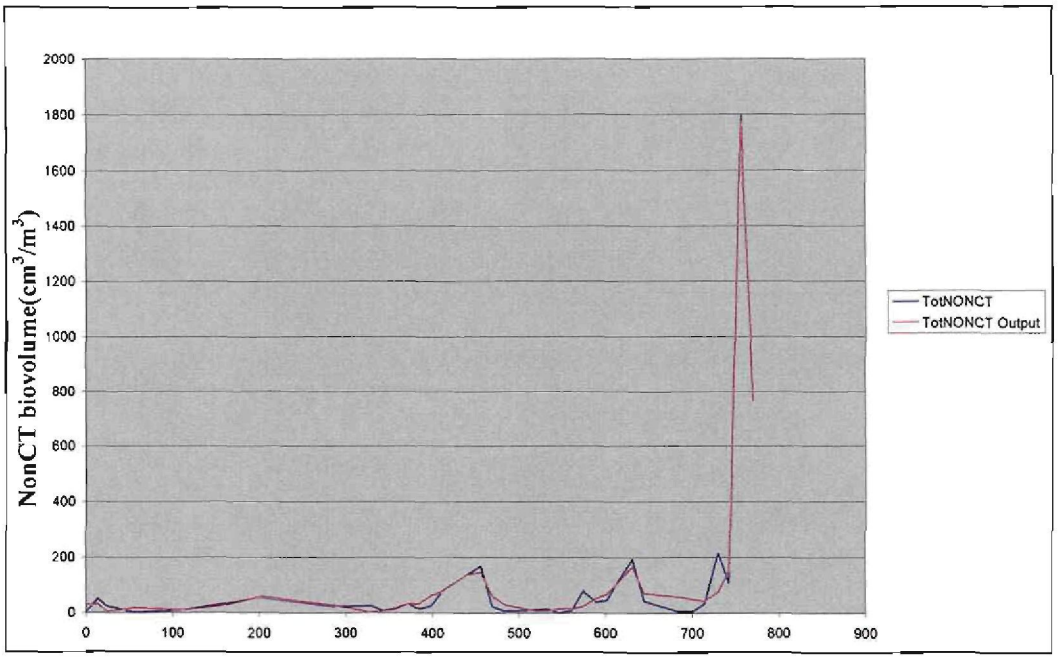


Figure 5.10 Results from the NeuroSolutions 5.04 ANN for the Non-CT producing group (TotNONCT), showing real time forecasting (pink line) versus the actual data (blue line).

5.2.3.1.2 Forecaster XL

At first only the data from Hartbeespoort Dam was modelled with the ANN (Forecaster XL). This modelling package achieved very good results and gave a better understanding of the results because of the build in verification. For each experiment, the results from this model are summarized in two graphs and a table, showing the errors and the relative tolerances of the training and testing data set. The percentage relative tolerance of the prediction versus actual data in the modelling package for the training set is 10%, and with the testing data set 30%. The model is, therefore, quite strict with itself in accepting the efficiency of the model.

At the end of this section a summary of the results are shown. This is done to compare the outcome of the experiments and to compare the important variables to be included in modelling the forecasts for the cyanotoxin producing cyanobacteria (CT) and the non-cyanotoxin producing cyanobacteria (Non-CT) groups.

Modelling using data from a single dam

Experiment 1: Non-cyanotoxin producer (Non-CT) group with only selected environment variables as input (single dam)

Using only the measured environmental variables for the period 2000 to 2005 for the modelling exercise, the following results were obtained. The r^2 value was 0.9595 and the correlation was 0.9807. Figure 5.11 shows the actual measurements against the forecasted results for the Non-CT producing group. The results show that the forecasted results do occasionally over, or under predict, yet have very good correlations as mentioned. According to the results in Table 5.8, the number of good results is 0, which means that more than a 30% error was found within predicting and comparing all the test results.

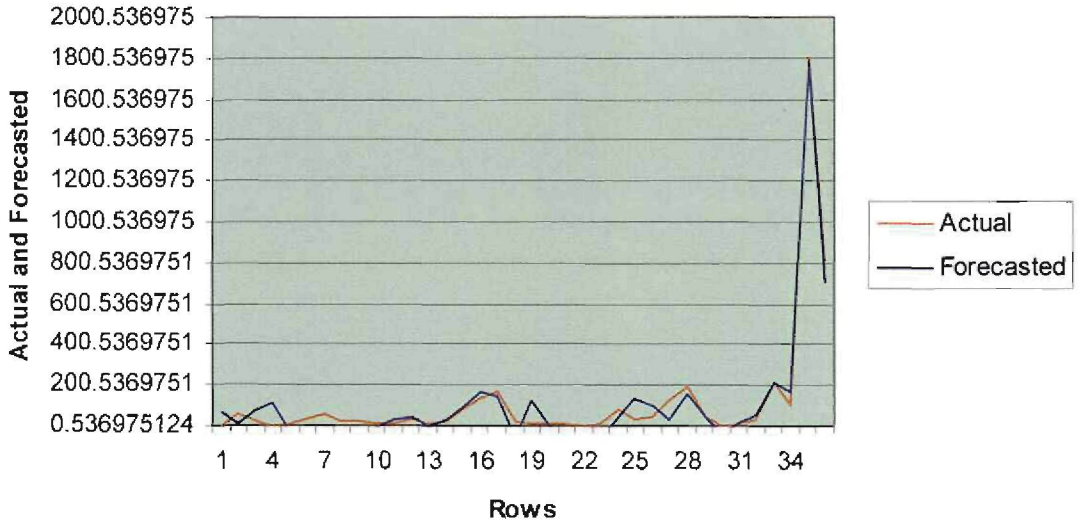


Figure 5.11 Actual (orange line) versus forecasted (blue line) results for Non-CT producing group biovolume using only environmental variables and Chl *a* concentration as input with the Forecaster XL ANN software package.

Table 5.8 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using only environmental variables and Chl *a* concentration as inputs.

	Training set	**Test set**
# of rows:	30	6
CCR:	n/a	n/a
Average AE:	46.791955	62.012307
Average MSE:	3862.6524	4339.2413
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	5 (17%)	0 (0%)
# of Bad forecasts:	25 (83%)	6 (100%)

One of the excellent features of this model is the graph that shows the percentage of importance of each variable that are used in forecasting the outcomes. Fig. 5.12 shows that with this first modelling attempt Chl *a*, was by far the most important variable in determining the outcomes of the modelling results.

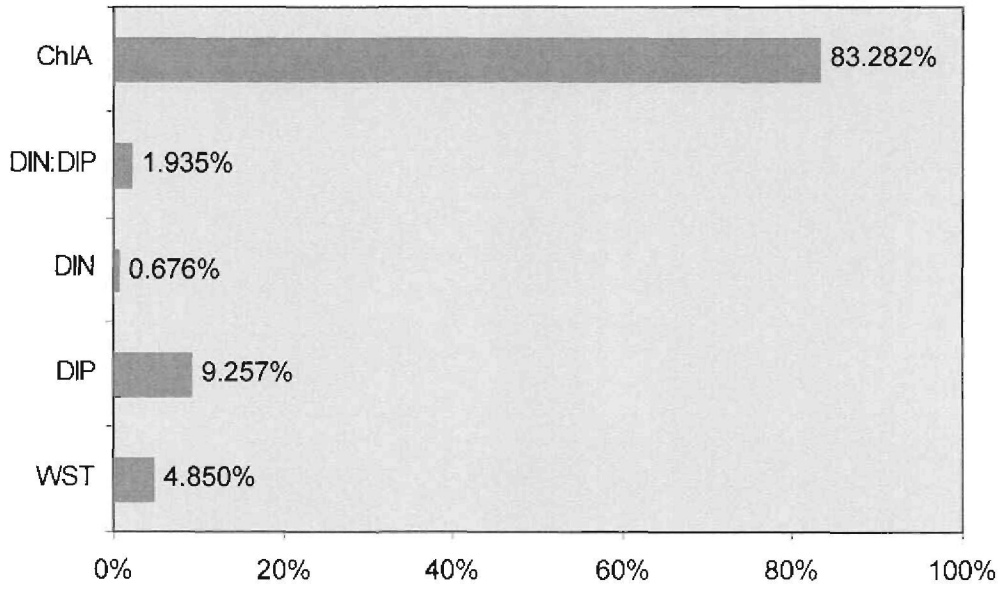


Figure 5.12 Percentage contributions of measured environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group in the Hartbeespoort Dam.

Secondly, the DIP concentration, and thirdly, the WST contributed to the outcome of the model. These results, therefore, indicate that the initial biomass is the most important determinant for future bloom development. Next are the nutrients, specifically in the form of DIP, and thirdly, the WST as a climate indicator in determining the forecasting of cyanobacterial biovolume.

Experiment 2: Cyanotoxin producer (CT) group with only selected environment variables as input (single dam)

In this second experiment with Forecaster XL only the selected environmental variables (See Section 4.2.5) were used to achieve results with predictive capacity for the CT group of the phytoplankton community. The correlation coefficient (r^2) value was 0.2746 and the correlation was 0.5620. The periods of dominance by the CT group was forecasted quite well although there are a number of over and under predictions (Fig. 5.13).

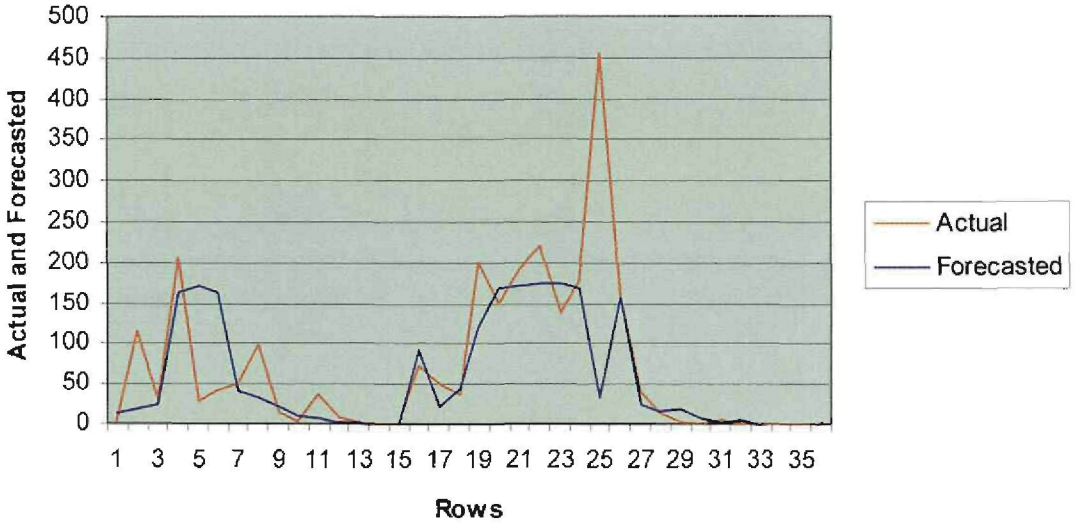


Figure 5.13 Actual (orange line) versus forecasted (blue line) results for CT producing group using only environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package on the Hartbeespoort Dam data.

In this case, the percentage of accurate forecasts with the training set was only 7%. With the testing set, the number of good predictions with a tolerance of 30% was 50% (Table 5.9). This is still not accurate enough for the model as the average MSE and the average actual error (AE) for the testing set was much higher than with the training set.

In this instance the results show that in predicting the CT group the most important environmental variable is WST with a 60.91% contribution to the modelling success and secondly the DIP with a contribution of 18.17% (Fig. 5.14). DIN:DIP and DIN is also showing more importance (8.303% and 7.028% respectively) in the contribution to the outcome of the modelling results than the initial biomass, as indicated by the Chl *a* concentration. Multivariate analyses has shown that WST, as the climatic indicator, (See Section 4.2.5) is the most important environmental variable. The importance of phosphorous and the DIN:DIP ratio is in accordance with existing literature (Vollenweider 1990, Rossouw 2000, Rossouw and Görgens 2005).

Table 5.9 A summary of the modelling result output with the Forecaster XL ANN software package for the cyanotoxin (CT) producing group using only environmental variables and Chl *a* concentration as input.

	Training set	**Test set**
# of rows:	30	6
CCR:	n/a	n/a
Average AE:	23.211721	102.33326
Average MSE:	1731.8002	31450.745
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	2 (7%)	3 (50%)
# of Bad forecasts:	28 (93%)	3 (50%)

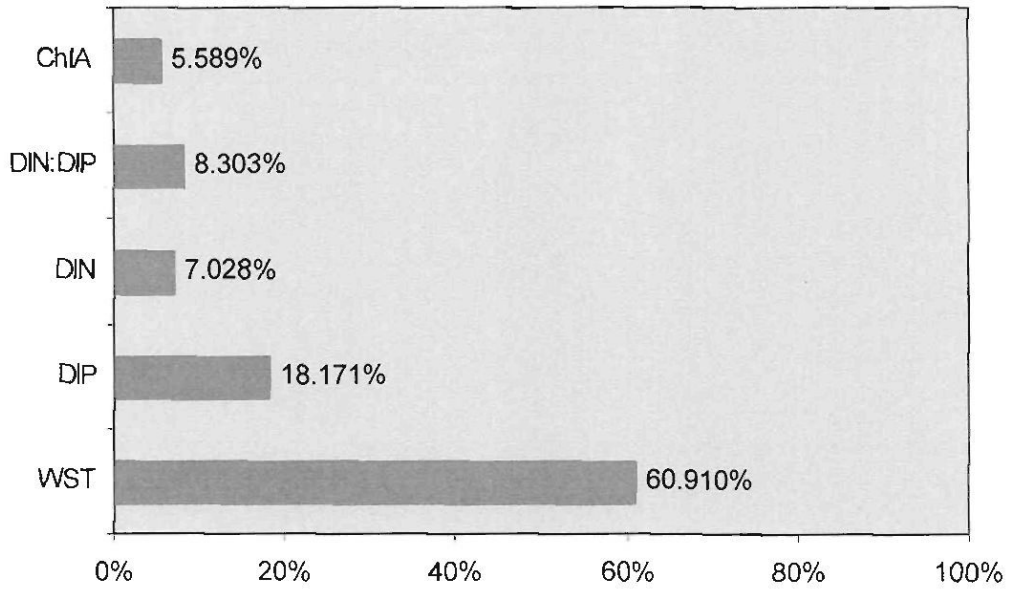


Figure 5.14 Percentage contributions of measured environmental parameters and Chl *a* concentration to ANN prediction of the CT group.

Experiment 3: Non-Cyanotoxin producer (Non-CT) group algae dominance, Chl *a* concentration and environment variables as input to the Forexter XL model (single dam)

With this experiment, the algal dominance was included in the modelling together with the measured environmental variables and the Chl *a* concentration to determine the predictive capacity for the Non-CT group. The actual data is the measured data, while the forecasted data is the predicted outcome of the model. The r^2 value was 0.9963 and the correlation was 0.9982. The results are shown in Fig. 5.15.

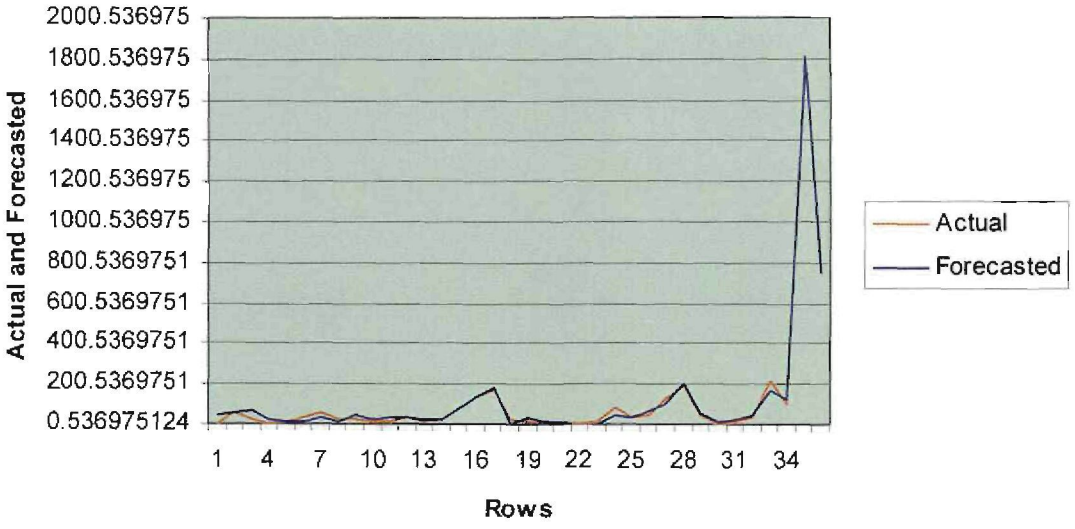


Figure 5.15 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package.

Table 5.10 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs.

	Training set	Test set
# of rows:	30	6
CCR:	n/a	n/a
Average AE:	13.922402	23.618138
Average MSE:	293.51034	692.71795
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	7 (23%)	2 (33%)
# of Bad forecasts:	23 (77%)	4 (67%)

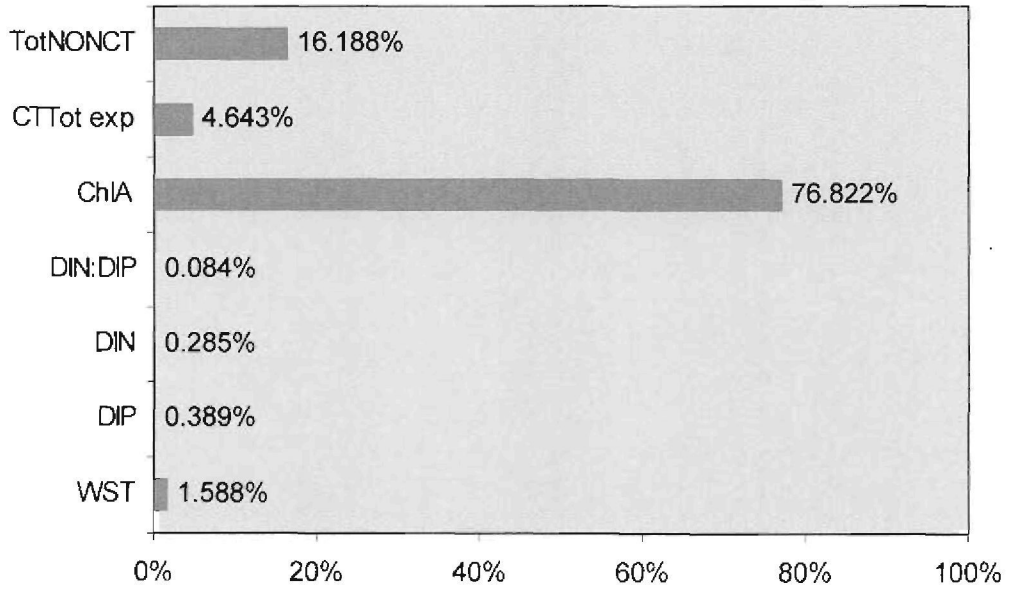


Figure 5.16 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group.

The results with experiment 3 indicate that with the inclusion of the algal group dominance the average actual error (Average AE) is much smaller than with the previous results (Table 5.10). The number of good predictions with the training set is only 23%, but the 10% tolerance might be

too strict. With the testing dataset the percentage good predictions for the Non-CT group is 33%. This is much higher than what was found with Experiment 1.

The most important variables in the prediction of the Non-CT group are again Chl *a*, with an importance of 76.822% (Fig. 5.16). Secondly, the dominance determination of the total Non-CT groups that was added in this experiment was used and the importance of the input was 16.188%. Thirdly, the CT group dominance values were used and had an importance of 4.643%. Fourthly, the WST had an importance of only 1.588%.

With experiment 3, the nutrient components did not feature as important factors in determining the algal group contribution in the phytoplankton community. There are two ways of looking at the results. Firstly, it may indicate that including the Chl *a* concentration as a variable in the Forecaster XL model, minimises the effect of the selected environmental variables for forecasting purposes. Secondly, it may indicate that the Chl *a* concentration as an indicator of the biomass/biovolume is closely related to the specific species dominant in the system.

Experiment 4: Cyanotoxin producer (CT) group algae counts and environment variables as input (single dam)

Using the measured environmental variables together with the algal group dominance in the reservoir the results shown in Fig. 5.17 were obtained. The r^2 value was 0.9843 and the correlation was 0.9932. The correlations show, therefore, good results.

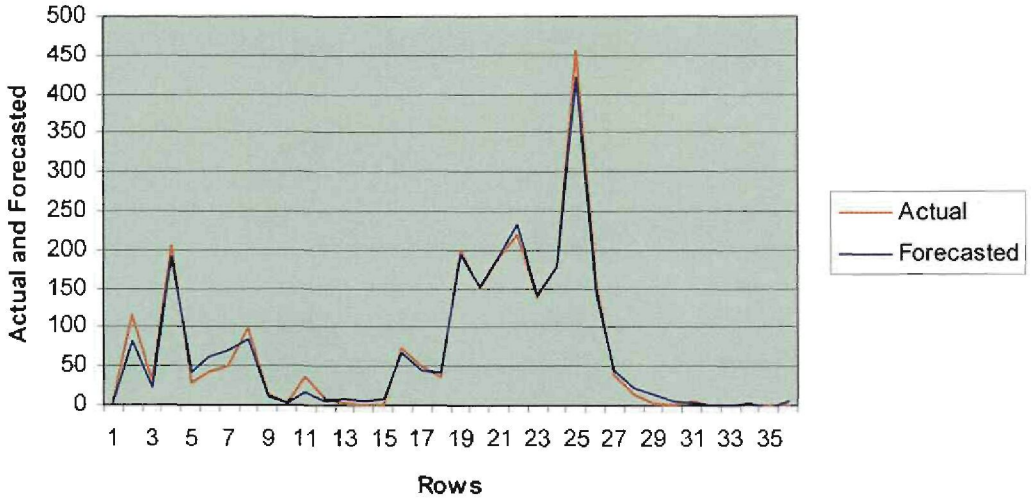


Figure 5.17 Actual (orange line) versus forecasted (blue line) results for the CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package.

The summary of the modelling results is shown in Table. 5.11. The results show that the average actual error (Average AE) is much lower than any of the previous results (Experiments 1-3). This indicates that the results are significant. However, the tolerance values for the training set are again not within the acceptable tolerance of the model of 10%. Only 37% of the results fall within the 10% tolerance limit. With the training data set, only 17% good forecasts are within the tolerance limit of 30%.

Table 5.11 A summary of the modelling result output with the Forecaster XL ANN software package for the CT producing group using algal dominance, environmental variables and Chl *a* concentration as inputs.

	Training set	**Test set**
# of rows:	30	6
CCR:	n/a	n/a
Average AE:	6.5620637	17.712396
Average MSE:	88.862874	421.084
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	11 (37%)	1 (17%)
# of Bad forecasts:	19 (63%)	5 (83%)

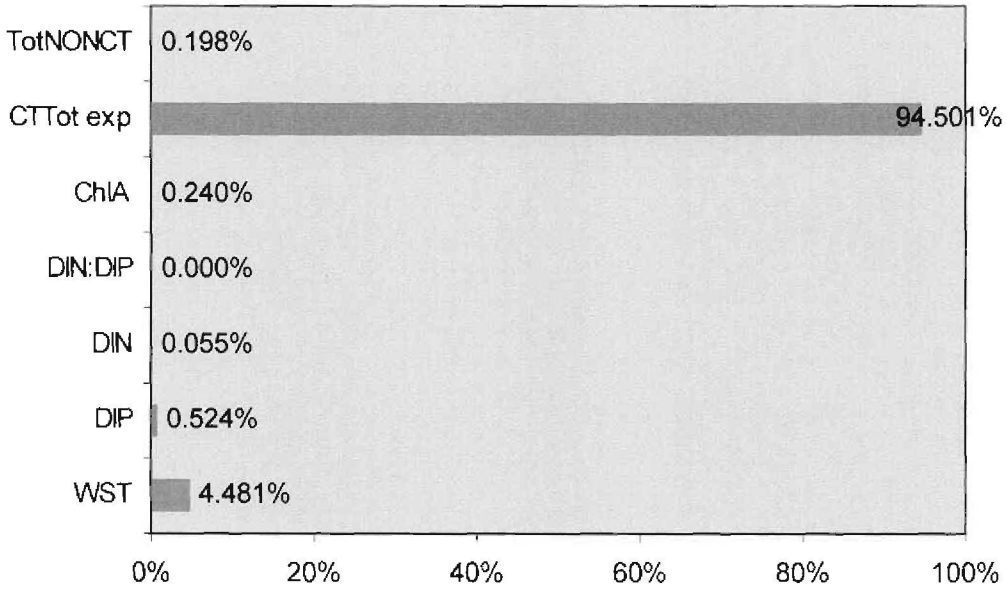


Figure 5.18 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the CT group.

The percentage contribution of the selected environmental and biological components that were used as input for experiment 4 to predict the CT group, are shown in Fig. 5.18. In this experiment

the total CT dominance contributed 94.501% to the outcomes of the results. This is basically using the forecasted variable to predict itself. WST was the second most important contributor to the outcomes with only 4.481% contribution and the DIP contributes only 0.524%. In this instance the DIN:DIP ratio did not contribute at all to the outcome of the model.

Modelling using data from all the dams

With the combining of all the dams' data, the experiments were again repeated as with the single dam data to determine if the outcomes are consistent with the first four experiments.

Experiment 5: Non-cyanotoxin producer (Non-CT) group with selected environment variables as input (all dams)

With Experiment 5, only the measured environmental variables were used as input into the model and the results obtained are shown in Fig. 5.19. The r^2 value was 0.9331 and the correlation was 0.9684. The correlation values indicate very good results and there is a linear relationship between the actual data and the forecasted data.

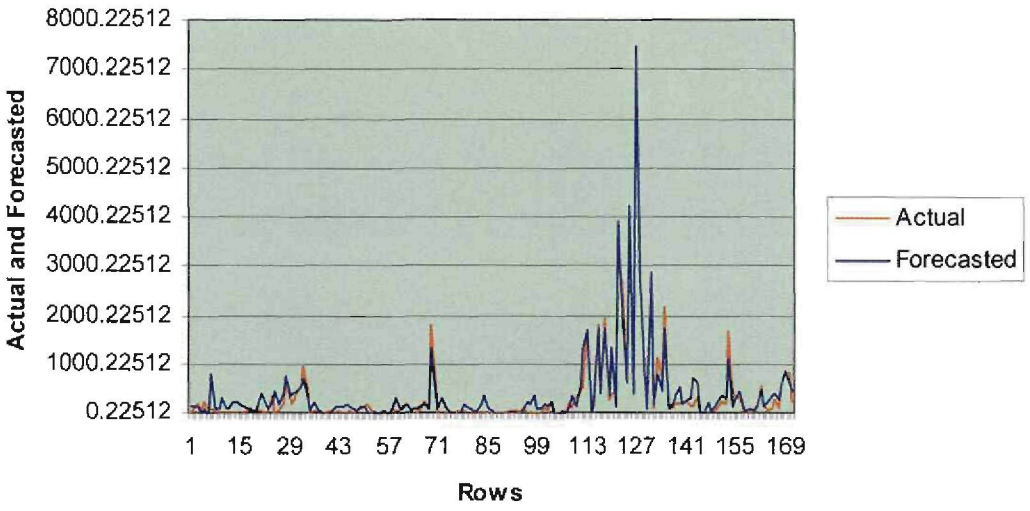


Figure 5.19 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using environmental variables and Chl *a* concentration as inputs for the Forecaster XL ANN Software package.

The Average AE (Table 4.12) is low and indicates that the results indicate good results. However, the number of good results with both the training dataset (13%) and with the testing dataset (17%) is not acceptable.

Table 5.12 A summary of the modelling result output with the Forecaster XL ANN software package for the Non-CT producing group using selected environmental variables d Chl *a* as inputs.

	Training set	**Test set**
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	136.34151	140.71186
Average MSE:	42747.538	54454.783
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	18 (13%)	5 (17%)
# of Bad forecasts:	124 (87%)	24 (83%)

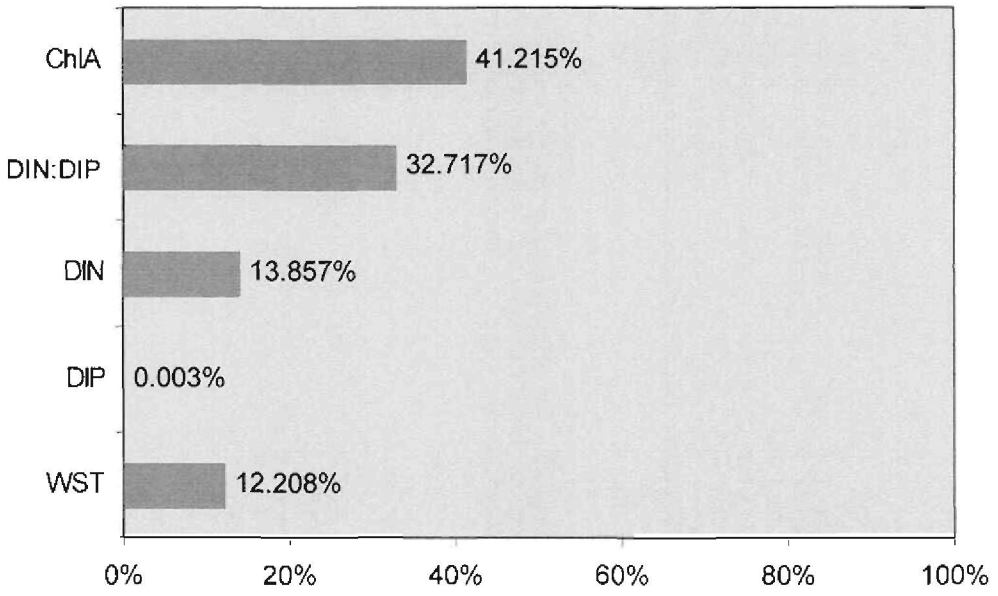


Figure 5.20 Percentage contributions of selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group.

Fig.5.20 shows the percentage contribution of the measured and input variables to determine the Non-CT group's forecasts. As indicated, the main contribution to the outcome is the Chl *a* concentration. Secondly, the DIN:DIP ratio contributed 32.717%. Thirdly, the DIN concentrations contributed 13.857% to the outcomes of the model. Fourthly, the WST contributed 12.208%. These results show that nitrogen may be more important than phosphorus in the forecasting of the Non-CT group of the phytoplankton community.

Experiment 5 (all dams) can be compared to Experiment 1 (single dam). For both experiments, the Chl *a* concentration is the dominant variable that determines the outcome. Secondly, the nutrients are important. However, for the single dam the DIP concentration is more important, while the combination of the data from the different dams shows that DIN is the more important of the two nutrients. For modelling the combined dams, the DIN:DIP ratio is more pronounced than with the single dam.

Experiment 6: Cyanotoxin producer (CT) group with only selected environment variables and Chl *a* concentration as input (all dams)

In Experiment 6 the environmental variables and the Chl *a* concentration were used as input into the model. The actual and forecasted results are shown in Fig. 5.21. The r^2 value was 0.6866 and the correlation was 0.8311, which indicate good correlation between the actual and forecasted data.

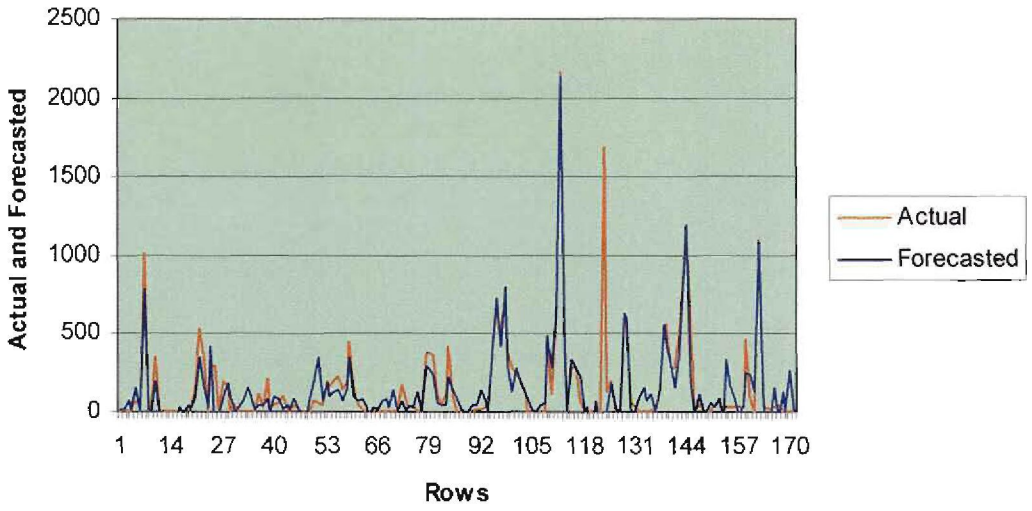


Figure 5.21 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables as inputs for the Forecaster XL ANN Software package.

The results shown in Table 5.13, show that here are larger errors than before, and this makes the model outcome not acceptable. The number of good forecasts for the training dataset is 11% and for the testing dataset was only 17%. Again, one needs to look at the tolerance level and determine whether the model is not too strict regarding the percentage tolerance allowed. It may also indicate that increasing the dataset do not necessarily improve the results.

Table 5.13 A summary of the modelling result output with the Forecaster XL ANN software package for the CT producing group using selected environmental variables and Chl *a* as inputs.

	Training set	Test set
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	66.752897	142.20048
Average MSE:	8735.5117	118584.66
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	15 (11%)	5 (17%)
# of Bad forecasts:	127 (89%)	24 (83%)

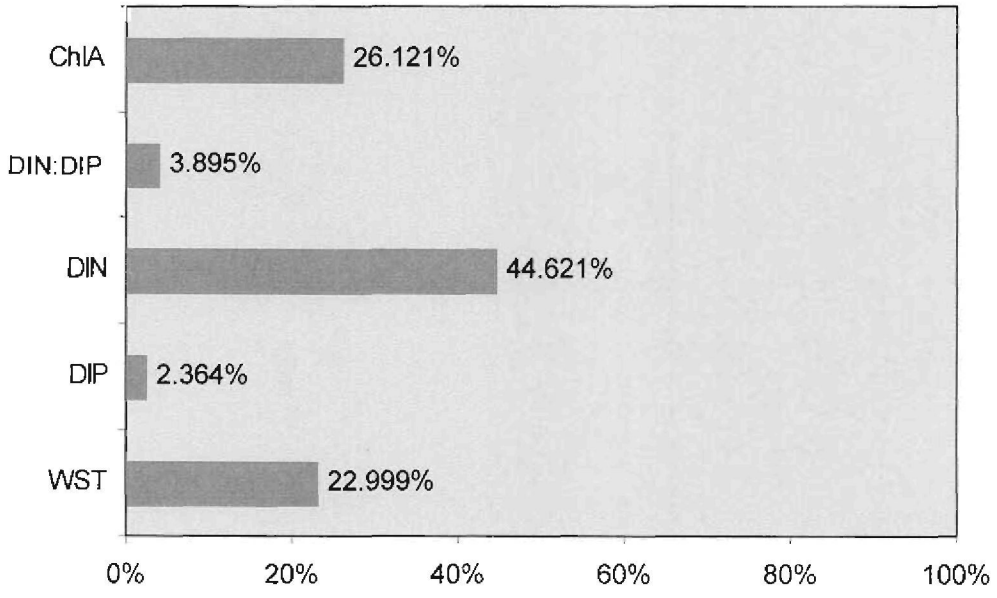


Figure 5.22 Percentage contributions of selected environmental parameters and Chl *a* concentration to ANN prediction for the CT group.

In the outcome of this modelling experiment the most important contributing variables (Fig. 5.22) are the DIN concentrations (44.621%), then the Chl *a* concentration (26.121% and lastly the WST values (22.999%). However, all the input variables were used to determine the result.

When the results of single dam experiment (Experiment 2) are compared to Experiment 6 (all dams), there is no trend in the important variables that determine the outcome of the forecast. For the single dam, the most important variable is temperature that contributes over 60% towards the outcome of the results. The DIP concentration is the second most important variable for the single dam scenario. In Experiment 6 the DIN concentration is the most important environmental variable, contributing more than 40% towards the outcome of the results. The Chl *a* concentration and the DIN concentration is almost equally important, contributing over 20% towards the outcome of the results.

Experiment 7: Non-Cyanotoxin producer (Non-CT) group using algae dominance, selected environment variables and Chl *a* concentration as input (all dams)

In Experiment 7, to forecast the Non-CT group the selected environmental variables, the algal dominance and the Chl *a* concentration were used as the input to the model. The results obtained are shown in Fig. 5.23. The r^2 value was 0.9863 and the correlation was 0.9932. This indicates an extremely good correlation between the actual and the forecasted datasets.

In Table 5.14, there is a summary of the modelling results regarding the extent of the errors, the tolerance type and the number of good and bad forecasts according to the model. Results show that the average actual error is 65 for the training dataset and 86 for the testing dataset. The number of good forecasts in the training set is 18% and for the testing data set 28%. Within the tolerance levels of 10% and 30%, respectively this is quite acceptable.

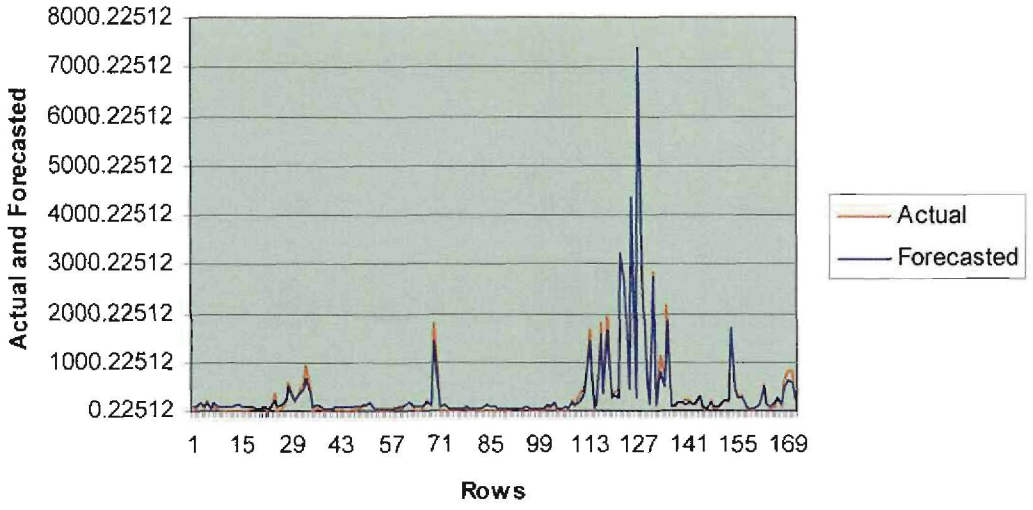


Figure 5.23 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only environmental variables as inputs for the Forecaster XL ANN Software package.

Table 5.14 A summary of the modelling results for the Non-CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs.

	Training set	**Test set**
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	64.584469	85.779019
Average MSE:	8357.8507	13143.261
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	25 (18%)	8 (28%)
# of Bad forecasts:	117 (82%)	21 (72%)

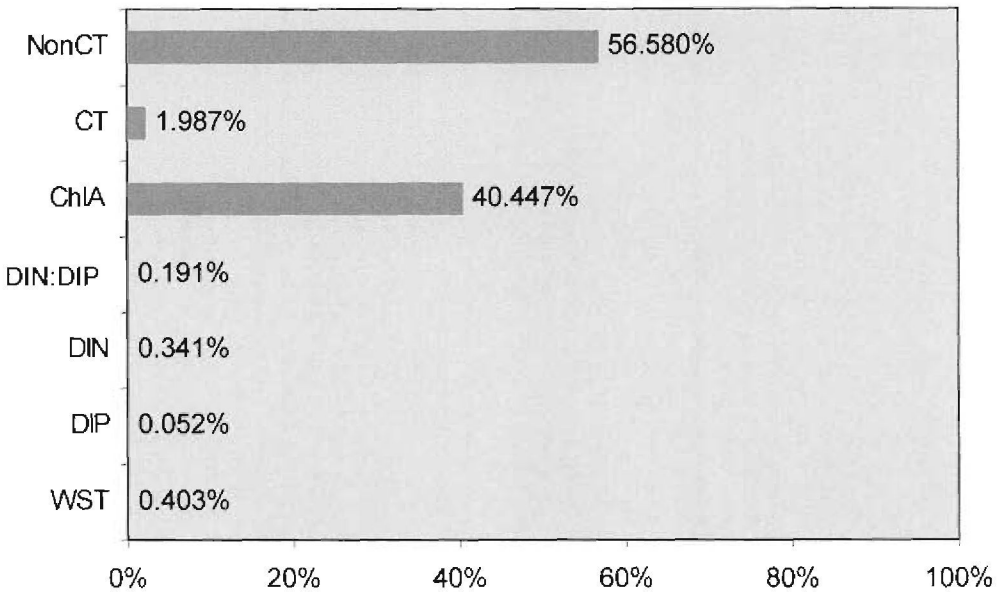


Figure 5.24 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT producing group in all the dams.

Results (Fig. 5.24) show that the environmental parameters that contribute most to the outcomes of the model are the Non-CT counts (56.58% and the Chl *a* concentration (40.447%). The CT algal counts contribute only 1.987% to the outcome. All the other input variables contribute very little to the outcomes of the model and the suggestion is that these variables are not important in establishing the Non-CT group contribution to the phytoplankton community.

Experiment 7 (for all dams) when compared to Experiment 3 (for the single dam) indicates that the Chl *a* concentration and the Non-CT dominance are important in forecasting real time outcomes of both the experiments. This indicates that the use of both these variables may not be applicable for future use.

Experiment 8: Cyanotoxin producer (CT) group algae counts and environment variables as input (all dams)

In Experiment 8 to forecast the CT Group, algae dominance, the selected environmental variables together with the Chl *a* concentration were used as input into the model. The results obtained are shown in Fig. 5.25. The r^2 value was 0.9911 and the correlation was 0.9962. This correlation is extremely good and statistically highly significant.

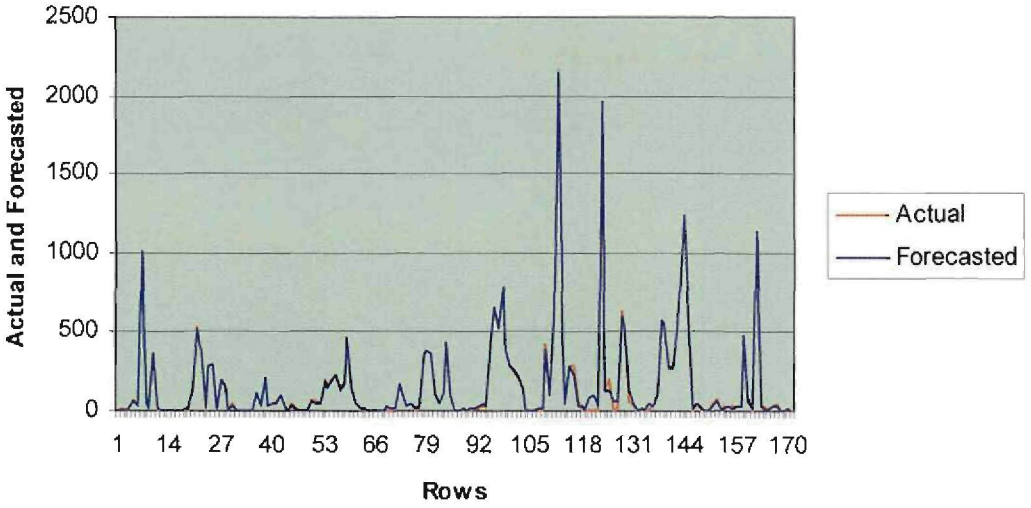


Figure 5.25 Actual (orange line) versus forecasted (blue line) results for the CT producing algae group using environmental variables and algae counts as inputs for the Forecaster XL ANN Software package.

The results of the model as shown in Table 5.15 also show that the model predicts the outcomes very well as the average actual error is very small and the average mean square error (MSE) is lower than any of the experiments before. The number of good forecasts has increased to 38% for the training dataset and to 52% with the testing dataset.

Table 5.15 A summary of the modelling results for CT producing group using algal dominance, selected environmental variables and Chl *a* concentration as inputs.

	Training set	Test set
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	12.951232	8.2859074
Average MSE:	889.96353	239.1688
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	54 (38%)	15 (52%)
# of Bad forecasts:	88 (62%)	14 (48%)

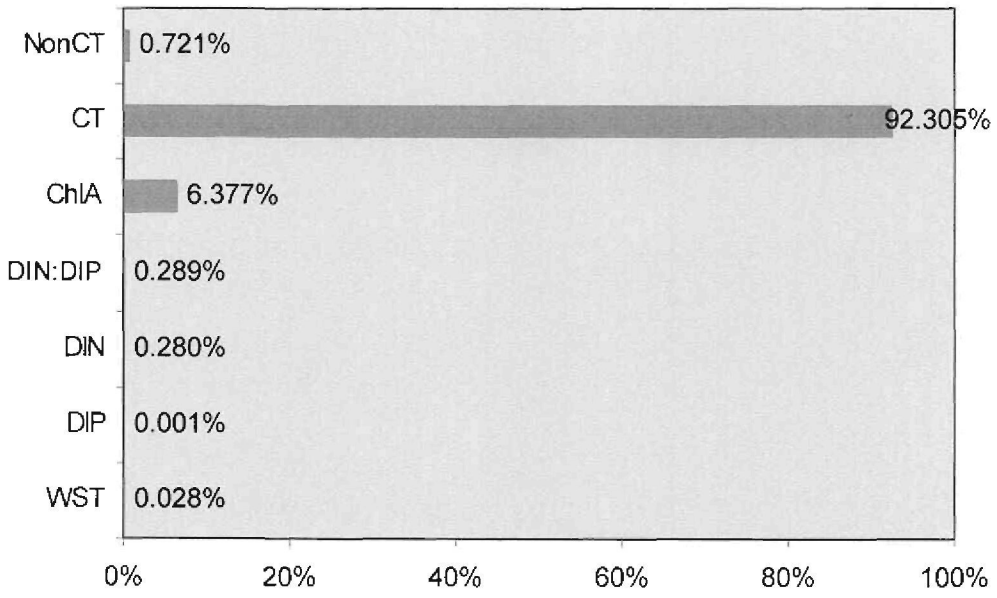


Figure 5.26 Percentage contributions of algal dominance data, selected environmental parameters and Chl *a* concentration as result of the Forecaster XL forecasting of the CT producing group.

In Fig. 5.26 it is shown that the CT group counts contributed 92.305% to the outcomes of the model. The Chl *a* concentrations contributed 6.377% and the non-CT group counts 0.721% to the outcomes. This show that it is basically the CT counts itself that are used to predict the CT biovolume and this is not quite what one would want from a prediction model. Other factors e.g.

the environmental variables need to contribute towards the successful forecasts of the outcome variable.

When Experiment 8 (all dams) is compared to Experiment 4 (single dam), the CT biovolume contribute in both instances over 90% to the outcome of the forecast. In the case of Experiment 4 the WST is an important environmental variable contributing to the results, while in the case of Experiment 8 the Chl *a* concentration is more important.

Experiment 9: Non-Cyanotoxin producer (Non-CT) group environment variables, chlorophyll omitted, as input (all dams)

In Experiment 9 the input variables used were only the measured environmental variables. Both the algae counts, and the Chl *a* concentrations, were omitted from the input variables. The results obtained are shown in Fig. 5.27. The r^2 value was 0.3973 and the correlation was 0.7233. This is not such a good correlation, however, the fact that the Chl *a*, and the algal counts have been left out as input variables makes it more usable as forecasting tool.

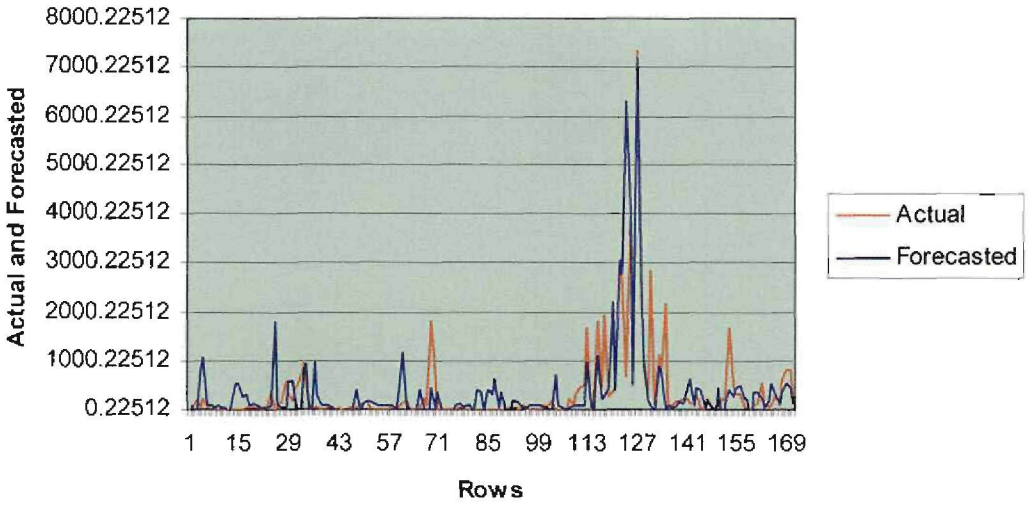


Figure 5.27 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only selected environmental variables as inputs for the Forecaster XL ANN Software package.

The errors and tolerances used in the model are shown in Table 5.16. The training dataset and the testing dataset errors were not as good as with experiment 8, but are still acceptable within the variability of the results. However, the number of good forecasts in the training set is only 6% and for the testing data set only 17%.

Table 5.16 A summary of the modelling results for the Non-CT producing group using only environmental variables as inputs.

	Training set	**Test set**
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	266.95715	484.59321
Average MSE:	238219.96	1281396.3
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	9 (6%)	5 (17%)
# of Bad forecasts:	133 (94%)	24 (83%)

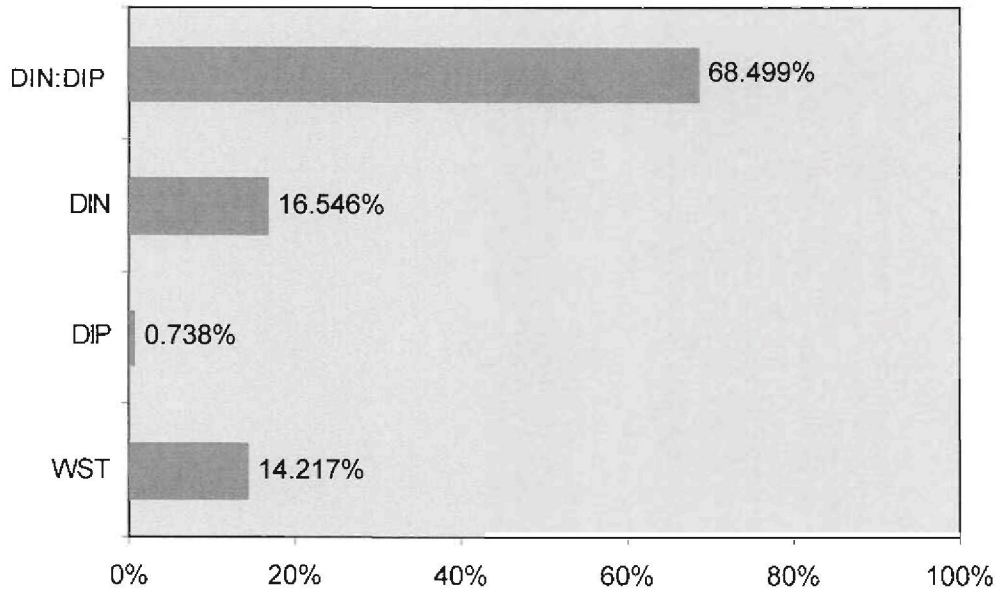


Figure 5.28 Percentage contributions of the selected environmental parameters to Forecaster XL forecasts for the Non-CT group.

In this case, where only the environmental variables were used as input to the model, the DIN:DIP ratio contributes 68.499% to the outcome(non-CT group) of the model (Fig. 5.28). This is contradictory to Chutter & Rossouw (1991) that suggests that the N:P ratio is a driving force for the presence of the CT group. The other environmental variables that contribute to the outcome are the DIN concentrations with a contribution of 16.546% and the WST with a contribution of 14.217%. The DIP concentrations indicate that only 0.738% of the outcome is caused by the contribution of DIP. Therefore, these results show that the driving forces behind the forecasts of the non-CT group are the N:P ration, DIN concentrations and the WST at a given time.

Experiment 10: Cyanotoxin producer (CT) group environment variables, Chl *a* omitted, as input (all dams)

In Experiment 10, the algae counts and the Chl *a*, concentrations are not included as input variables. Only the measured environmental variables are used as input into the model and the results obtained are shown in Figure 5.29. The r^2 value was 0.4681 and the correlation was 0.7018.

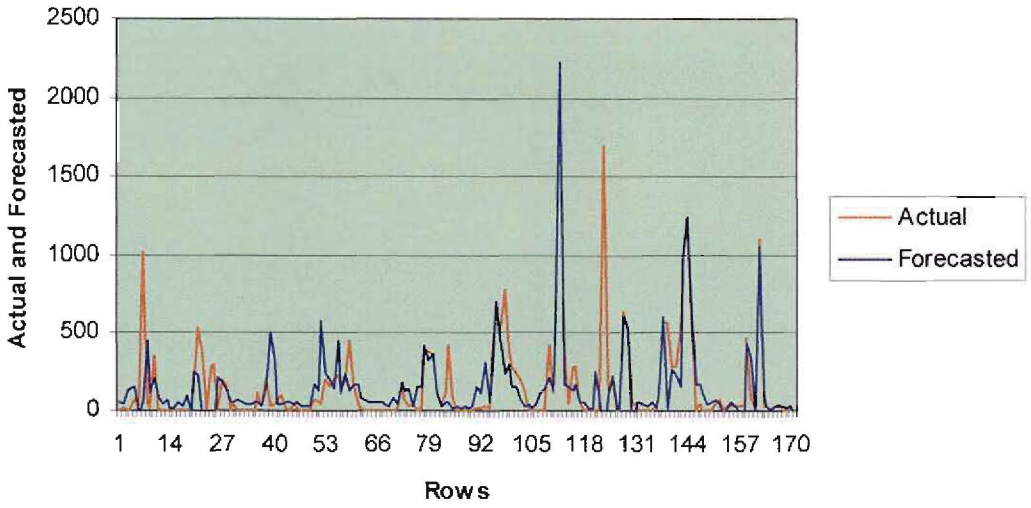


Figure 5.29 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables as inputs for the Forecaster XL ANN Software package.

The errors and the tolerances of the results of the model for Experiment 10 are shown in Table 5.17. The average actual errors (AE) may be within acceptable limits but the average MSE are quite high. The number of good forecasts according to the model is only 12% in the training dataset and 10% for the testing dataset.

Table 5.17 A summary of the modelling results for the CT producing group using only the selected environmental variables as inputs.

	Training set	**Test set**
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	90.393001	171.97215
Average MSE:	41676.174	69798.824
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	17 (12%)	3 (10%)
# of Bad forecasts:	125 (88%)	26 (90%)

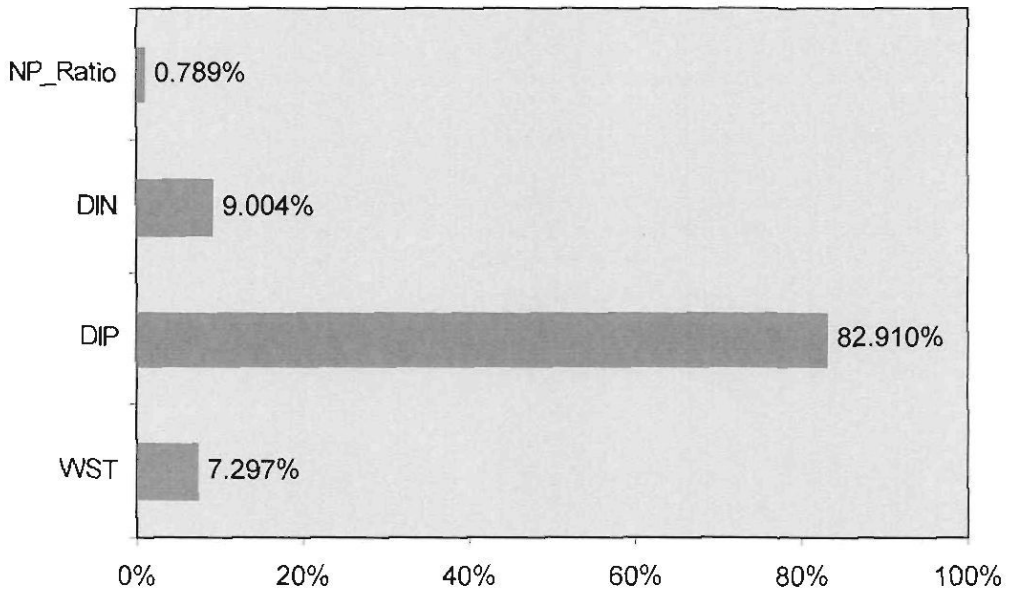


Figure 5.30 Percentage contributions of selected environmental parameters to Forecaster XL forecast for the CT producing group.

In the forecast of the CT group, using only the selected environmental variables as input it is shown in Fig. 5.30 that the DIP concentrations contributed most to the model's outcome. The second most important environmental variable was the DIN concentrations with a contribution of 9.004%. This indicates the importance of the nutrients in forecasting the composition of the algal community. It seems that the DIP is the most important environmental variable in determining the presence of the CT group. The WST is also important in the outcome of the model by predicting the CT Group. With this experiment the N:P ratio was the least important variable in predicting the CT Group.

Experiment 11: Prediction of the Non-cyanotoxin producer (Non-CT) group including environment variables and algae dominance; chlorophyll omitted, as input (all dams)

In Experiment 11 the Chl *a* concentrations are not included as input variables. Only the measured environmental variables and the algae dominance are used as input into the model and the results obtained are shown in Figure 5.31. The r^2 value was 0.9799 and the correlation was 0.9904.

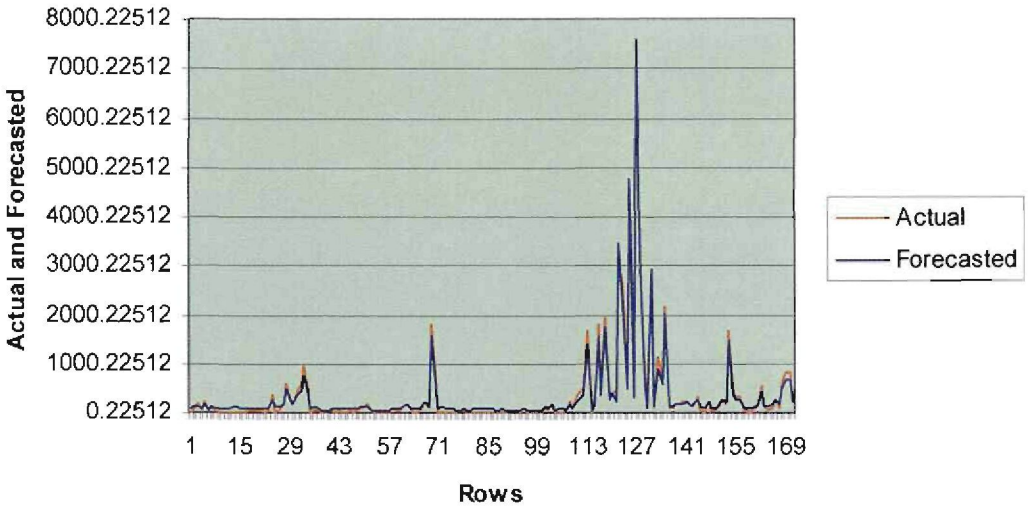


Figure 5.31 Actual (orange line) versus forecasted (blue line) results for the Non-CT producing group using only environmental variables and algal counts as inputs for the Forecaster XL ANN Software package.

The errors and the tolerances of the results of the model for Experiment 11 are shown in Table 5.18. The average actual errors (AE) may be within acceptable limits but the average MSE are quite high. The number of good forecasts according to the model is only 17% in the training dataset and 24% for the testing dataset.

Table 5.18 A summary of the modelling results of the Non-CT producing group using the selected environmental variables, algal dominance and Chl *a* concentration as inputs.

	Training set	Test set
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	79.179105	73.15253
Average MSE:	14788.918	6900.0463
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	24 (17%)	7 (24%)
# of Bad forecasts:	118 (83%)	22 (76%)

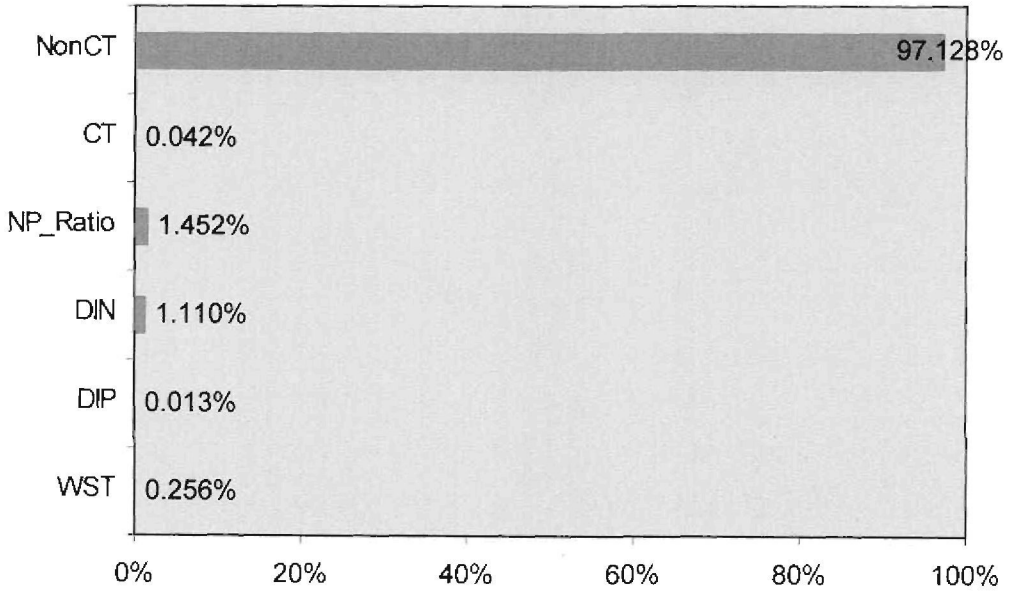


Figure 5.32 Percentage contributions of algal dominance, selected environmental parameters and Chl *a* concentration to ANN prediction for the Non-CT group.

In the forecast of the Non-CT group using only the environmental variables and algal dominance as inputs, it is shown in Fig. 5.32 that the algal dominance contributed most (97.128%) to the model outcome. The second most important environmental variable was the N:P ratio (1.452%)

and thirdly, the DIN concentrations contribute 1.110%. The WST is used in determining the outcome of the model by predicting the Non-CT Group, however the contribution is very small. This indicates that the DIN concentrations may be more important in the development and forecasting of the Non-CT group.

Experiment 12: Prediction of the Cyanotoxin producer (CT) group including the environment variables and algae dominance; chlorophyll omitted, as input (all dams)

In Experiment 12 the Chl *a* concentrations are not included as input variables. Only the measured environmental variables and the algae counts are used as input into the model and the results obtained are shown in Figure 5.33. There is very little distinction between the actual and predicted values. The r^2 value was 0.9811 and the correlation was 0.9911. This is a very good correlation.

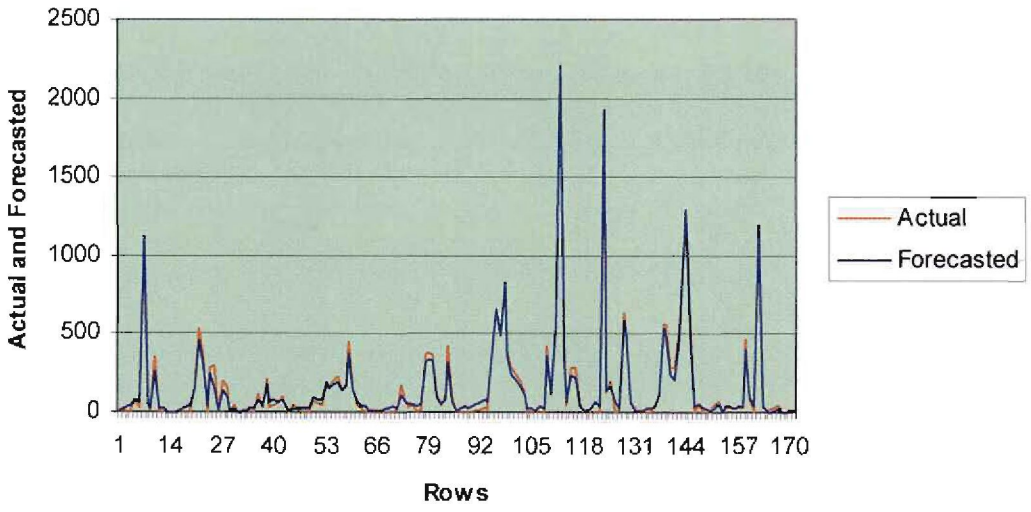


Figure 5.33 Actual (orange line) versus forecasted (blue line) results for the CT producing group using only environmental variables and algal counts as inputs for the Forecaster XL ANN Software package.

The errors and the tolerances of the results of the model for Experiment 12 are shown in Table 5.19. The average actual errors (AE) may and the average MSE are quite high. The number of good forecasts according to the model is 20% in the training dataset and 31% for the testing dataset. The acceptance of a 30% tolerance is quite strict and does not give very acceptable results.

Table 5.19 A summary of the modelling results for the CT producing group using algal dominance and selected environmental variables as inputs.

	Training set	Test set
# of rows:	142	29
CCR:	n/a	n/a
Average AE:	30.124946	32.220362
Average MSE:	1597.2205	1915.5839
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
# of Good forecasts:	28 (20%)	9 (31%)
# of Bad forecasts:	114 (80%)	20 (69%)

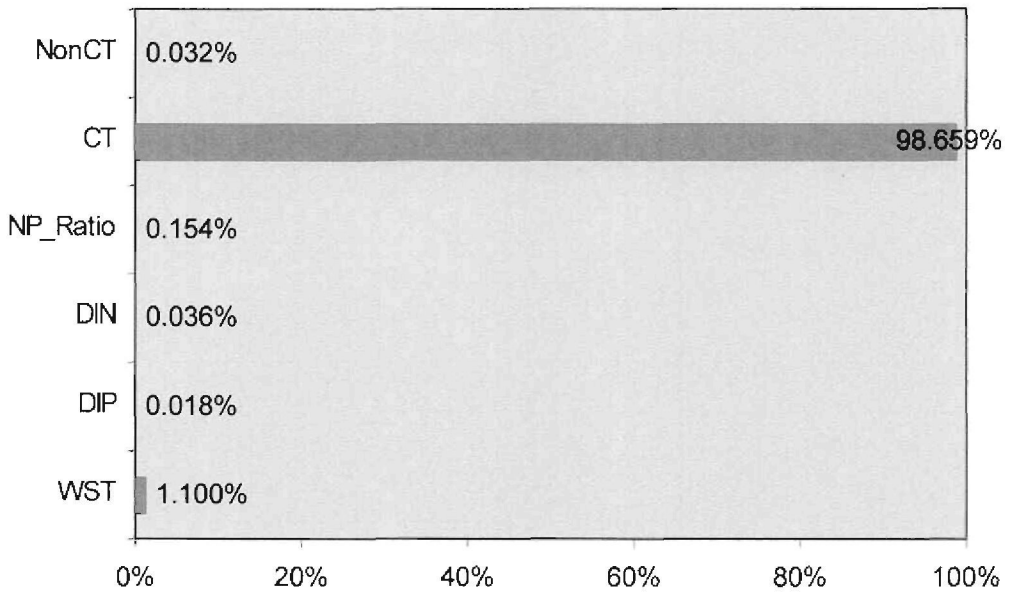


Figure 5.34 Percentage contributions of the algal dominance and the selected environmental parameters to ANN prediction for the CT group.

In the forecast of the Cyanotoxin group using the environmental variables and algal dominance as inputs it is shown in Fig. 5.34 that the algal dominance contributed most (98.659%) to the model outcome. The second most important environmental variable is the WST (1.100%) and thirdly the

N:P ratio contributed 0.154%. The DIN and DIP concentrations do not contribute significantly to the outcome of this experiment.

The results achieved with the MFNN (Forecaster XL) look promising, and give an output that shows the successes and failures of the results. The percentage tolerance of the modelling package with the training set is 10% and with the testing set is 30%. The presentation of the most important factors in determining the output of the model, does give a researcher insight into the driving environmental variables that were fed into the model.

The modelling results did not show very high successes, however, the user friendliness of the program and the ease with which it can be used within the Microsoft background make it a good choice for further use and testing. The one main problem of the program is that it does not supply the user with the developed algorithms used within the model.

5.2.3.1.3 Comparison of the Forecaster XL experiments

A comparison of the 12 experiments with Forecaster XL is shown in Table 5.20. In forecasting the CT group, the experiments show that whenever either the algal dominance and/or the Chl *a* concentration is omitted as input variable to Forecaster XL, WST, DIP and DIN is the most important variables in determining the outcome of the model. Only in one experiment (Exp 9) did the DIN:DIP ratio show up as an important variable in determining the CT biovolume.

Whenever Chl *a* concentration is included as an input variable and the algal dominance is omitted in the model, it usually came up as the most important variable in predicting the outcome of the Non-CT group. This indicates that the extent of the Chl *a* is important in forecasting new biomass, specifically, for the Non-CT producing group. Chl *a* seem to be not so important in the forecasting of the CT group. There is no consistency regarding this group (Table 5:20).

If algal dominance is used as input variable, it dominates the contribution of the input variables. This is using the forecasted variable to predict the biovolume, and it should not be used in this type of modelling exercise.

Table 5.20 Comparison of the experiments done with the MFF ANN modeling results of ForecasterXL showing the average actual errors, the good and bad forecasts and the contribution of the environmental, biomass and Chl *a* towards the output of each experiment.

Experiment	Average Actual Error	Good Forecasts (%)	Bad Forecasts (%)	Contribution of Environmental Variables to modelling Results (%)						
				Non-CT	CT	Chl *a*	DIN:DIP	DIN	DIP	WST
Single Dam										
Experiment 1: Non-CT, Chl *a* & Environmental Variables				-	-	83.28	1.935	0.676	9.257	4.850
- Testing Data Set	46.792	17	83							
- Training Data Set	62.012	0	100							
Experiment 2: CT, Chl *a* & Environmental Variables				-	-	5.589	8.303	7.028	18.171	60.910
- Testing Data Set	23.212	7	93							
- Training Data Set	102.333	50	50							
Experiment 3: Non-CT, Environmental Variables, Algal dominance & Chl *a*				16.188	4.643	76.822	0.084	0.285	0.389	1.588
- Testing Data Set	13.922	23	77							
- Training Data Set	23.618	33	67							
Experiment 4: CT, Environmental Variables, Algal dominance & Chl *a*				0.198	94.501	0.240	0.000	0.055	0.524	4.481
- Testing Data Set	6.562	37	17							
- Training Data Set	17.712	63	83							

Experiment	Average Actual Error	Good Forecasts (%)	Bad Forecasts (%)	Contribution of Environmental Variables to modelling Results (%)						
				Non-CT	CT	Chl *a*	DIN:DIP	DIN	DIP	WST
All dams										
Experiment 5: Non-CT, Environmental Variables & Chl *a*				-	-	41.215	32.717	13.857	0.003	12.208
- Testing Data Set	136.342	13	87							
- Training Data Set	140.712	17	83							
Experiment 6: CT, Environmental Variables & Chl *a*				-	-	26.121	3.895	44.621	2.364	22.999
- Testing Data Set	66.753	11	89							
- Training Data Set	142.200	17	83							
Experiment 7: Non-CT, Environmental Variables, Algal dominance & Chl *a*				56.58	1.987	40.447	0.191	0.341	0.052	0.403
- Testing Data Set	64.584	18	82							
- Training Data Set	85.779	28	72							
Experiment 8: CT, Environmental Variables, Algal dominance & Chl *a*				0.721	92.305	6.377	0.289	0.280	0.001	0.028
- Testing Data Set	12.952	38	62							
- Training Data Set	8.286	52	48							
Experiment 9: Non-CT & Environmental Variables				-	-	-	68.499	16.549	0.738	14.217
- Testing Data Set	266.957	6	94							
- Training Data Set	484.593	17	83							

Experiment	Average Actual Error	Good Forecasts (%)	Bad Forecasts (%)	Contribution of Environmental Variables to modelling Results (%)						
				Non-CT	CT	Chl *a*	DIN:DIP	DIN	DIP	WST
All Dams										
Experiment 10: CT & Environmental Variables				-	-	-	0.789	9.004	82.910	7.297
- Testing Data Set	90.393	12	88							
- Training Data Set	171.97	10	90							
Experiment 11: Non-CT & Environmental Variables & Algal Dominance				97.128	0.024	-	1.452	1.110	0.013	0.256
- Testing Data Set	79.179	17	83							
- Training Data Set	73.153	24	76							
Experiment 12: CT & Environmental Variables & Algal dominance				0.032	98.659	-	0.154	0.036	0.018	1.100
- Testing Data Set	30.125	20	80							
- Training Data Set	32.220	31	69							

5.2.3.2 Self-Organising Map (SOM)

The primary application for the SOM is clustering and data segmentation. Non-supervised neural networks learn to ordinate, cluster or group the data patterns only by inspecting the similarities between the inputs (Smith 2002; Recknagel *et al.* 2006). Clustering involves classifying or ordinating the data into groups based upon the natural structure of the data, rather than known pre-defined classifications. The data from the Roodeplaat Dam was divided into two groups (before 1991 and after 1991) to represent the before and after Zeekoegat Waste Water Treatment Works (WCW) conditions. The WCW discharge directly into the Roodeplaat Dam since 1991 and the impact of the establishment of a WCW on a reservoir can be expressed in the ordination and clustering mapping of the unsupervised SOM.

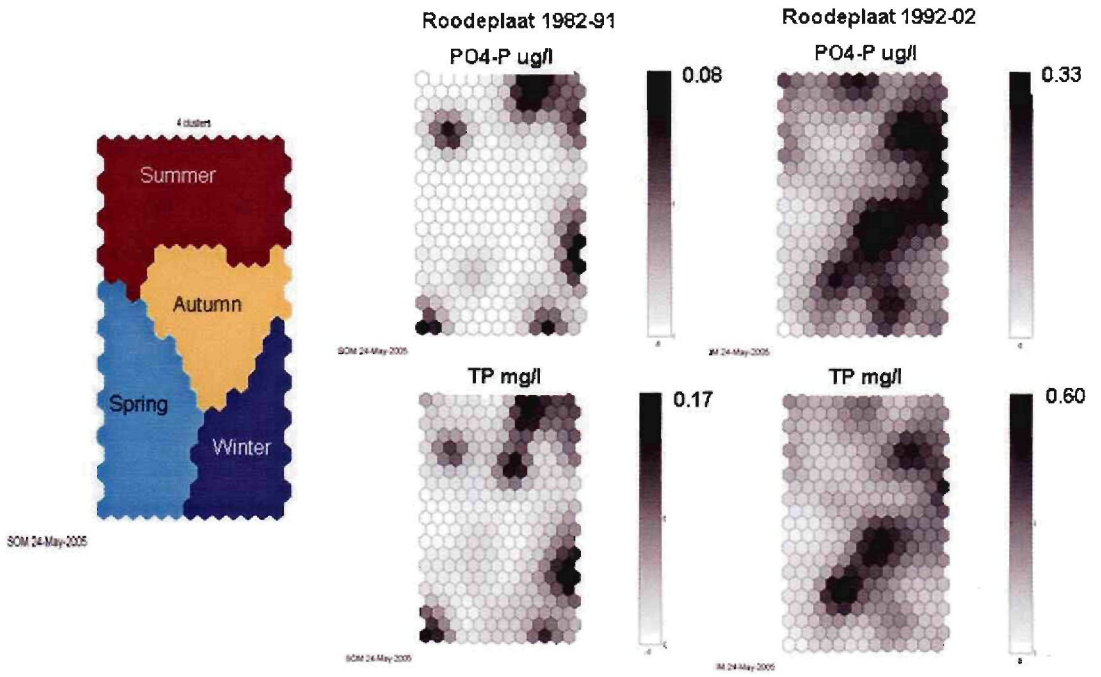


Figure 5.35 The ordination and clustering map of the PO_4-P (DIP) and TP concentrations in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WCW that show the changes in the water quality.

In Fig 5.35, it is shown that there was a large increase in the maximum concentrations of both the PO_4-P and TP in the Roodeplaat Reservoir in response to the impact of the Zeekoegat WCW. The maximum PO_4-P concentration increased from 0.08mg/L to 0.33mg/L. The maximum TP concentration increased form 0.17mg/L to 0.60mg/L. The seasonal occurrence of the higher PO_4-

P concentrations shifted from small peaks in all seasons, but primarily in summer to extreme high concentrations throughout the year. The changes in TP concentration showed similar patterns. The impact of the WCW was severe and the Roodeplaat Dam has since been shifted into hypertrophic conditions, as can be seen from this data as well.

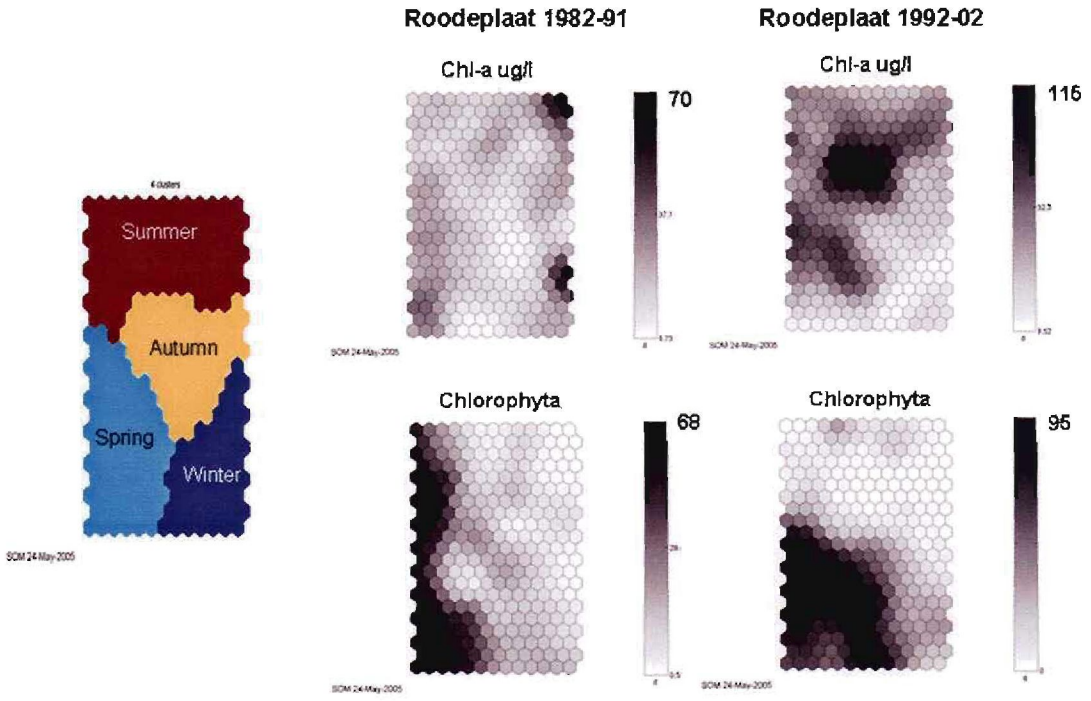


Figure 5.36 The ordination and clustering map of the Chl *a* compared to the Chlorophyta dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WCW that show the changes in the water quality.

The modelling results also show an increase in the Chl *a* concentration between the before and after Zeekoegat WCW establishment. Chl *a* concentrations, increased from a maximum of 70ug/L to a maximum of 115ug/L (Fig 5.36). The occurrence of maximum Chl *a*, concentrations shifted from small peaks in the summer and the winter to large peaks during summer and smaller peaks throughout the year. The dominance of the Chlorophyta (green algae) shifted form being dominant during spring and summer before the establishment of the WCW to being dominant primarily during spring. The dominance of the Chlorophyta during spring was replaced by dominance of cyanobacteria (Cyanophyta) (Fig. 5.37). Before the WCW was established, cyanobacteria were prominent throughout the year but the dominant period shifted to primarily

summer through autumn with much higher biovolume overall as shown by the Chl *a* concentrations.

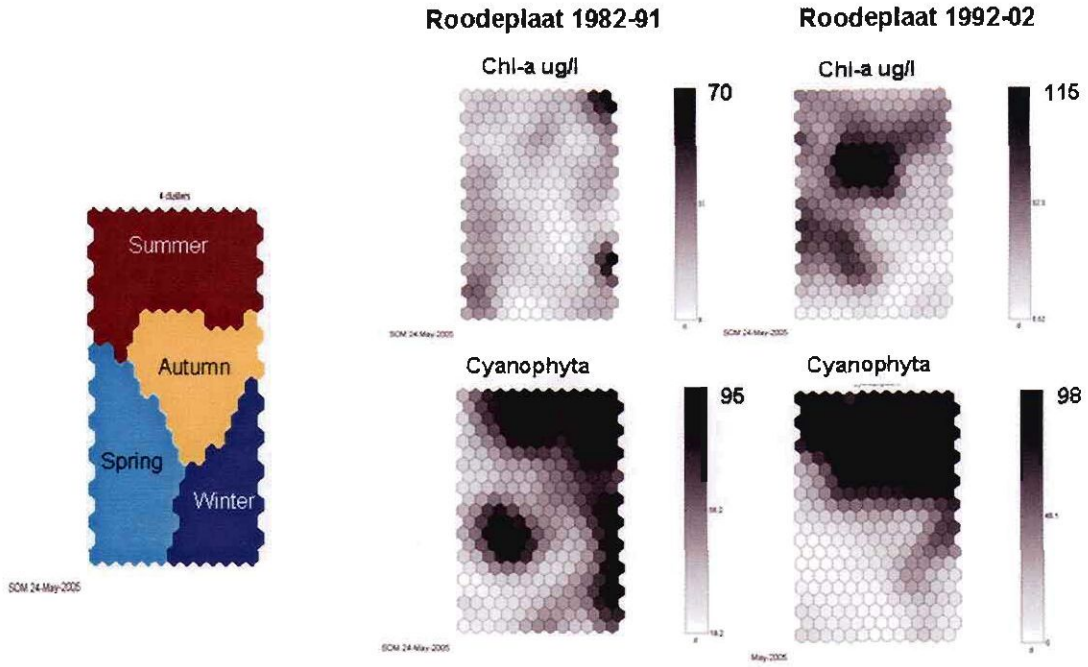


Figure 5.37 The ordination and clustering map of the Chl *a* compared to the Cyanophyta dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WCW that show the changes in the water quality.

The ordination and clustering map as output of the SOM method showed an increase in the maximum dominance of the diatoms from 41% to 52%. The diatoms have been and still are limited to be dominant primarily in the winter periods when biovolume is low (Fig. 5.38). The impact of the WCW on the diatoms was therefore not so severe.

The SOM modelling method showed potential for further use in South Africa as it can be used to potentially:

- Investigate, before and after situations,
- To determine the potential effects of management actions,
- Or changes in the environmental conditions.

This method is not a prediction model but can assist in the search for answers in cause and effect investigations.

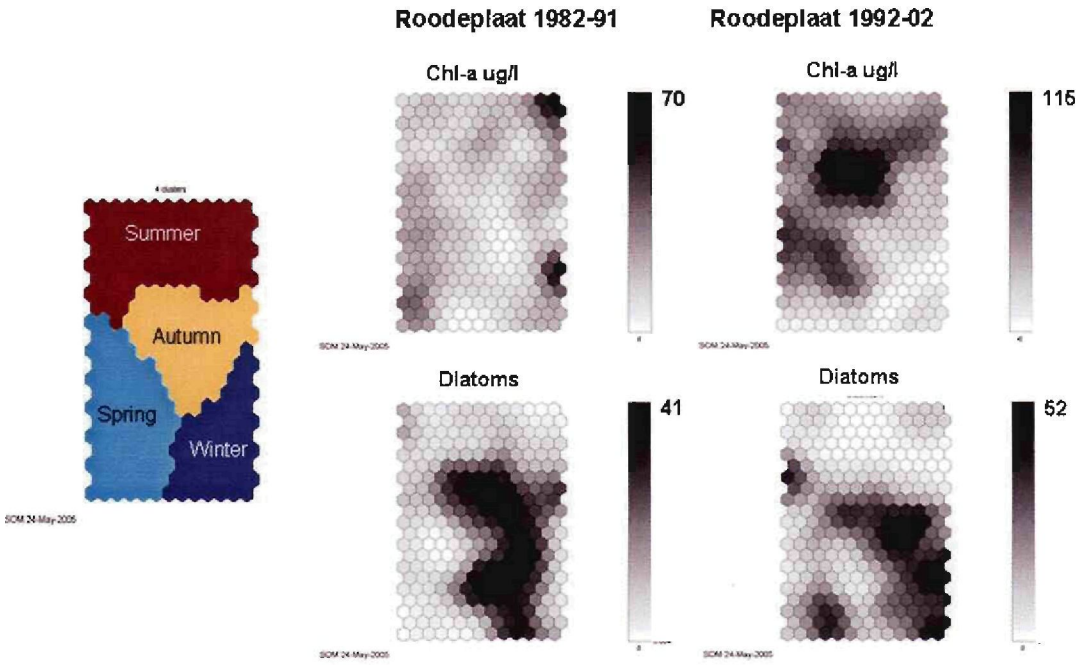


Figure 5.38 The ordination and clustering map of the Chl *a* compared to the diatom dominance in the Roodeplaat Reservoir for the before and after construction of Zeekoegat WCW that show the changes in the water quality.

5.2.4 Rule Set Development by Hybrid Evolutionary Algorithms

Previous studies on algal communities have demonstrated that highly complex ecological time series can be successfully probed to develop rule sets as prediction tools, by using hybrid evolutionary algorithms (HEA) (Cao *et al.* 2005). The HEA is an adaptive method that mimics ecological evolutionary processes, natural selection and genetic variations. A number of experiments were done to determine the most applicable and usable rule set in the prediction of *Microcystis* and the dinoflagellate, *Ceratium*.

Each experiment will be discussed separately, showing the best-developed rule set for each experiment. Selected outcomes of each experiment are shown in Appendix A. All the results from the HEA experiments are available on the CD attached.

5.2.4.1 HEA Experiment 1: Real-time prediction simulation and rule-set discovery for *Microcystis* including Chl *a*, as an environmental factor.

The 36 years of real-time daily data (See Section 5.1.4) of the Hartbeespoort, Rietvlei and Roodeplaat reservoirs was used as training data to develop a rule set for the prediction of *Microcystis* biovolume as shown in Figure 5.39.

The correlation coefficient (r^2 = 0.88) and the root mean square error (RMSE = 9.3357) for the training data and the correlation coefficient (r^2 = 0.73) and the root mean square error (RMSE = 7.4620) for the tested data indicate that the predictions are quite good. The model manages to predict the peaks and lows as well, as shown in Fig.5.35 and Fig. 5.36. The predictions for the Hartbeespoort and the Roodeplaat Dam were much better than for the Rietvlei Reservoir.

More validation was done by determining the Average Actual Error (A AE) and the mean square error (MSE) as shown in Table 5.21. To further test the results of the experiment and to determine the impurities between the actual data and the test data set (Greenberg 2003), the relative tolerance of the training set and of the testing data set was determined and is shown in Table 5.21.

The results show that with a set relative tolerance of 10% in the training set and with a relative tolerance of 30% for the testing data set, there are respectively 38.5% and 44.8% good predictions

61.5% and 55.2% bad predictions. This is a very strict relative tolerance when compared to the up to 300% that is allowed in the Vollenweider Model (IETC 2000).

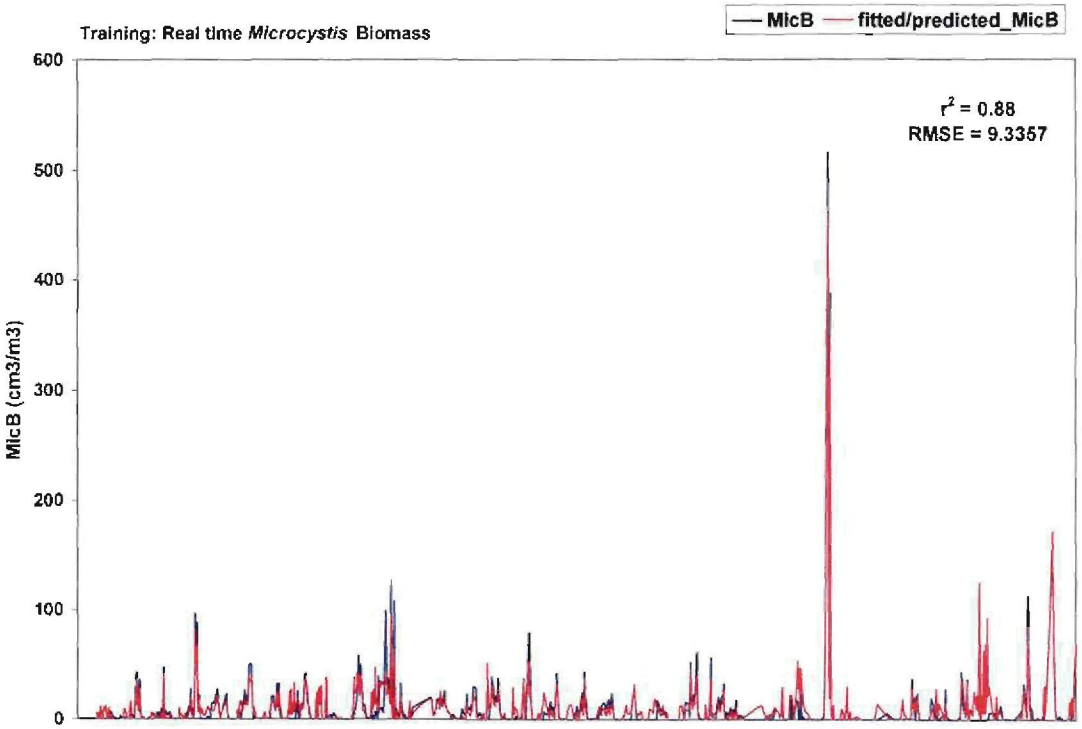


Figure 5.39 Training results of the real-time forecasting of *Microcystis* biovolume including Chl *a*, as an environmental variable in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the rule set for HEA Experiment 1.

Fig. 5.40 shows the results of the tested data and it is clear that the rule set fits the Hartbeespoort and Roodeplaat reservoirs data better than the Rietvlei data, where under and over predictions are found. This may be due to the fact that Rietvlei also experienced dinoflagellate blooms during the study period. However, the magnitudes of the maximum *Microcystis* biovolume peaks are predicted quite well in all three of the reservoirs.

Table 5.21 Summary of the average actual error (AE), the mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume including Chl *a* as an environmental variable in real time rule set development.

	Training set	Test set
# of rows:	13148	2193
Average AE:	0.10422	0.974735
MSE:	8.414296	7.063562
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	38.5%	44.8%
Bad predictions:	61.5%	55.2%

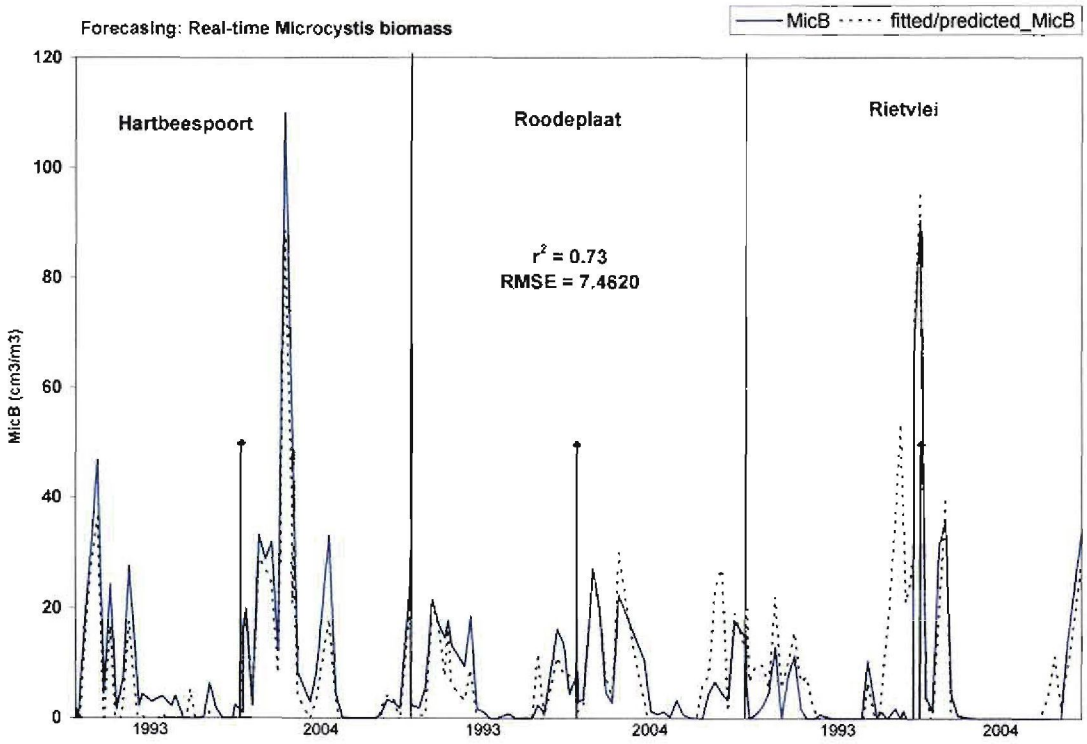


Figure 5.40 Testing results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the rule set as shown in Figure 5.41.

The conditions for the developed rule set (Fig. 5.41) include TP and Chl *a* concentrations that are used to determine if the THEN branch or the ELSE branch of the rule set is to be used for forecasting *Microcystis* biovolume. The sensitivity analysis that was done on the real-time testing data for the three reservoirs to determine the *Microcystis* biovolume show that with the THEN branch of the rule set surface temperature and total nitrogen are the variables to which the determination of the *Microcystis* biovolume is most sensitive to. Low TN concentrations and high surface temperatures are associated with high *Microcystis* biovolume. This suggests that the *Microcystis* biovolume decreases with an increase in the TN concentrations. The Tsurf has the expected opposite reaction. The higher the temperature rise, the larger the *Microcystis* biovolume grow.

The ELSE branch of the rule set is used when Chl *a* concentrations are below 120.519 ug/L and total phosphorous concentrations are higher than 227.982 ug/L. In the ELSE branch of the rule set Chl *a*, is the variable that the *Microcystis* biovolume is most sensitive to. Both pH and Tsurf are also used in the developed RULE set to predict the *Microcystis* biovolume but is not so important in determining changes in the *Microcystis* biovolume under these high total phosphorous conditions.

These results indicate that under phosphorus concentrations smaller than 431.729 ug/L (expressed as TP) the increases in *Microcystis* biovolume are driven negatively by high nitrogen concentrations and positively by the surface temperature (Tsurf).

Under the ELSE rule set, where TP concentrations exceed 431.729 ug/L the *Microcystis* biovolume that develops is primarily driven by the initial biomass (as indicated by the Chl *a* concentration). Temperature is still also impacting on the developing *Microcystis* biovolume, but is not as important as the Chl *a*.

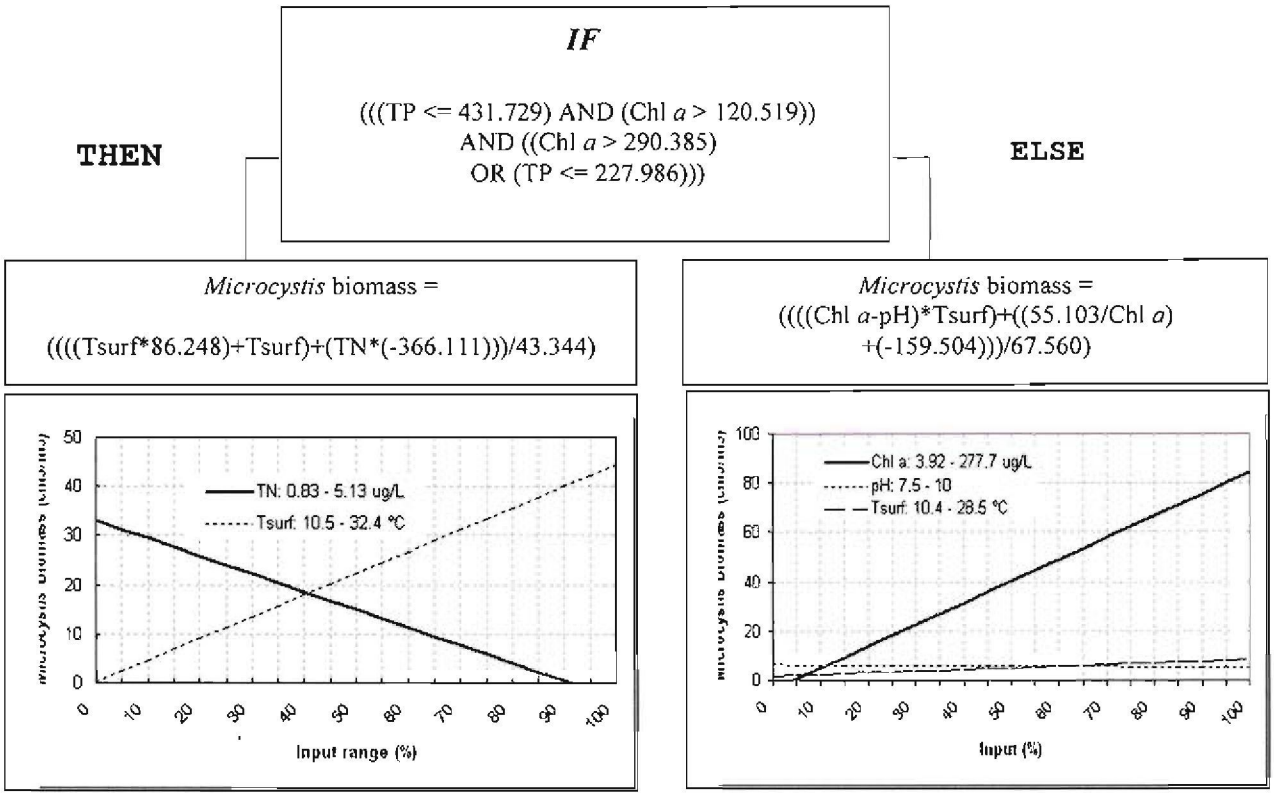


Figure 5.41 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for real-time forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

5.2.4.2 HEA Experiment 2: Real-time prediction simulation and rule-set discovery for *Microcystis* excluding Chl *a*, as an environmental variable.

In the set-up of this experiment, no Chl *a* values, were included as an environmental variable. In Fig. 5.42 it is shown that, the predicted results are not as good as the previous experiment (HEA Experiment 1), therefore, for all subsequent experiments Chl *a* was included. A number of the peaks were not predicted correctly. However, the correlation coefficient (r^2 = 0.81) and the root mean square error (RMSE = 11.7002) for the training data and the correlation coefficient (r^2 = 0.46) and the root mean square error (RMSE = 10.3676) for the tested data indicate that the predictions are still quite good. The model does manage to predict the peaks and lows quite accurately as shown in Fig. 5.42 and Fig. 5.43. Again, the predictions for the Hartbeespoort and the Roodeplaat Dam were better than for the Rietvlei Reservoir.

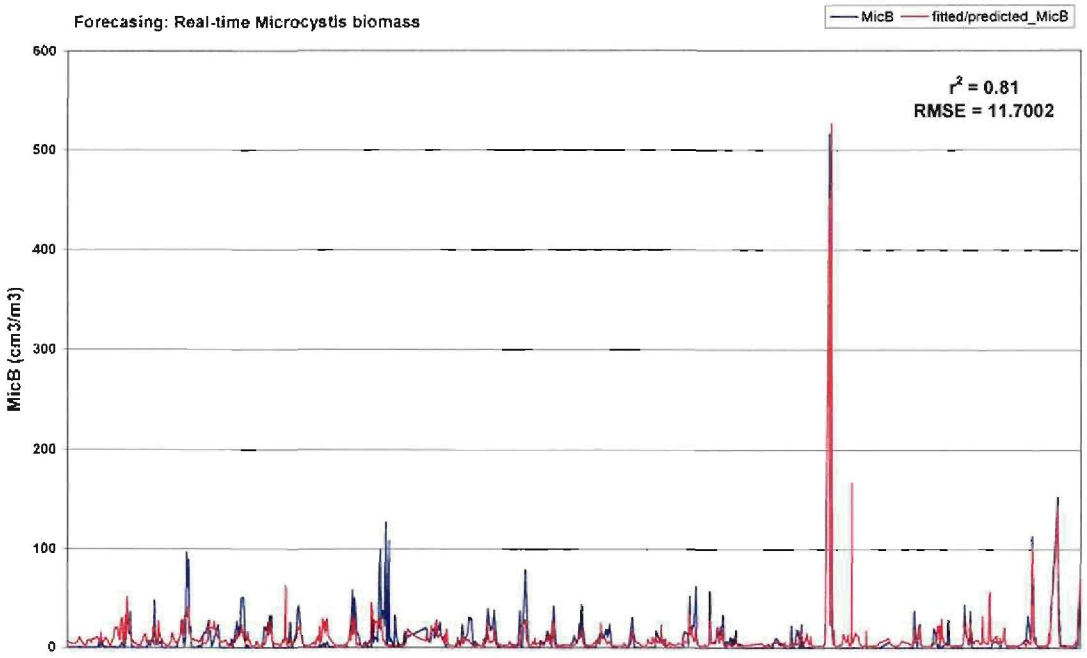


Figure 5.42 Training results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 2 that excluded Chl *a*, as an environmental factor.

Further validation was done by determining the Average Actual Error (A AE) and the average square error (MSE) as shown in Table 5.22. The impurities between the actual data and the test

data set (Greenberg 2003) was determined by the relative tolerance of the training set, and of the testing data set, and is shown in Table 5.22.

The results show that with a set relative tolerance of 10% in the training set and with a relative tolerance of 30% for the testing data set, there are respectively 42.3% and 62.9% good predictions and respectively 57.7% and 37.1% bad predictions. Although this is a very strict relative tolerance, these results are better than the previous experiment, and the number of acceptable predictions is quite good.

Table 5.22 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume real time rule set development excluding Chl *a* as an environmental variable.

	Training set	**Test set**
# of rows:	13148	2193
Average AE:	0.342782	1.630976
MSE:	10.52243	7.503751
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	42.3%	62.9%
Bad predictions:	57.7%	37.1%

A number of times the predictions were under predicting. This was especially the case in the predictions for Rietvlei. The Rietvlei Reservoir is known to lie within a colder area when compared to the Hartbeespoort and the Roodeplaat reservoirs.

The conditions for the developed rule set (Fig. 5.44) in the absence of Chl *a* concentrations include the TN and Secchi depth as conditional parameters that are used to determine if the THEN branch or the ELSE branch of the rule set is to be used for forecasting *Microcystis* biovolume. For the rule set to use the THEN branch the TN concentration needs to be greater than 0mg/L and smaller than 7.77mg/L. The Secchi depth to comply with the condition set for the THEN branch should be less than 0.7m. Otherwise, the ELSE branch applies to determining the *Microcystis* biovolume.

The sensitivity analysis that was done on the real-time testing data for the three reservoirs to determine the *Microcystis* biovolume show that with the THEN branch of the rule set, total phosphorous (TP), total nitrogen (TN), pH and surface temperature (Tsurf) are the variables to which the outcome of the *Microcystis* biovolume is most sensitive to. Within this rule set TP has the highest impact, TN the second highest and Tsurf has a smaller impact on the extent of *Microcystis* biovolume development. The pH, although used in the rule set does not seem to play a significant role in determining the outcome. This phenomenon may be because of the small changes in pH in general. This suggests that the *Microcystis* biovolume increases with increase in the TP, TN and Tsurf.

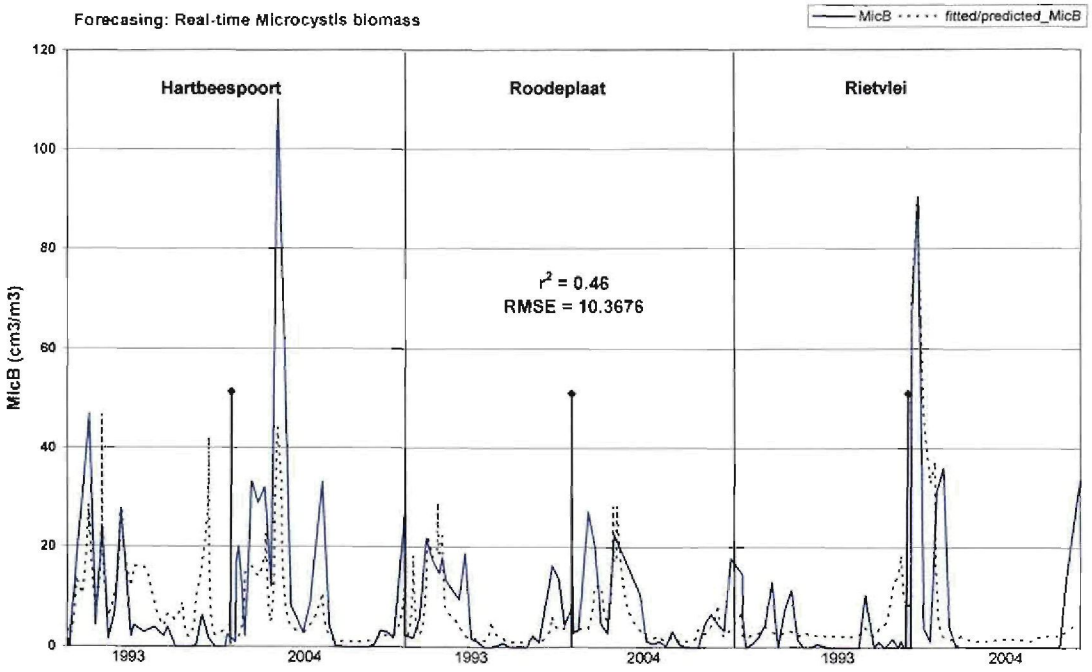


Figure 5.43 Testing results of the real-time forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.44, excluding Chl *a*, as an environmental factor.

In the ELSE branch of the rule set, the Secchi depth and the pH are the variables that determine the outcome of the *Microcystis* biovolume. The Secchi depth is the most important variable that negatively impacts on the prediction of the *Microcystis* biovolume. Under low light conditions (shallow Secchi depths), there is a strong decrease in the *Microcystis* biovolume. Again, the pH seems to have no significant impact on the *Microcystis* biovolume outcome when the input ranges increase stepwise.

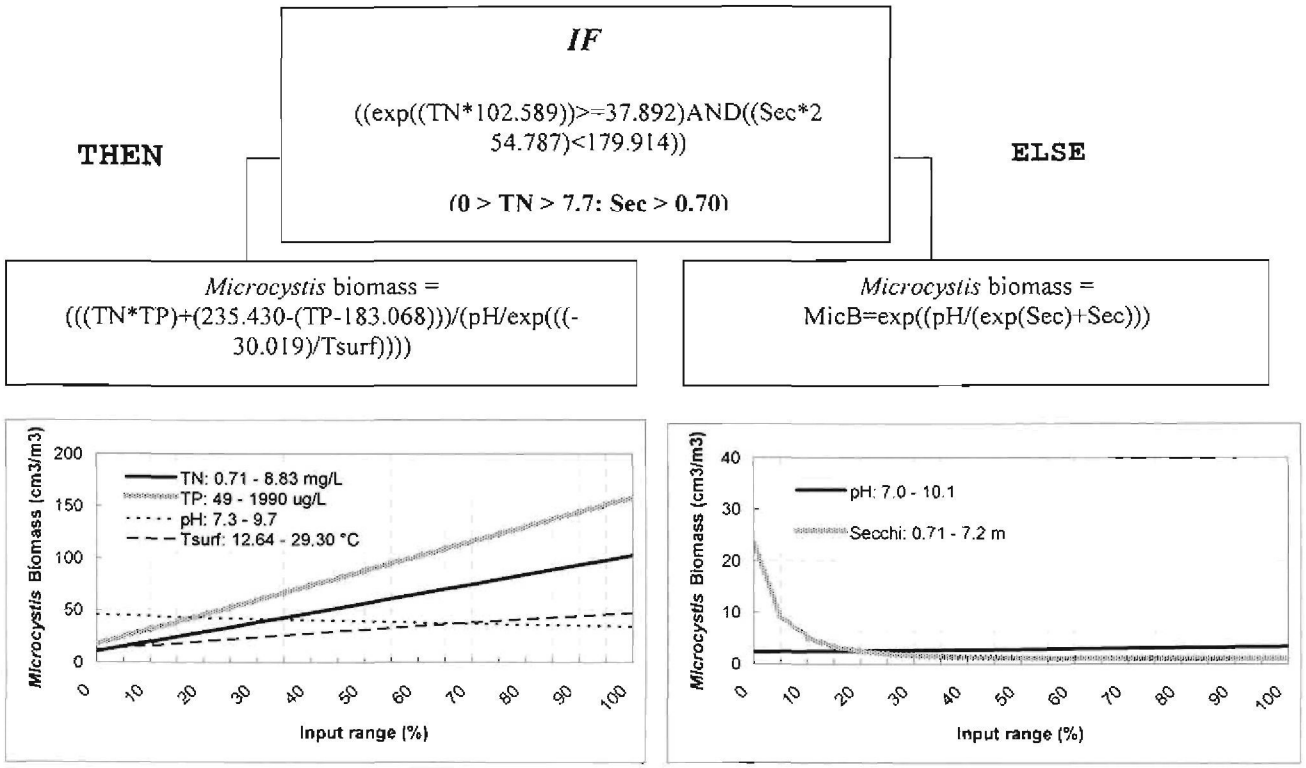


Figure 5.44 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for real-time forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables (excluding Chl *a*) to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

5.2.4.3 HEA Experiment 3: 7-Days forward prediction simulation and rule set discovery for *Microcystis*

To develop a rule set for the 7-days forward prediction of *Microcystis* biovolume the daily interpolated data of the 36 years data set was used for training on the Hartbeespoort, Rietvlei and Roodeplaat reservoir. The *Microcystis* biovolume measured data was shifted backwards to create the condition of 7-days forward training. The training results from this experiment are shown in Figure 5.45.

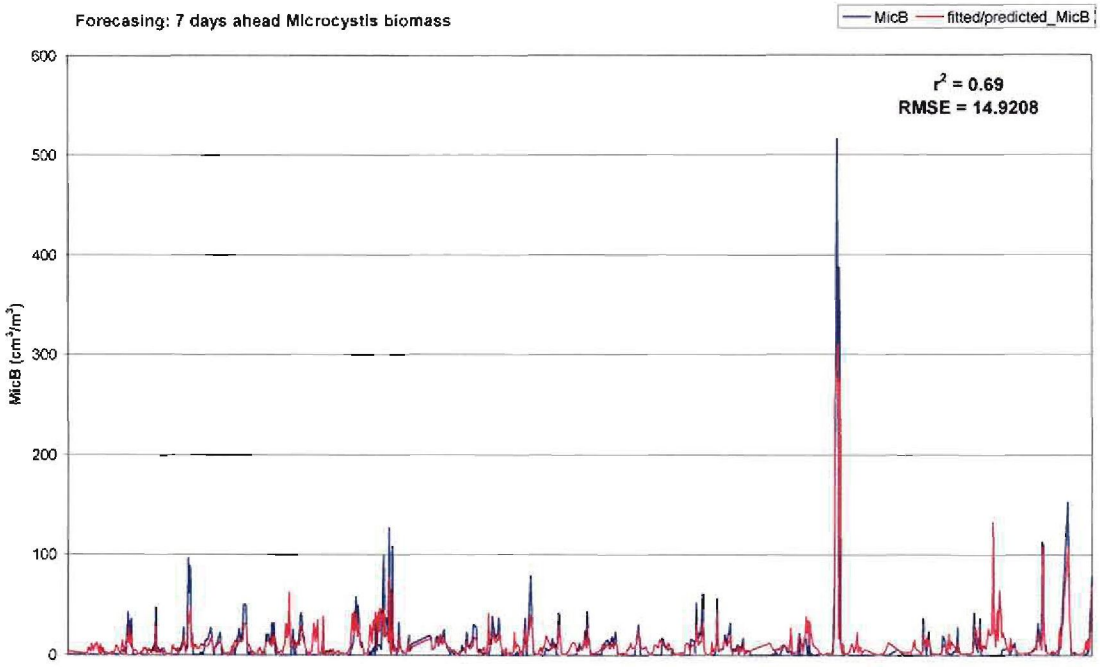


Figure 5.45 Training results of the 7-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the developed rule set of experiment 3.

The r^2 = 0.69 and the RMSE = 14.9208 for the training data and the r^2 = 0.52 and the RMSE = 9.3582 for the tested data are very good results and show that this 7-days forward prediction can be used on reservoirs with similar trophic status and climatic conditions. Fig. 5.46 shows the results of the tested data and it is clear that the rule set fits the Hartbeespoort and Roodeplaat reservoirs data better than the Rietvlei data, where under and over predictions are found. However, the magnitudes of the maximum *Microcystis* biovolume peaks are predicted quite well in all three the reservoirs.

Table 5.23 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the 7-days forward forecasting of *Microcystis* biovolume RULE set development.

	Training set	Test set
# of rows:	13148	2193
Average AE:	-1.43515	-0.2281
MSE:	12.22073	6.978118
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	53.9%	70.3%
Bad predictions:	46.1%	29.7%

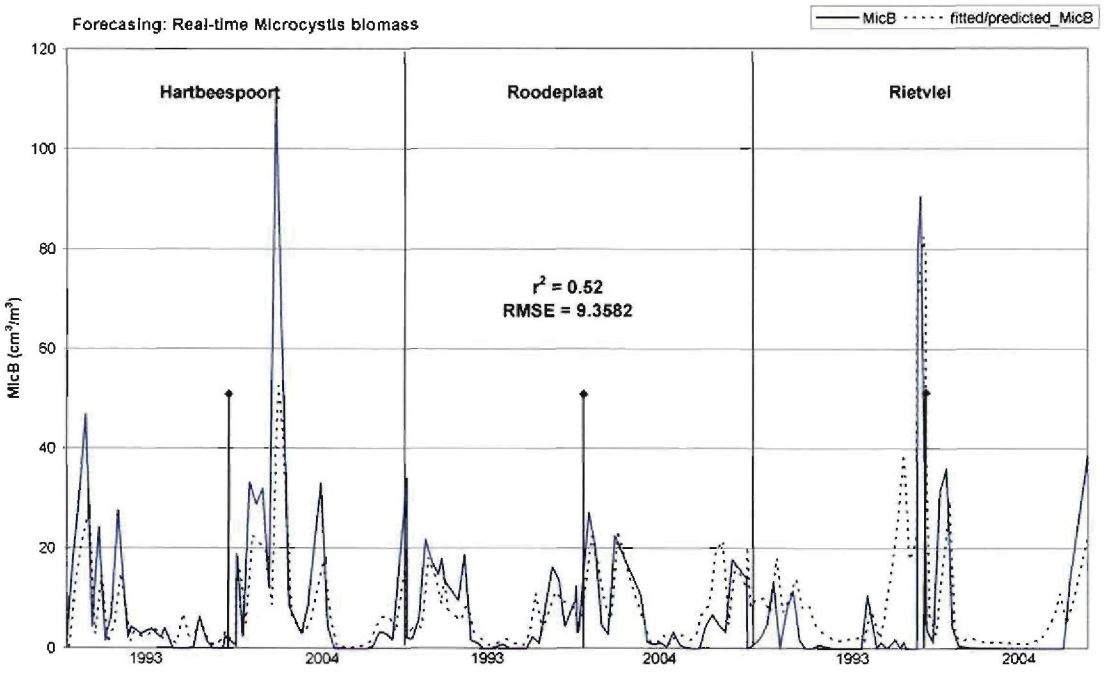


Figure 5.46 Testing results of the 7-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.43.

The further validation shows that the Relative Tolerance of 10% do give an outcome of 53.9% good predictions for the training data set (Table 5.23). The Relative Tolerance of 30% for the testing set shows an outcome of 70.3% good predictions. The smaller MSE found within the The

testing data set, when compared to the training data set, show that the predictive capability of the RULE set is better.

Fig. 5.47 illustrates the IF-THEN/ELSE RULE Set and the sensitivity analysis regarding the THEN- and ELSE- branches of the developed RULE Set for the 7-days-ahead prediction of *Microcystis* biovolume in the three reservoirs. Chl *a* concentrations, lower than 175.634ug/L is the condition to use the RULE Set of the THEN-branch.

In the THEN branch, the Chl *a* concentration and the surface temperature are the variables that are used to predict the *Microcystis* biovolume. The outcome of the *Microcystis* biovolume is sensitive to both these environmental variables. However, Chl *a*, is the most important environmental variable in determining the outcome. The input of Chl *a* ranges between 0.6 and 175.66ug/L. The surface temperature is the second most important variable regarding the sensitivity of the *Microcystis* biovolume and the surface temperature ranges between 10.4°C and 32.4°C for the study.

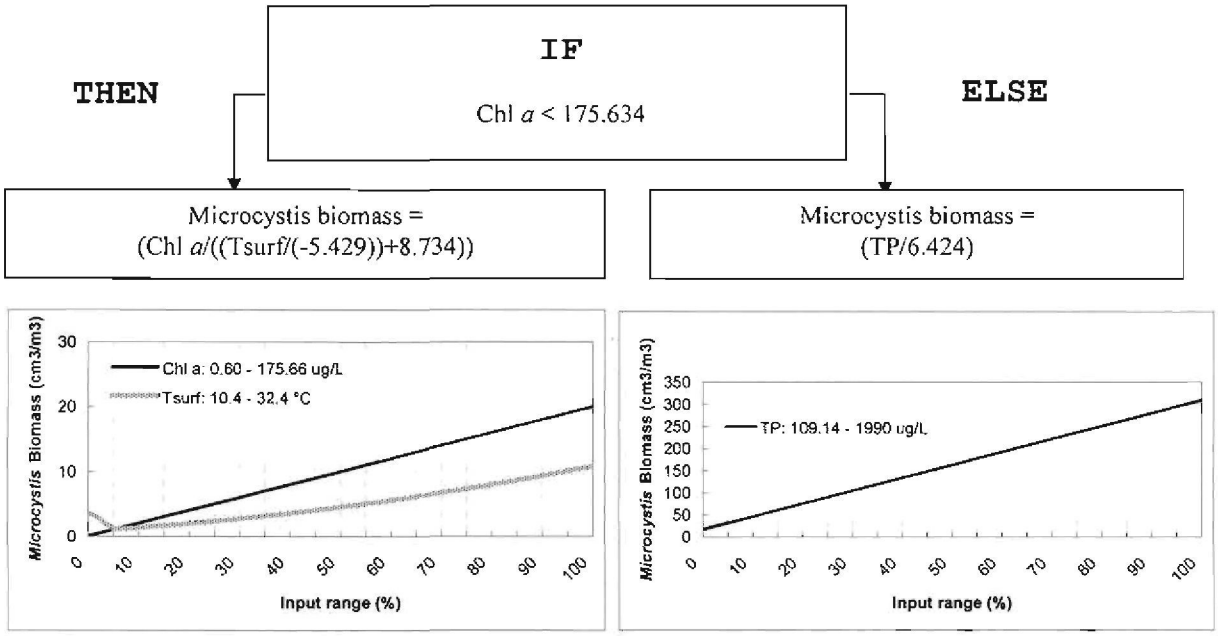


Figure 5.47 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 7-days forward forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

For the ELSE branch, where Chl *a* exceeds 175.634ug/L, TP is the most important environmental variable that determines the *Microcystis* biovolume (Fig. 5 43).

These results indicate that the hypertrophic reservoirs have probably such an excess of nutrients available in the system that temperature and the presence of high primary productivity, as indicated by the Chl *a* concentrations, are the main factors that drive the development of further *Microcystis* biovolume. Thus, the initial inoculum of cyanobacteria leads to the extent and further development of *Microcystis* biovolume.

Under high biovolume conditions, it is the TP concentrations present in the system, which drives the ELSE rule set. This is the case during high primary production biovolume conditions and TP is the sole variable used in the ELSE rule set to predict *Microcystis* biovolume. This can be because TP is used at a high rate, which determine the growth rate of the population.

5.2.4.4 HEA Experiment 4: 14-Days forward prediction simulation and rule set discovery for *Microcystis*

To develop a rule set for the 14-days forward prediction of *Microcystis* biovolume the daily interpolated data of the 36 years data set was used for training on the Hartbeespoort, Rietvlei and Roodeplaat reservoir. The *Microcystis* biovolume, measured data, was shifted backwards to create the data set of the 14-days forward training. The training results from this experiment are shown in Figure 5.48.

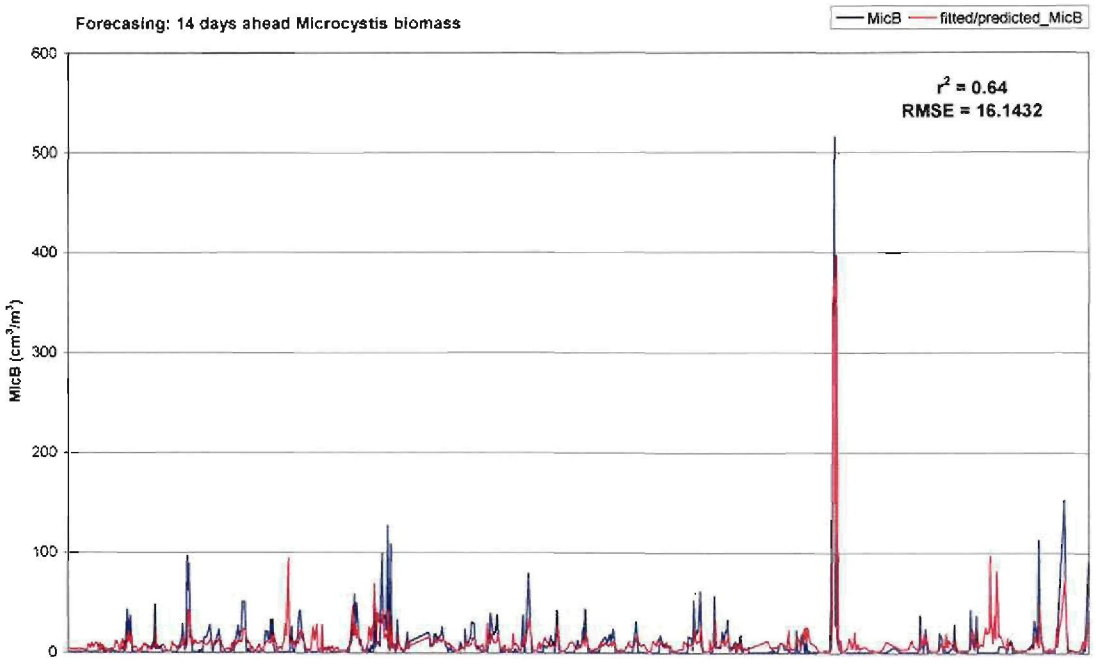


Figure 5.48 Training results of the 14-days forward forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 4, including Chl *a* as an environmental factor.

The r^2 of 0.64 and the RMSE of 16.1432 for the training data and the r^2 of 0.29 and the RMSE of 11.0926 for the tested data are still good results. However, the results are not as good as with the 7-days forward predictions. The results show that the 14-days forward prediction can be used on reservoirs with similar trophic status and climatic conditions with relative success. Fig. 5.49 shows the results of the tested data and it is clear that the rule set does predict the peaks, however the extent of the peaks are not predicted accurately.

Table 5.24 Summary of the average actual error (AE), the mean square error (MSE) and the relative tolerance of the training and testing data sets, of the *Microcystis* biovolume 14-days ahead RULE set development.

	Training set	Test set
# of rows:	13148	2193
Mean AE:	-0.34259	1.302998
MSE:	13.85358	5.687873
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	55.4%	67.5%
Bad predictions:	44.6%	32.5%

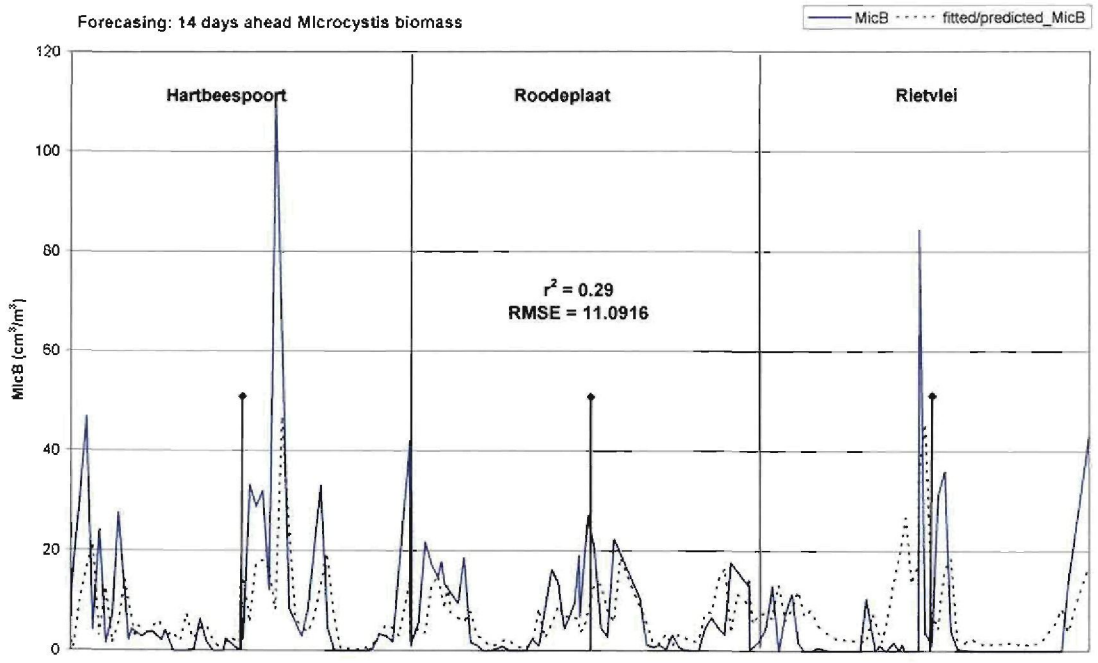


Figure 5.49 Testing results of the 14-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best developed RULE set as shown in Figure 5.50 including Chl *a* as an environmental variable.

The further validation show that the Relative Tolerance of 10% do give an outcome of 55.4% good predictions for the training data set. The Relative Tolerance of 30% for the testing data set shows an outcome of 67.5% good predictions (Table 5.24). The decrease in MSE found between

the training and the Testing data set show that the predictions capability of the RULE set is still quite good.

Fig. 5.50 illustrates the IF-THEN/ELSE RULE Set and the sensitivity analysis regarding the THEN- and ELSE- branches of the developed RULE Set for the 14-days forwards prediction of *Microcystis* biovolume in the three reservoirs. Chl *a* concentration lower than, or equal to 194.455ug/L, or a dissolved inorganic phosphorous concentration of less than 498.225 ug/L, is the condition to use the RULE Set of the THEN-branch.

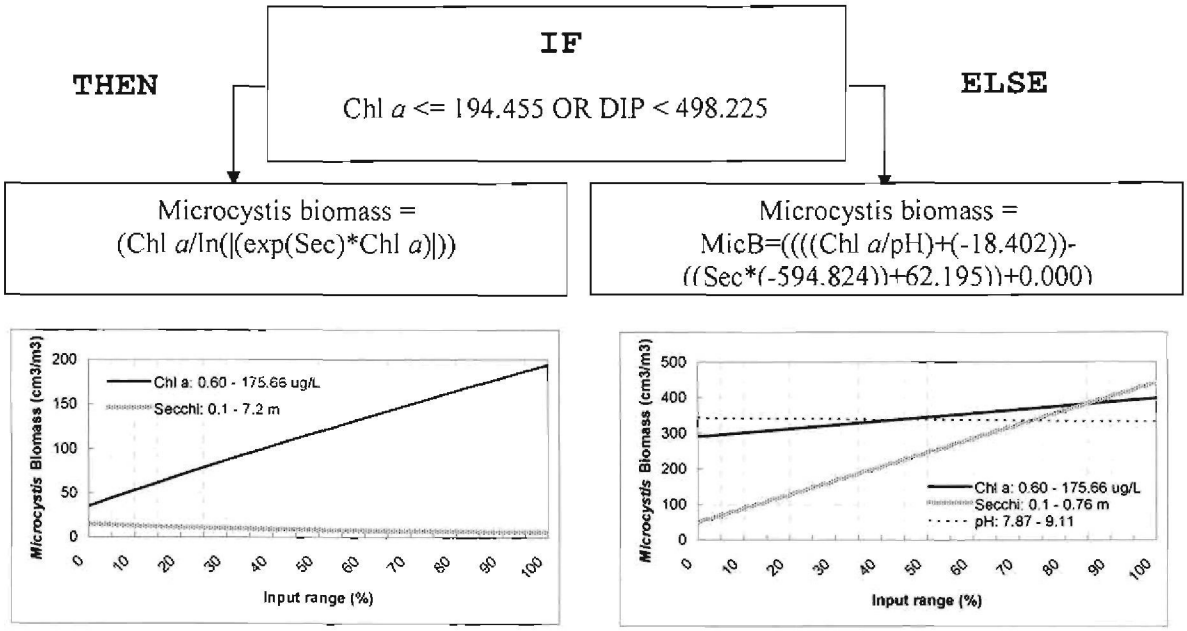


Figure 5.50 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 14-days forward forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

In the THEN branch, the Chl *a* concentration and the Secchi disc depth are the two environmental variables that are used to predict the *Microcystis* biovolume. The outcome of the *Microcystis* biovolume is sensitive to these environmental variables are shown in Fig. 5.46. Chl *a* is the most important environmental variable in determining the outcome. The input of Chl *a* ranges between 0.6 and 175.66ug/L. The Secchi disc do have a slight negative impact on the outcome of the *Microcystis* biovolume and the Secchi disc readings ranges between 0.1m and 0.76m for the

THEN RULE set. This indicates that the THEN RULE set is applicable during all turbidity conditions. For the ELSE branch, where exceeds 194.455ug/L or dissolved, inorganic phosphorous is greater than 498.225ug/L, the RULE set make use of the Chl *a*, Secchi depth readings and the pH to determine the *Microcystis* biovolume. In this case the Secchi disc reading is the environmental variable that the most important environmental variable in determining the *Microcystis* biovolume (Fig. 5 50). The ELSE RULE set is applicable to low light conditions as shown by the range of Secchi disc readings in Fig. 5.50. The Chl *a* concentrations are the second most important variable in the determination of the *Microcystis* biovolume.

As previously shown, these results indicate that these hypertrophic reservoirs have such an excess of nutrients available in the system that the light conditions in the systems (measured as Secchi disc readings) and the presence of high primary productivity (as indicated by the Chl *a* concentrations) is the main environmental variables that is used to determine the predicted development of *Microcystis* biovolume.

5.2.4.5 HEA Experiment 5: 21-Days forward prediction simulation and rule set discovery for *Microcystis*

To develop a rule set for the 21-days forward prediction of *Microcystis* biovolume the daily interpolated data of the 36 years data set was used for training on the Hartbeespoort, Rietvlei and Roodeplaat reservoir. The *Microcystis* biovolume measured data was shifted backwards to create the condition of 21-days forward training. The training results from this experiment are shown in Figure 5.51.

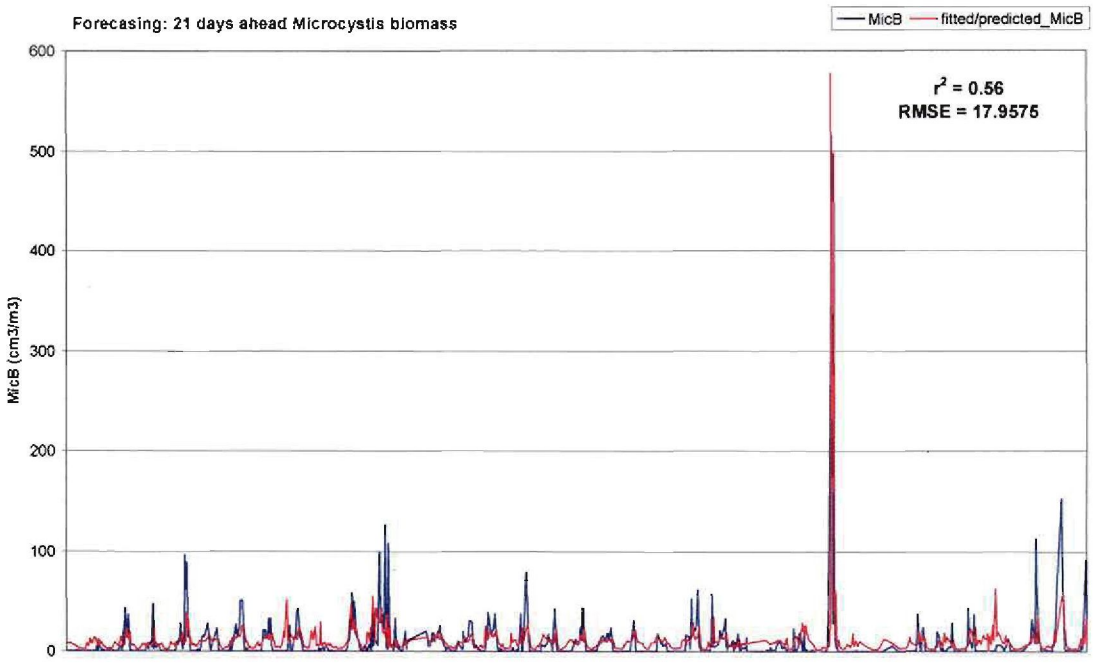


Figure 5.51 Training results of the 21-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 5 including Chl *a* as an environmental factor.

The r^2 of 0.56 and the RMSE of 17.9575 for the training data and the r^2 of 0.15 and the RMSE of 12.0300 for the tested data are still good results. However, the results are not as good as with the previous forward predictions. Despite the lower correlations, the results shown in Fig 5.51 and Fig 5.52 of the 21-days forward prediction can still be used on reservoirs with similar trophic status and climatic conditions with relative success. Fig. 5.52 shows the results of the tested data and it is clear that the rule set does predict the peaks, however, the extent of the peaks was not

reached and the predictions of this experiment was, therefore, not nearly as good as with the previous experiments.

Table 5.25 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume 21-days ahead rule set development.

	Training set	**Test set**
# of rows:	13148	2193
Average AE:	-0.9561	-0.10506
Average MSE:	13.21759	5.892838
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	55.4%	64.6%
Bad predictions:	44.6%	35.4%

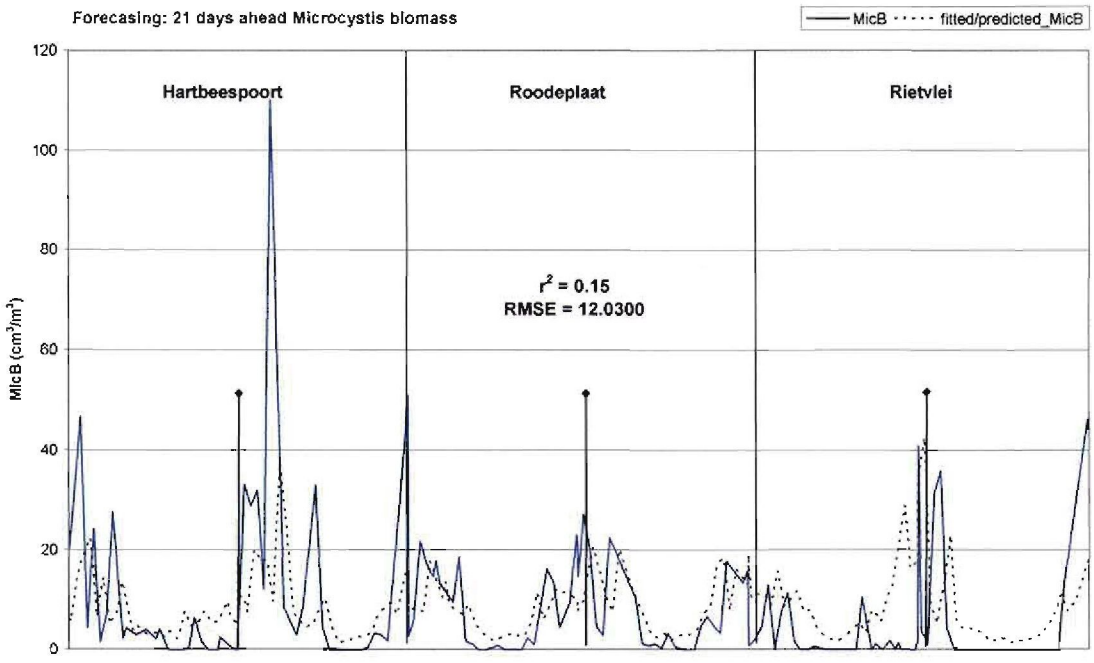


Figure 5.52 Testing results of the 21-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.53 including Chl *a* as an environmental variable.

The further validation show that the Relative Tolerance of 10% do give an outcome of 55.4% good predictions for the training data set. The Relative Tolerance of 30% for the testing data set shows an outcome of 64.6% good predictions (Table 5.25). The decrease in MSE found between the training and the Testing data set show that the predicting capability of the RULE set is still quite good.

Fig. 5.53 illustrates the IF-THEN/ELSE RULE Set and the sensitivity analysis regarding the THEN- and ELSE- branches of the developed RULE Set for the 21-days forwards prediction of *Microcystis* biovolume in the three reservoirs. There is a number of conditions that are used to determine if the THEN or ELSE branch of the RULE set are used in the prediction. Within the first condition, Chl *a* concentration can range between 0.6ug/L and 969ug/L and the TN concentrations vary between 0mg/L and 8.3mg/L. High DIP and TP concentrations are set as potential conditions (Fig. 5.53). If any of these four conditions as shown in Fig. 5.53 are met the THEN branch is used to determine the predicted outcome. Only when none of these conditions are met is the ELSE branch used in the predicted outcome.

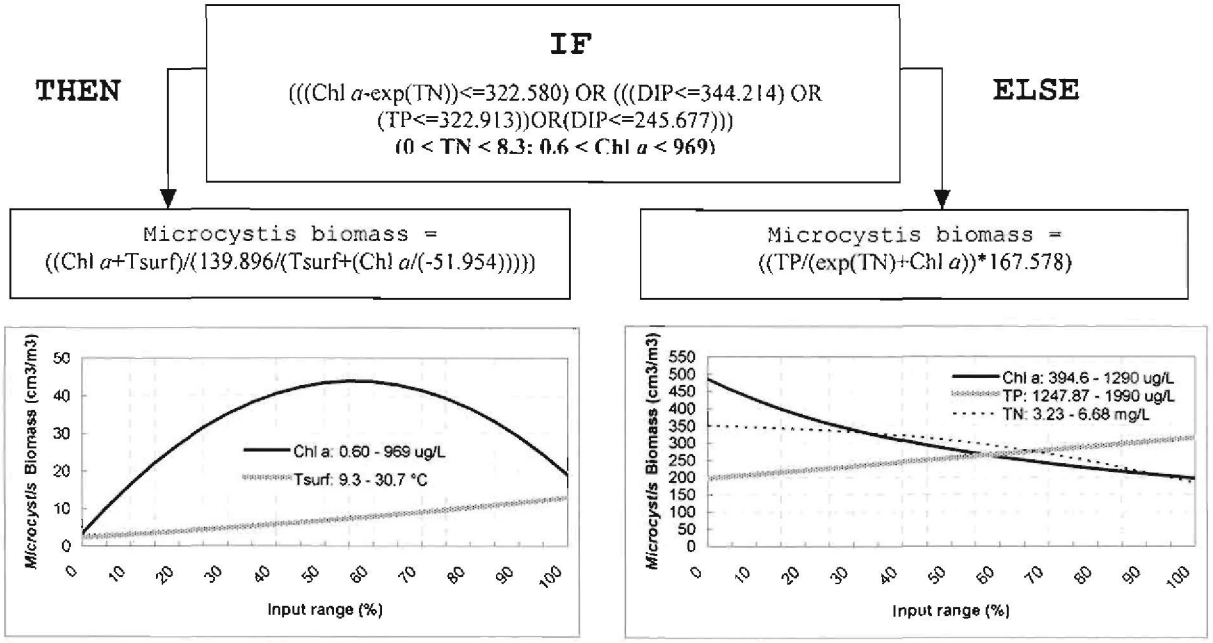


Figure 5.53 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 21-days ahead forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

In the THEN branch, the Chl *a* concentration and the Tsurf measurements are the two environmental variables that are used to predict the *Microcystis* biovolume. The outcome of the *Microcystis* biovolume is sensitive to these environmental variables as shown in Fig. 5.53. Chl *a* concentration is, once again, the most important environmental variable in determining the outcome. The input of Chl *a* concentration ranges between 0.6ug/L and 969ug/L. The Tsurf measurements do have a much smaller impact on the outcome of the *Microcystis* biovolume and the Tsurf measurements ranges between 9.3°C and 30.7°C for the THEN RULE set. It is interesting that at large changes in the Chl *a*, the sensitivity testing for the THEN branch in Fig 5.53 indicates that there is much lower *Microcystis* biovolume output, while the Tsurf input changes the development of *Microcystis* biovolume constantly. This may be explained by the shading effect of high biovolume, as indicated by high Chl *a*, on further growth. The THEN RULE set is applied under the lower phosphorus conditions although the set concentrations are still extremely high and typical of the hypertrophic conditions that is found in the set of reservoirs.

For the ELSE branch Chl *a* ranges between 394.6ug/L and 1290ug/L, the TP is in excess of 1247.87ug/L up to 1990ug/L and TN ranges between 3.23mg/l and 6.68mg/L. The nutrients are therefore in very high concentrations present in the systems during the ELSE RULE set determinations. The *Microcystis* biovolume outcome is sensitive to all three these variables under the ELSE RULE set conditions (Fig. 5 53). The ELSE RULE set is applicable to low light conditions as shown by the range of Secchi disc readings in Fig. 5.53. The Chl *a* concentrations are the second most important variable in the determination of the *Microcystis* biovolume. Both the high Chl *a* and the TN input has a negative impact on the outcome of the ELSE RULE set application, while the TP input has a positive increasing impact on the *Microcystis* biovolume outcome.

5.2.4.6 HEA Experiment 6: 28-Days ahead prediction simulation and rule set discovery for *Microcystis*

To develop a rule set for the 28-days forward prediction of *Microcystis* biovolume the daily interpolated data of the 36 years data set was used for training on the Hartbeespoort, Rietvlei and Roodeplaat reservoir. The *Microcystis* biovolume measured data was shifted backwards to create the condition of 28-days forward training. The training results from this experiment are shown in Figure 5.54.

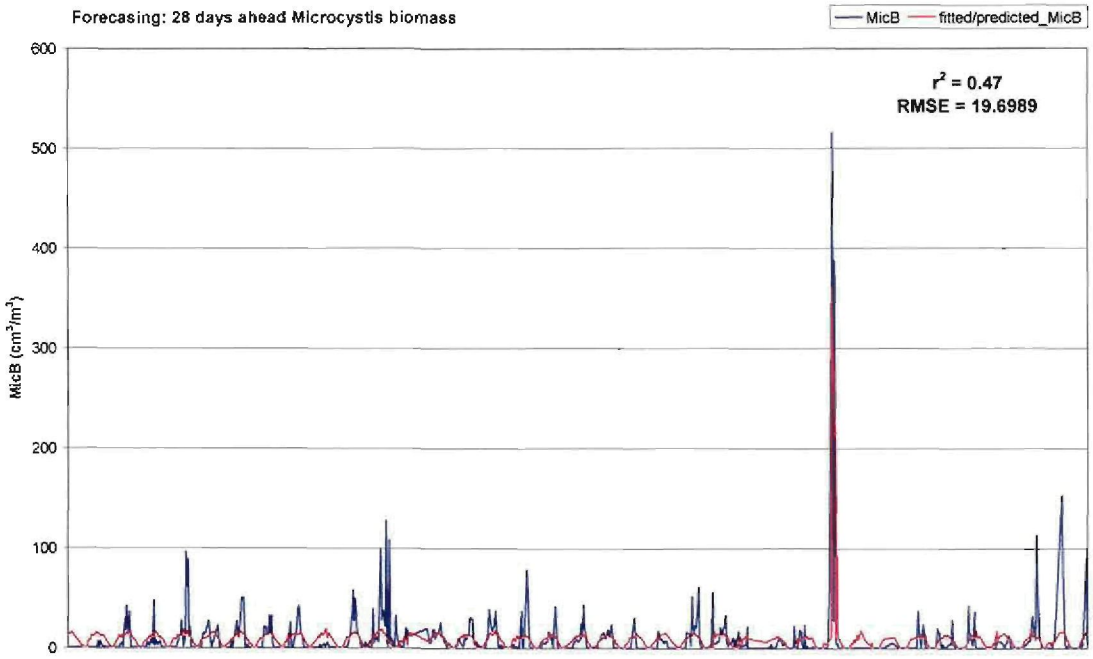


Figure 5.54 Training results of the 28-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs during the development of the best rule set for HEA Experiment 6 including Chl *a* as an environmental factor.

The r^2 of 0.47 and the RMSE of 19.6989 for the training data and the r^2 of 0.20 and the RMSE of 12.0657 for the tested data are still good results. However, the results are not as good as with the previous forward predictions. Despite this results in Fig 5.54 and Fig 5.55 show that the 28-days forward prediction can still be used on reservoirs with similar trophic status and climatic conditions with relative success. Fig. 5.55 shows the results of the tested data and it is clear that the rule set does predict the peaks, however, none of the highest peaks were predicted correctly.

Therefore, the 28-days forward predictions were not nearly as good as with the previous experiments, but do predict the cyclic occurrence of *Microcystis* correctly.

Table 5.26 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the training and testing data sets of the *Microcystis* biovolume 28-days ahead rule set development.

	Training set	**Test set**
# of rows:	13148	2193
Average AE:	0.942485	1.805445
Average MSE:	13.55385	4.613223
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	53.6%	55.5%
Bad predictions:	46.4%	44.5%

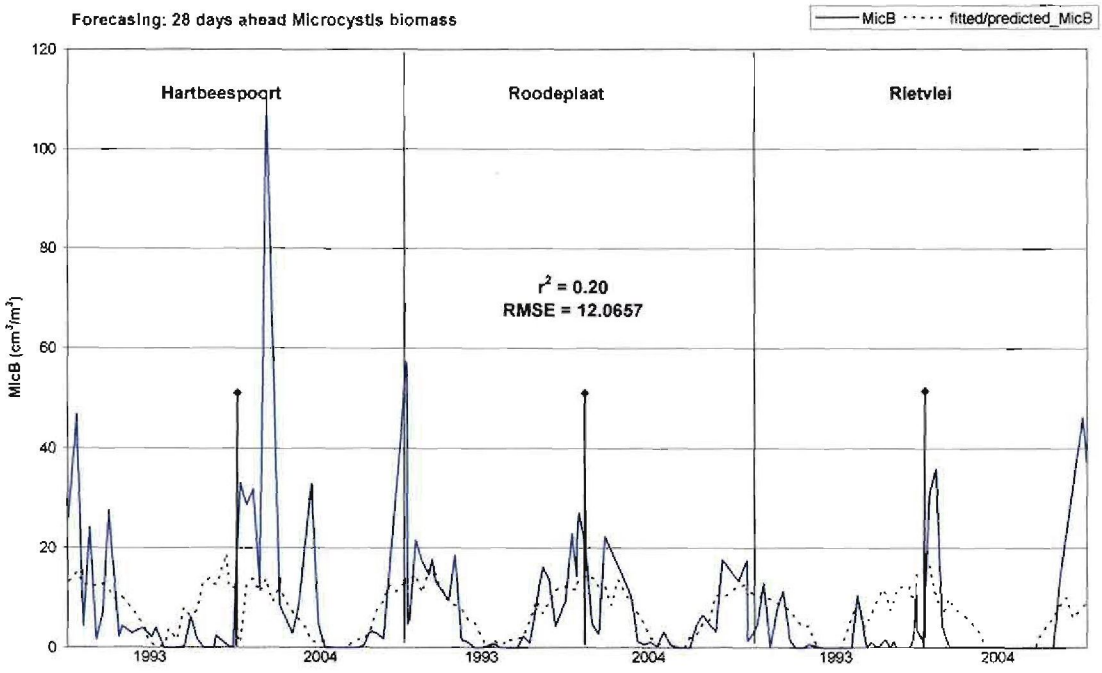


Figure 5.55 Testing results of the 28-days ahead forecasting of *Microcystis* biovolume (MicB) in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs using the best-developed rule set as shown in Figure 5.56 including Chl *a* as an environmental variable.

The further validation (Table 5.26) shows that the Relative Tolerance of 10% do give an outcome of 53.6% good predictions for the training data set. The Relative Tolerance of 30% for the testing data set shows an outcome of 55.5% good predictions. The decrease in MSE found between the training and the Testing data set show that the predicting capability of the RULE set is still quite good.

Fig. 5.56 illustrates the IF-THEN/ELSE RULE Set and the sensitivity analysis regarding the THEN- and ELSE- branches of the developed RULE Set for the 28-days forwards prediction of *Microcystis* biovolume in the three reservoirs. There is a two conditions that are required for the THEN RULE set to be used, namely DIP concentrations must be higher or equal to 355.453 ug/L and the Chl *a* concentration should be higher or equal to 180.962 ug/L.

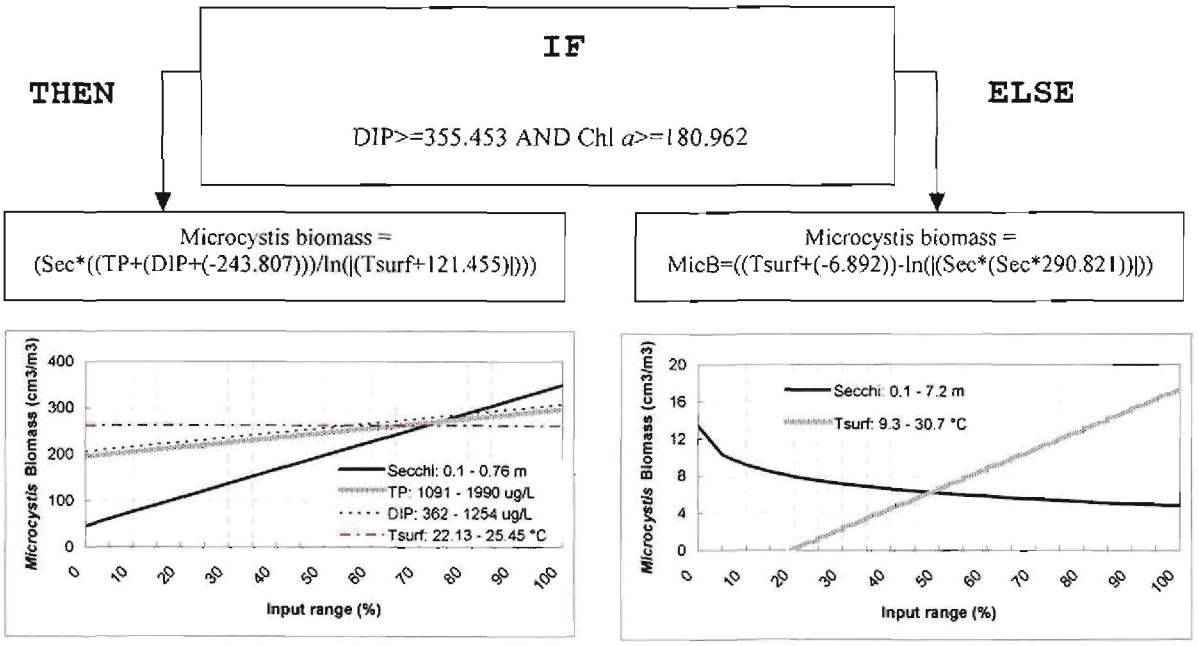


Figure 5.56 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the best developed rule set for 28-days ahead forecasting of *Microcystis* biovolume using the median concentrations of the most important environmental variables to determine the sensitivity of the *Microcystis* biovolume prediction to each variable.

To use the ELSE RULE set, the DIP and Chl *a* concentrations need to be lower than the mentioned conditions as shown in Fig. 5.56.

In the THEN branch the Secchi depth reading, the TP and DIP concentrations and the Tsurf measurements are the environmental variables that are used to predict the *Microcystis* biovolume. The outcome of the *Microcystis* biovolume is sensitive to these environmental variables as shown in Fig. 5.56. In the RULE set for the THEN branch, the changes in Secchi depth readings are the most important variable in determining the outcome of the predicted *Microcystis* biovolume. The second and third environmental variables that cause large changes in the predicted *Microcystis* biovolume outcome is the TP and DIP concentrations. Fourthly, the Tsurf input changes do have a much smaller impact on the outcome of the *Microcystis* biovolume and the Tsurf measurements range between 22.13°C and 25.45°C for the THEN RULE set. It is interesting that the THEN branch is found in this very small temperature range as shown in Fig 5.56, indicating optimum temperatures for the development of *Microcystis* biovolume in these reservoirs. The THEN RULE set is applied under extremely high phosphorus (TP and DIP).

For the ELSE branch it is the Tsurf readings and the Secchi disc depth that are the most important environmental variables that affects the large changes in the *Microcystis* biovolume development. The Tsurf ranges between 9.3°C and 30.7°C and has the greatest impact on the model outcome. The Secchi disc readings that range between 0.1m and 7.2m have a negative impact on the *Microcystis* biovolume development.

The ELSE RULE set conditions covers a very wide section of the data and model outcomes, while the THEN RULE set covers only a small section of the data as the Secchi disc reading range and the Tsurf range are very small. The THEN RULE set is applicable to very low light conditions as shown by the range of Secchi disc readings in Fig. 5.56.

5.2.4.7 HEA Experiment 7: Real-Time Simulation and Rule Set Discovery for the dinoflagellate *Ceratium*

The rule set for the dinoflagellate, *Ceratium*, real time prediction was developed as discussed in the methods and the results were tested on the unseen data of Bon Accord and Klipvoor reservoirs. The method proved effective in the development of a rule set for predicting *Microcystis* biovolume and the method was thus applied to the prediction of dinoflagellate blooms that consists primarily of *Ceratium*. This species seem to be more pronounced since 1999 in the studied reservoir systems (Van Ginkel *et al.* 2001).

Testing of the Rule Set

The developed rule set was tested for five reservoirs, namely the three reservoirs that it was trained and tested on (Hartbeespoort, Rietvlei and Roodeplaat) (Fig. 5.57 & Fig. 5.58) and two unseen reservoirs (Bon Accord and Klipvoor) (Fig. 5.60 & Fig. 5.61). Both the latter reservoirs experienced annually large blooms of dinoflagellates (consisting of *Ceratium*) during the period 2000 to 2005. The rule set was applied and tested to 1993 and 2004 on the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs (Fig. 5.58). The rule set was also tested and applied to the 2000 to 2005 data of the two unseen reservoirs, Bon Accord and Klipvoor (Fig.5.60 & Fig. 5.61).

To validate the results of the rule set the correlation coefficient of the measured and fitted data were determined. The Root Mean Square Error (RMSE) of the testing error was also calculated for the different reservoirs to indicate the standard error of the estimate.

Sensitivity Analysis of Developed Rule Set

Sensitivity analyses were done on the tested data for the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs for the period 1993 and 2004. The sensitivity analysis was done for both the THEN and the ELSE rule set (Fig. 5.59).

The input range changes (as percentage) were determined by applying the rule set to calculate changes in each variable that was used in the rule set, while the other variables median values were used. The starting point of the environmental variable was the minimum and the maximum values measured within the data set. This gives an indication of the importance of each variable in

driving the changes (increases or decreases) of the outcome, namely the Dinoflagellate biovolume.

Real-time rule-set discovery for *Ceratium* biovolume in Hartbeespoort, Roodeplaat and Rietvlei Reservoirs

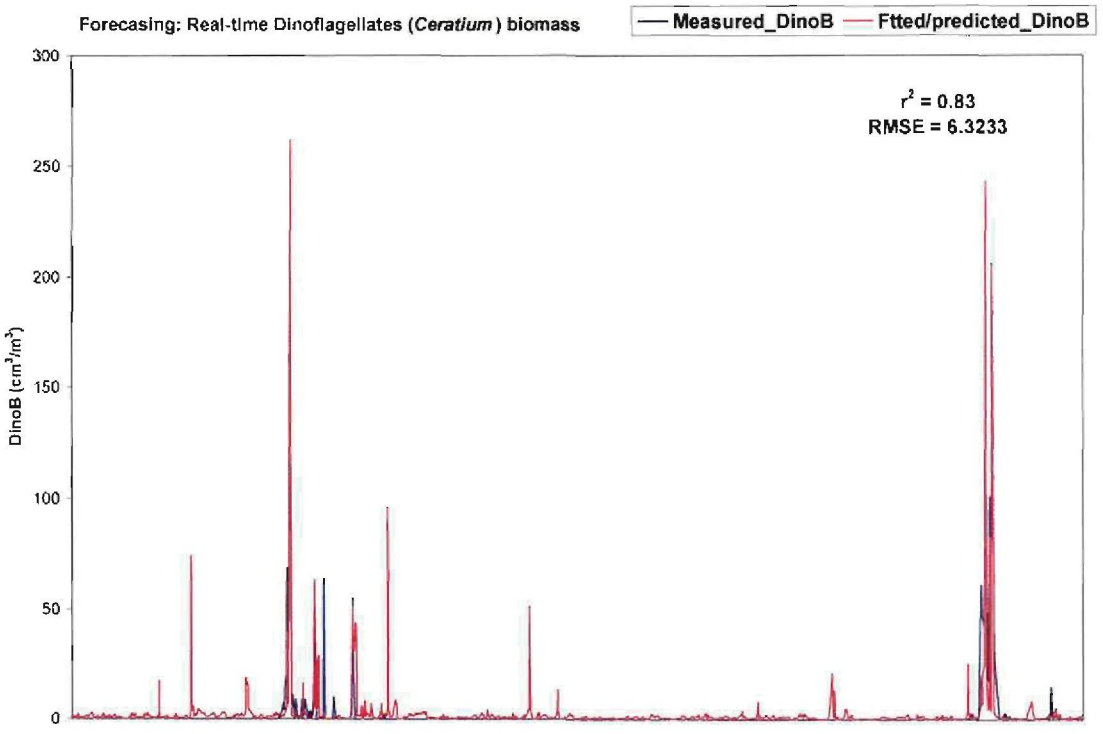


Figure 5.57 Training results of the real-time forecasting of *Ceratium* biovolume (DinoB), in the Hartbeespoort, Roodeplaat and Rietvlei reservoirs during the development of the best rule set for HEA Experiment 7 including Chl *a* as an environmental factor.

The 36 years of real-time training on the Hartbeespoort, Rietvlei and Roodeplaat reservoir daily data produced a rule set for the prediction of *Ceratium* biovolume as shown in Figure 5.38. The correlation coefficient (r^2 = 0.83) and the root mean square error (RMSE = 6.3233) for the training data and the correlation coefficient (r^2 = 0.85) and the root mean square error (RMSE = 6.7857) for the tested data are still significant. Fig. 5.58 shows the results of the tested data on the three reservoirs for the period 1993 and 2004 for the Hartbeespoort, Roodeplaat and Rietvlei reservoirs. It is clear that the rule set fits the Hartbeespoort Reservoir better than the Roodeplaat and the Rietvlei data, where under and over predictions were found. However, the

magnitudes of the maximum *Ceratium* biovolume peaks were predicted quite well in all three the reservoirs.

Table 5.27 Summary of the average actual error (AE), the average mean square error (MSE) and the relative tolerance of the real time training and testing data sets of the *Ceratium* biovolume rule set development.

	Training set	**Test set**
# of rows:	13148	2193
Average AE:	-0.59789	-0.63201
Average MSE:	5.763118	6.593074
Tolerance type:	Relative	Relative
Tolerance:	10%	30%
Good predictions:	78.4%	85.5%
Bad predictions:	21.6%	14.5%

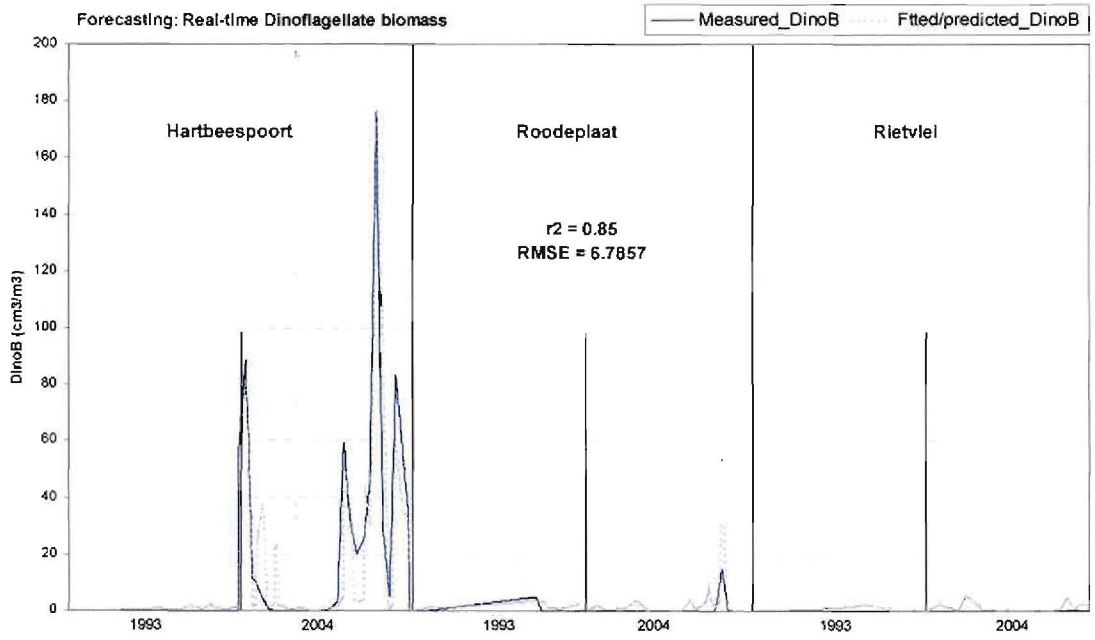


Figure 5.58 Testing results of the real-time forecasting of *Ceratium* biovolume (DinoB), in the Hartbeespoort, Roodeplaat and Rietvlei reservoirs using the rule set as shown in Figure 5.59, including Chl *a* as an environmental factor.

Further validation shows that the Relative Tolerance of 10% does give an extremely good outcome of 78.4% good predictions for the training data set. The Relative Tolerance of 30% for the testing data set shows an outcome of 85.5% good predictions (Table 5.27). The increase in MSE found between the training and the testing data set show that the predicting capability of the RULE set decreased from the training to the testing data set.

The conditions for the rule set included TP and Chl *a* concentrations that were used to determine if the THEN branch or the ELSE branch of the rule set was to be used for forecasting the *Ceratium* biovolume. The THEN branch is applicable to situations when the Chl *a* concentrations and the Dinoflagellate biovolume are very high (Fig. 5.40). The sensitivity analysis that was done on the real-time testing data for the three reservoirs to determine the *Ceratium* biovolume shows that with the THEN branch of the rule set Chl *a*, TN and TP are the variables that the Dinoflagellate biovolume is most sensitive to. The TN concentrations vary from 2.1mg/L to 4.53mg/L and TP concentrations vary between 110.86ug/L and 272ug/L. Under these high nutrient conditions, other variables were not important in the forecasting of the *Ceratium* biovolume.

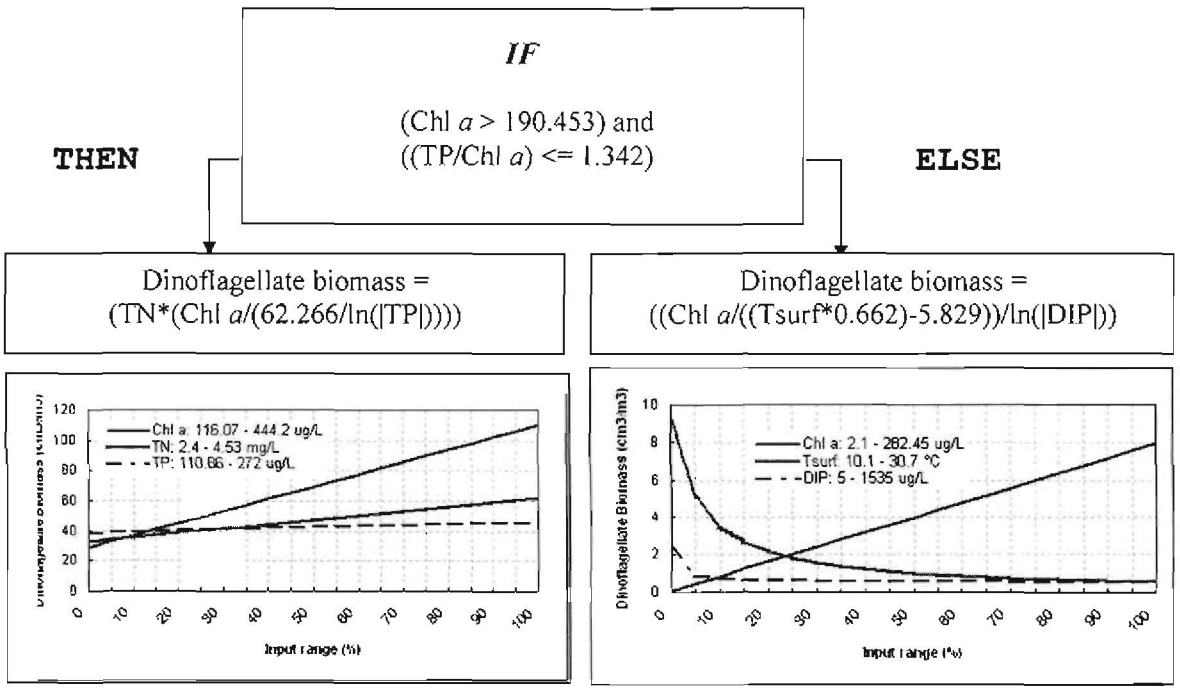


Figure 5.59 Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the rule set for real-time Dinoflagellate biovolume forecasting.

The ELSE branch of the rule set is used when Chl *a* concentrations varied between 2.1ug/L and 282.45ug/L and the *Ceratium* biovolume was below 10 cm^3/m^3. The temperature change of the surface water is important in determining the *Ceratium* biovolume. Increases in the temperature input were important in the decrease of the *Ceratium* biovolume (Fig. 5.59). This showed that the optimum growth temperature that ranged from 5°C to 30°C according to Buck (1989) is only important under low Chl *a* concentrations and up to 17°C. Temperatures higher than 17°C showed no significance on the dinoflagellate biovolume. Higher *Ceratium* biovolume is more regulated by the availability of sufficient nutrients (Reynolds, 1978; Buck, 1989) within the water body as shown by the THEN rule set. Under lower Chl *a* concentrations, the availability of DIP was important up to 82ug/L after which the effect of the change in input on the determination of the *Ceratium* biovolume was insignificant.

Real-time rule-set for Ceratium biovolume testing in Bon Accord and Klipvoor Reservoirs

Five years of data from both the Bon Accord and Klipvoor Reservoirs were used to test the applicability of the rule set to unseen data. The correlation of the measured and predicted *Ceratium* biovolume in the Bon Accord Reservoir (Fig. 5.60) was 0.62, which is statistically highly significant ($P < 0.001$). The RMSE of 6.803 also indicates that the results were significant. The extent of the peaks were under-predicted or over-predicted in certain instances, but all the peaks were predicted.

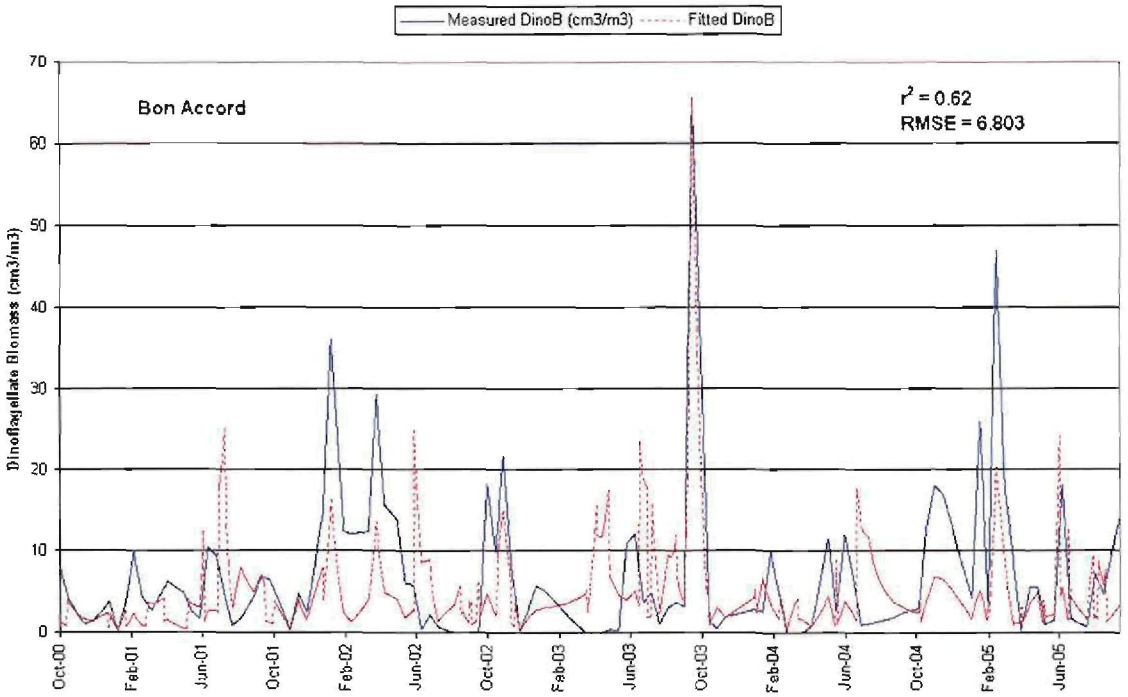


Figure 5.60 Testing of the developed rule set on Bon Accord Dam for real-time Dinoflagellate (*Ceratium*) biovolume forecasting.

On two occasions, peaks were predicted that did not occur (Autumn 2003 and Spring 2004). This may be due to the occurrence of other phytoplankton taxa (e.g. *Microcystis*) that dominated the community and the rule set used the Chl *a* concentration in determining the predicted biovolume of the *Ceratium* in the Bon Accord Reservoir. The maximum peak of *Ceratium* biovolume in the spring of 2004 was predicted very well.

In the Klipvoor Reservoir (Fig. 5.61), the correlation of the measured and predicted *Ceratium* biovolume was 0.48, which is significant ($P < 0.001$). Although the correlation coefficient is less than for Bon Accord the results is still statistically significant. The RMSE, as an 'estimation' of the standard deviation, of 2.53 also indicated that results were significant. The extents of the peaks were under-predicted or over-predicted, but all the peaks were predicted. Peaks were predicted in the spring of 2002 and the autumn of 2003, which did not occur.

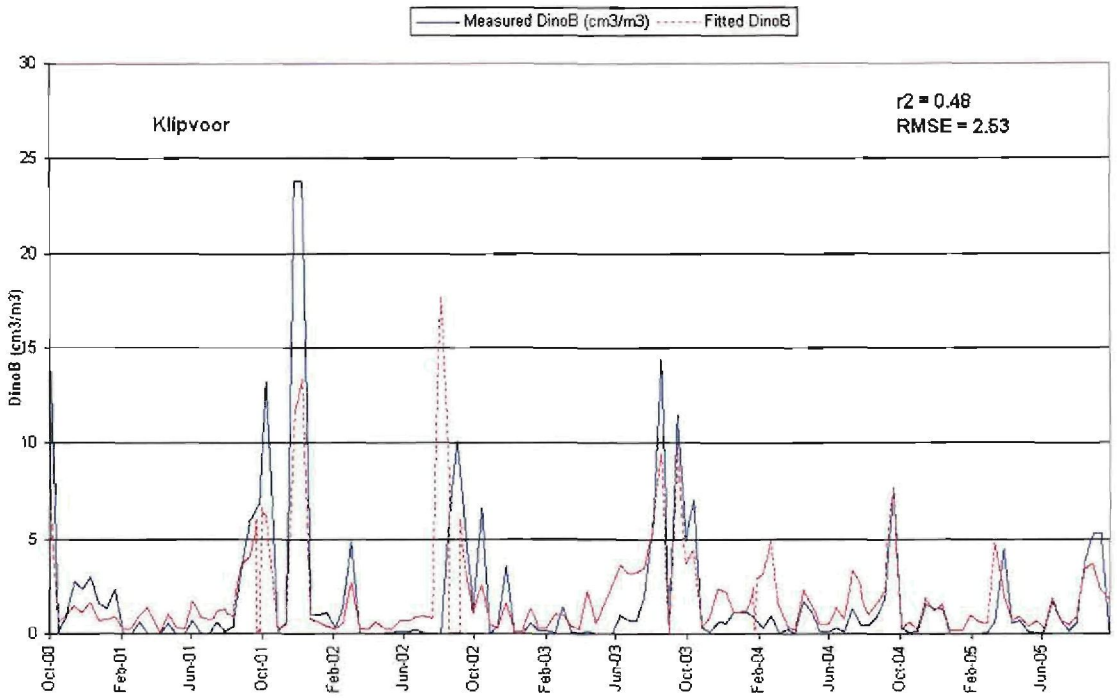


Figure 5.61 Testing of the developed rule set on Klipvoor Dam, for real-time *Ceratium* biovolume (DinoB) forecasting.

These results indicate that the RULE set developed on hypertrophic reservoirs in the summer rainfall and temperate region of South Africa is applicable to reservoirs within the same climatic region and of the same hypertrophic status. The methods used may further be investigated for applicability in other climatic regions of South Africa and on reservoirs with different trophic status to determine if a separate RULE set needs to be developed for different climatic zones or for reservoirs of different trophic status.

The developed RULE sets made use of the most important environmental variables as have been determined through years of limnological research (Table 5.28). For the *Microcystis* RULE Set development, the inclusion of the Chl *a* concentrations gave the best results. The most important environmental variables that were used in determining the final most applicable RULE sets for the different experiments are the existing biovolume that are measured as Chl *a*, Tsurf, TP, DIP, Secchi (as indication of light conditions), TN and lastly pH. In no circumstances was DIN used in the most applicable RULE sets for the prediction of *Microcystis* or in the prediction of *Ceratium*. This may indicate that although DIN is the food source, the TN is more important and determine the eventual availability of nitrogen and contribution to growth.

Table 5.28 The summarised results from the HEA Experiments, showing the correlation coefficient (r^2), Root Mean Square Error (RMSE), the relative tolerances (10% for training and 30% for testing), the Environmental Variables important in Conditions to apply the THEN or ELSE Rule set, and the Environmental variables (in order of importance) that were used in the developed RULE sets. Experiment 1-6 was the development of RULE sets for *Microcystis*, and Experiment 7 was done for the development of a RULE set for the dinoflagellate, *Ceratium*.

Experiment	r^2	RMSE	Relative Tolerance	Environmental Conditions	Environmental Variables Used	
					THEN	ELSE
Experiment 1:				TP	1. Tsurf	1. Chl *a*
Real time + Chl *a*				Chl *a*	2. TN (neg)	2. Tsurf
Training:	0.88	9.3357	38.7%			3. pH
Testing	0.73	7.4120	44.2%			
Experiment 2:				TN	1. TP	1. Secchi
Real time – Chl *a*				Secchi	2. TN	(neg)
Training:	0.81	11.7002	42.3%		3. Tsurf	2. pH
Testing	0.46	10.3676	62.9%		4. pH	
Experiment 3:				Chl *a*	1. Chl *a*	1. TP
7 days forward					2. Tsurf	
Training:	0.69	14.9208	53.9%			
Testing	0.52	9.3582	70.3%			
Experiment 4:				Chl *a*	1. Chl *a*	1. Secchi
14 days forward				DIP	2. Secchi	2. Chl *a*
Training:	0.64	16.1432	55.4%			3. pH
Testing	0.29	11.0916	67.5%			
Experiment 5:				TN	1. Chl *a*	1. Chl *a*
21 days forward				Chl *a*	2. Tsurf	2. TP
Training:	0.56	17.9575	55.4%			3. TN
Testing	0.15	12.0300	64.6%			
Experiment 6:				DIP	1. Secchi	1. Tsurf
28 days forward				Chl *a*	2. TP	2. Secchi
Training:	0.47	19.6989	53.6%		3. DIP	(neg)
Testing	0.20	12.0657	55.5%		4. Tsurf	
Experiment 7:				Chl *a*	1. Chl *a*	1. Chl *a*
Real time + Chl *a*				TP	2. TN	2. Tsurf
Training:	0.83	6.3233	78.4%		3. TP	3. DIP
Testing 1	0.85	6.7857	85.5%			
Testing 2	0.62	6.8030	43.2%			
Testing 3	0.48	2.5300	70.1%			

For the dinoflagellate, *Ceratium,* it was shown that a temperature range of 5°C to 17°C and DIP concentrations below 82 ug/L are important environmental conditions during the development of the initial low Dinoflagellate biovolume. In the development of excessive Dinoflagellate blooms TN is the most important variable in determining the extent of the Dinoflagellate bloom.

In all the cases where pH was used in the developed RULE set, the change in pH input had little effect on the output algal biovolume, and it can be considered to be not a major contributor to the final results. Tsurf seems to be one of the most important environmental variables together with the nutrient concentrations (Table 5.28). TP and DIP are the most important nutrients in determining the outcome of the algal biovolume. TN when used has mostly a negative impact on the algal biovolume output of the model.

It is interesting, that for the forward prediction of *Microcystis* the 7-days ahead forecasting gave the best results (Table 5.28), followed shortly by the 14-days, the 21-days and the 28-days ahead forecasting. The worst predictions (28-days), as found with the real time results, can be explained by the fact that real time environmental conditions are not causing the biovolume, but are rather the result of the biovolume.

From these results, it is shown that the existing biovolume, temperature, nutrients specifically phosphorous, and the existing light conditions are the most important environmental variables in the determination or prediction of the algal biovolume of *Microcystis*. In the case of *Ceratium* the nitrogen and temperature are the most important environmental variables in determining the predicted biovolume.

5.3 Conclusions and Evaluation

Five modelling techniques were included in the study and tested on hypertrophic reservoirs in South Africa. The data and results generated with the different models showed variable successes to predict chlorophyll *a* (Chl *a*), phosphorous, algal groups and specific species. The testing of management options in some of the models proved to be quite useful.

Firstly, the relatively simple Vollenweider Model, a well-known and widely tested model was used. The Vollenweider model (Vollenweider 1976) was applied in South Africa on a number of occasions (Thornton and Walmsley 1981; Grobler and Silberbauer 1984; Grobler 1985). This

model was originally developed as an eutrophication management tool. It did not provide information on the algal or cyanobacterial species that may pose a problem in the system. The ease of application of the model using total phosphorus as major input to determine the Chl *a* outcome still makes this model a favourite with managers as the effect of decreases or increases in total phosphorus does give a quick result for management purposes. It does, however, give no information on the phytoplankton species that may be present in the system and hence management problems.

The model does not take any of the complex characteristics of a freshwater system into consideration but uses a simple linear correlation that exist between the TP concentrations in a system and the potential to develop phytoplankton biovolume, as measured by Chl *a*. This model is giving the manager a quick reference on the trophic conditions within the system. It supplies the manager with an easy to use tool, for possible consideration of potential management options for phosphorous removal.

The second model tested is the SALMO-OO model, a complex deterministic model that takes growth equations of the different algal groups into consideration and that takes the main impacting factors like inflow and nutrient loads, as inputs, into consideration. The simulation library SALMO-OO has been developed by the object-oriented implementation of the lake ecosystem model SALMO (Recknagel and Benndorf 1982; Benndorf and Recknagel 1982; Recknagel *et al.* 1995) as the core of a lake simulation library by means of JAVA. It provides optional access to alternative causal representations of ecological processes in lakes such as: a) photosynthesis and respiration of diatoms, green algae and cyanobacteria; b) grazing of diatoms, green algae and cyanobacteria by zooplankton; c) growth of herbivorous zooplankton; and d) predation of zooplankton by planktivorous fish. Alternative process representations were adopted from published lake ecosystem models different to SALMO-OO such as Park *et al.* (1974), Hongqing and Jianyi (2002) and Arhonditsis and Brett (2005) to test applicability in hypertrophic reservoirs of South Africa.

SALMO-OO was tested on data from 2003 to 2004 on three hypertrophic reservoirs, Hartbeespoort, Klipvoor and Roodeplaat with regards to the simulation of seasonal abundances of diatoms, green algae and cynaobacteria. It proved to be valid for the temperate hypertrophic lakes and it can assist in optimising eutrophication control of these lakes based on complex management scenario analysis. The one problem foreseen with this model is the fact that it does

not make provision for dinoflagellates, which are becoming more dominant in the hypertrophic systems of South Africa.

The nutrient loads are used in the model to predict the outcomes. The model is designed to test the application of different potential management options, which make it applicable for use by mangers that need to determine management options.

The predicted outcomes of the method are not quite as accurate as for lakes in the literature, but do predict the highs and lows of bloom development to a certain degree of accuracy. Discrepancies in the predictions may be because not all algal groups are included in the model. Especially the Dinoflagellate group is not included, probably because major blooms was not pronounced during the development of SALMO-OO, and may account for the discrepancies in the predictions, as all five the reservoirs did experience dinoflagellate blooms during the study period.

Thirdly, the Artificial Neural Network modelling methods tested for the study included two techniques. Firstly, the multilayered fast-forward neural network model (MFNN), Forecaster XL, and secondly, a Self Organising Mapping (SOM) technique were used to determine if it is applicable for usage in South African conditions.

The results achieved with Forecaster XL gave an output that shows the successes and failures of the results. The percentage tolerance of the modelling package with the training data set is 10% and with the testing data set 30%. The model is, therefore, quite strict with itself, in accepting the efficiency of the model. The modelling results did not show very high successes, however, the user friendliness of the program makes it a good choice for further use and testing. Because of the hidden layer of the ANN modelling technique the main problem of the program is that it does not provide the user with the mathematical methods used within the model for easy application on new data for prediction purposes.

The MFNN model, Forecaster XL, is easy to use and fits well within the wide use of Microsoft Excel as background. Data import, and the output from the model make this model very user friendly. The output does give one results that show the success and failures of the model. The two negative features of the model are:

a) The modeller do not know what algorithms were used to determine the output,

b) The methods to determine the tolerance of the results within the model are unknown.

The SOM modelling method, which is not a forward predictive modelling technique, but rather gives insight into historical events, showed potential for further use in South Africa as it can be used to:

- Investigate the before and after situations,
- To determine the potential causes of management actions,
- To determine changes in environmental conditions.

The SOM method is not a prediction method but is an ecological informatics method that can be used to determine cause and effect of changes that occur within a reservoir or catchment system. This method may need to be tested more intensively in South Africa for applicability to other systems and data sets. Using SOM, to test the impact of management and environmental changes that may have occurred within a catchment, reservoir or other system is highly applicable to this method.

Lastly, with the hybrid evolutionary algorithm (HEA) development, a number of tests were done to develop real time RULE agents for the prediction of *Microcystis* spp. and the Dinoflagellate (*Ceratium hirundinella*) bloom events in South African hypertrophic reservoirs. Validation of the results was done on the limnological time-series data. The tests to develop a prediction tool for *Microcystis* spp. was done with real time prediction, 7-days forward, 14-days forward, 21-days forward and 28-days forward predictions. The 7 day-forward prediction rule-based agent for *Microcystis* spp. has proven to be most accurate. The HEA was designed to assemble and optimise both the structure and parameters of predictive rules using genetic programming and evolutionary computation. In order to develop the rule-based agent for *Microcystis* and the Dinoflagellate group, merged limnological time-series data of the hypertrophic reservoirs Hartbeespoort, Rietvlei and Roodeplaat dams have been used for training. Rigorous leave-*k*-out cross-validation for a total of 36 years (12 years from each reservoir during the period 1991 - 2004) of data was used to do the rule based development training, excluding the years 1993 and 2004 (from each reservoir) to use for testing the developed rule.

The developed RULE sets of the hybrid evolutionary algorithms (HEA) for forecasting *Microcystis* and the dinoflagellate (*Ceratium*) biovolume proved to be highly applicable to complex unseen ecological data of South African Reservoirs with the same trophic status as the

reservoirs that were used to develop the RULE Set. The five reservoirs used for the testing of the RULE Set are all within the same temperate climatic region of South Africa and all reservoirs had the presence and dominance of either *Microcystis* or the dinoflagellate, *Ceratium*, during the study period, even though the sizes and other limnological characteristics differed between the reservoirs.

The sensitivity analysis and the best RULE set correspond well with theoretical hypotheses and experimental findings in previous studies. It can also refine ranges of variables that are deterministic in the development of the Dinoflagellate biovolume. This study indicates that the HEA methods for the development of RULE sets by machine learning of complex ecosystems such as man made Reservoirs is extremely applicable. The method should also be tested for further application in other ecological environments within South Africa.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary of Conclusions

The main aim of this project was to investigate the applicability of ecological infomatics modelling techniques to develop a predictive tool for harmful algal blooms in South Africa.

The Reservoirs studied are hypertrophic systems and are situated downstream of the largest urban development in the summer rainfall area of South Africa. These systems are subjected to severe algal blooms, dominated primarily by Cyanobacteria (*Microcystis*) and lately, also Dinoflagellates (*Ceratium hirundinella*).

The results from this study show that the cyanobacterial blooms in the Crocodile West/Marico Water Management Area are dominated by *Microcystis*, and although *Anabaena* and *Oscillatoria* are found occasionally in large numbers, they do form only a negligibly small component of the annual phytoplankton community. Furthermore, it was found that *Microcystis* is the most important cyanobacterial species, associated with the presence of cyanobacterial toxins in the five reservoirs. However, *Anabaena & Oscillatoria* are both also known TM producers and could contribute to the toxin production in all the reservoirs.

All the sites showed regular cyanobacterial, and increasingly dinoflagellate, blooms. The dinoflagellate blooms are exclusively *Ceratium hirundinella* and these dinoflagellate blooms are associated with extremely high Chl *a* concentrations in the five Reservoirs. *Ceratium* dominated Bon Accord and Klipvoor Dams during the 5-year study period, while Hartbeespoort and Rietvlei experienced *Ceratium* blooms during 2000 and again in 2005. Roodeplaat Dam experienced for the first time a *Ceratium* bloom in 2005.

The PCA of CANOCO indicated that the reservoirs are similar in both algal community and physico-chemical variables. This support the assumption that there are no major differences between the reservoirs and that the data could be combined in the modelling exercises that followed. The assumption that a reservoir is one site, and taking samples at five different sites as compared to the regular sampling in the reservoir, to determine toxin presence, may have led to

discrepancies in the cyanobacteria association with toxin results. However, it does give one a better perspective of when to potentially expect cyanobacterial toxins to be present in these hypertrophic reservoirs. This will enable Management to issue warnings in time, regarding the potential health risk that cyanobacterial blooms may pose to recreational users.

The multivariate analysis shows that of the environmental factors, temperature is the most important factor in the development of cyanobacteria in these hypertrophic systems. Temperature is, therefore, one of the main driving forces of the blooms and is a good indictor of climate and weather impacts on a system. According to the multivariate analysis, the main driving force that correlates quite well with the TM is the DIP concentrations in these systems. This variable will thus play an important role in the prediction of the cyanobacterial toxin production (TM).

The depth distribution of the toxins measured in the Hartbeespoort and Roodeplaat Reservoirs indicated that during the periods of excessive *Microcystis* biovolume, toxins are often found all the way through the water column. This may necessitate WCW's that produce potable water, to monitor at different depths, on the condition that they have the option to withdraw water at different depths, as is the case at the Roodeplaat Reservoir.

Since the multivariate statistical analysis with the CANOCO program showed that the five reservoirs were so similar that it enabled us to combine all the data for further use in applying the modelling techniques, especially the Artificial Neural Network modelling technique and the Hybrid Evolutionary Algorithm development technique. The most important environmental variables that were also used in further predictive capability development were water surface temperature (Tsurf), nutrients (TP, TN, DIP &DIN), underwater light conditions (Secchi depth readings) and existing biovolume (Chl *a*).

A number of different types of models were tested to determine the applicability to South African conditions:

(1) The relatively simple Vollenweider model was tested for Hartbeespoort, Klipvoor and Roodeplaat reservoirs. This model is easy to apply and provide a manager with a quick answer and relatively little information is needed to apply the model. It also provides the manager with the possibility of testing different management scenarios. However, over

and under predictions of 30% to 300% (IETC 2000) is possible and acceptable within the context of the model.

(2) The simulation library, SALMO-OO, allowing forecasting abundances of blue-green algae, green algae and diatoms in response to eutrophication control scenarios for Hartbeespoort, Klipvoor and Roodeplaat reservoirs. The model is data intensive as it works with a ten-day time step and inflow and outflow data are necessary to apply the model. This information was not available for the Bon Accord and Rietvlei reservoirs. This model takes the complex limnological characteristics of reservoirs into consideration and it supplies the manager with a tool to test different management scenarios to assist in decision-making. The results were, however, only partly successful with large over and under predictions, even after the growth equations were adapted.

(3) With the Artificial Neural Network modelling techniques, both the supervised multilayered feed forward neural network and the non-supervised self-organising map methods were tested. The Software Package Forecaster XL was used to predict the biovolume of the non-cyanotoxin producing group and the cyanotoxin producing group of algae. The five-year data set was used. The model provides visual successes but the strict tolerances, set by the model, to determine acceptable prediction as part of the outcome of the model, may be a problem to validate the results and ensure that an acceptable number of good predictions were found. The main disadvantage of the Forester XL model is that the outcomes do not provide the modeller with the algorithms and validation methods used, because of the black box nature of the technique.

The Self Organising Mapping (SOM) method was used on the Roodeplaat Dam data for a period of 20 years. This was done to determine the differences of the before and after conditions of the influence of the Zeekoegat WCW. The seasonal changes and increases in nutrients and Chl *a* concentration are pronounced. This modelling technique is thus applicable to investigate before and after scenarios. This is more of a knowledge development tool than a predictive tool. Only one test was done and the applicability to the impacts of other environmental variables on the development of harmful algal blooms needs to be further investigated.

(4) The RULE set discovery by Hybrid Evolutionary Algorithm (HEA) was tested on different scenarios of real time and 7-days forward, 14-days forward, 21-days forward and 28-days forward forecasting of the abundance of the cyanobacterium, *Microcystis* in the Hartbeespoort, Rietvlei and Roodeplaat reservoirs. The developed rule sets are highly applicable to the hypertrophic reservoirs of South Africa. These methods need, however, to be tested in other reservoirs to determine the applicability under different trophic status and different climatic conditions.

The same method was used to develop a real time algorithm for the dinoflagellate, *Ceratium hirundinella*, for the Hartbeespoort, Rietvlei and Roodeplaat reservoirs. The developed RULE set was then tested on the Bon Accord and Klipvoor Reservoirs that both experienced extreme blooms during the study period. This application was highly applicable to these reservoirs. This further suggests that the developed RULE set may potentially be applicable to reservoirs in other climatic areas of South Africa. This needs further investigation.

6.2 Recommendations

The study showed that eutrophication and the associated problems is a real threat to South African fresh water resources. The list of recommendations need to be taken further by a number of stakeholders, e.g. the Department of Water Affairs and Forestry, future CMA's, Universities and other researchers:

a) Monitor all the necessary variables for future modelling exercises. These include temperature and volumes at inflow. Some of the reservoirs lacked essential variables and certain models, e.g. SALMO-OO could not be tested on Bon Accord and Rietvlei reservoirs due to a lack of flow data. WST is one of the most important variables to measure in-lake and should be included in all monitoring programmes, in addition to the regular chemical and biological monitoring.
b) Include TM's monitoring in impacted fresh water resources at least during the summer periods to enable resource managers to issue warnings to all potential impacted stakeholders and also to provide data for future modelling exercises.

c) Initiate and test available management options to minimise the serious eutrophication levels in the Crocodile-West/Marico Water Management Area. Information and data collected during this testing should be available on the DWAF database for future use.

d) The cause and effects for the changing composition of the phytoplankton of these five reservoirs, that was previously dominated by *Microcystis* during summer periods and by diatoms during winter periods to change to *Ceratium* dominance need to be investigated.

e) Reservoirs that have the option of extracting water at different depths and that are subject to eutrophication related problems should monitor regularly at different depths, to determine the best depth for abstraction. This will enable them to abstract the best quality water and it will minimise the associated financial implications of treating water with potentially bad quality.

f) Of all the modelling methods used the Hybrid Evolutionary Algorithm (HEA) RULE set development proved to very effective. It is recommended that capacity is developed in South Africa regarding the use of this technique. Funding should be made available for the implementation and use of this technique in all research spheres, as the method is applicable to any type of numerical data.

g) Manage the risk imposed by the cyanobacterial blooms and the associated toxins produced in the water resources, on drinking water facilities and the health of recreational users.

h) In view of the successes of the modelling results the next step would be to develop short-term forecasting tools, for the algal blooms of *Microcystis* and *Ceratium,* with on-line water quality monitoring for early-warning and real-time forecasting for reservoir managers.

CHAPTER 7

REFERENCES

Allanson BR (2004) Limnology in South Africa: Past and present status and future needs. In: Limnology in Developing Countries Volume 4 Gopal B and Wetzel RG Eds. International Association for Theoretical and Applied Limonology. International Scientific Publications. New Delhi. India. ISBN 81-86047-23-9. 244pp.

Andradóttir HO and Nepf HM (2000) Thermal mediation by littoral wetlands and impacts on lake intrusion depth. *Water Resources Research*, **36**(3), 725-735.

Aoyama K, Kawamura N, Saitoh M, Magara Y and Ishibashi Y (1995) Interactions between bacteria-free *Anabaena macrospora* clone and bacteria isolated from unialgal culture. *Wat. Sci. Technol.* **31** (11), 121 – 126.

Arhonditsis, G.B., and M.T. Brett 2005 Eutrophication model for Lake Washington (USA): Part I - Model description and sensitivity analysis. *Ecological Modelling.* 187, 140-178.

Atlas RM and Bartha R (1998) *Microbial Ecology: fundamentals and applications*. Forth Edition. Addison Wesley Longman, Inc. Menlo Park, California.

Bäck T, Fogel DB and Michalewicz Z (1997) *Handbook on Evolutionary Computation*. Oxford University Press.

Barbanti A, Bergamini MC, Frascari F, Miserocchi S, Ratta M and Rosso G (1995) Diagenetic processes and nutrient fluxes at the sediment-water interface, northern Adriatic Sea, Italy, *Mar. Freshwater Res.*, **46**, 55-67.

Bendorf J and Recknagel F (1982) Problems of application of the ecological model SALMO to lakes and reservoirs having various trophic states. *Ecological Modelling,* **17** (2), 129-145.

Blenckner T, Pettersson K & Padisák J (2002) Lake plankton as tracer to discover climate signals. *Verh. Internat. Verein. Limnol.* **28** (3), 1324-1327.

Blomqvist P, Petterson A and Hyenstrand P (1994) Ammonium-nitrogen: A key regualtory factor causing dominance of non-nitrogen fixing cyanobacteria in aquatic systems. *Arch Hydrobiol.*, **132**, 141-164.

Burroughs W Ed. (2003) *Climate into the 21st Century.* World Meteorological Organization 2003. University Press, Cambridge. ISBN 0 521 79202 9. 240pp.

Cao H, Recknagel F, Joo G-J, and Kim D-K (2005) Discovery of predictive rule sets for chlorophyll-a dynamics in the Nakdong River (Korea) by means of the hybrid evolutionary algorithm HEA. *Ecological Informatics*, **1**, 43-53.

Cao H, Recknagel F, Welk A, Kim B and N Takamura, 2006. Hybrid Evolutionary Algorithm for Rule Set Discovery in Time-Series Data to Forecast and Explain Algal Population Dynamics in Two Lakes Different in Morphometry and Eutrophication. In: Recknagel, F. (ed.), 2006. *Ecological Informatics. 2nd Edition.* Springer-Verlag Berlin, Heidelberg, New York, 347-367.

Catalan J and Fee EJ (1994) Interannual variability in limnetic ecosystems: origin, patterns and predictability. In: Margalef R (ed.) *Limnology Now: a Paradigm of Planetary Problems*, 81-94. Elsevier.

Cetin L, Zhang B and Recknagel F (2005) Process-based simulation library SALMO-OO for lake ecosystems. http://www.mssanz.org.au/modsim05/papers/cetin.pdf.

Chriswell B, O'Halloran K and Wall J (1995) The role of sulfur in phosphorus release from sediments. *RACI Proceedings 4th Environmental Chemistry Conference*, Darwin, July 1995, EO16-1 – 3.

Chutter FM (1989) *Evaluation of the 1mg/l phosphate-P standard on the water quality and trophic state of Hartbeespoort Dam.* WRC Report No 181/1/89. 69pp.

Chutter FM and Rossouw JN (1991) *The management of phosphate concentrations and algae in Hartbeepoort Dam.* WRC Report No 289/1/91. xii + 37pp.

Codd GA, Azevedo SMFO, Bagchi SN, Burch MD, Carmichael WW, Harding WR, Kaya K and Utkilen HC (2005) CYANONET A global network for cyanobacterial bloom and toxin risk management. IHP-IV. Technical Documents in Hydrology, **76**. UNESCO. Paris. 138pp.

Cooke GD, Welch EB, Peterson SA and Newroth PR (1993) Restoration and management of lakes and reservoirs. Second edition. Lewis publishers, Boca Raton. 548pp.

Coombe D and Bell P (eds.) (1965) *Strasburger's Textbook of Botany.* Longmans. Bungay, Suffolk. 846pp.

Daintith J (ed.) (1991) A Concise dictionary of Chemistry. New Edition. Oxford. Oxford University Press.314pp.

Department of Water Affairs (1980) *Bon Accord Dam Capacity Determination.* Report No. S 0123/10/8000. Department af Water Affairs and Forestry. Pretoria.

Department of Water Affairs and Forestry (1997) *Resource Quality Services, Biology Laboratory. Method 2002005, Algal identification Method.* Revision 4, 8 pages, Approval date 2003-10-09, Pretoria. South Africa.

Department of Water Affairs and Forestry (2002) *National Eutrophication Monitoring Programme. Implementation Manual.* Compiled by K Murray, M du Preez and CE van Ginkel. Pretoria. South Africa.

Department of Water Affairs and Forestry (2003a) *Resource Quality Services, Biology Laboratory. Method 2002005, Chlorophyll a – Spectrophotometric Method.* Revision 4, 8 pages, Approval date 2003-10-09, Pretoria. South Africa.

Department of Water Affairs and Forestry (2003b) *Resource Quality Services, Biology Laboratory. Method 2006001, Quantiplate Microcystin Method.* Revision 1, 5 pages, Approval date 2003-10-17, Pretoria. South Africa.

Department of Water Affairs and Forestry (2006a) *Resource Quality Services, Macro Laboratory. Method 0007003, Automated determination of dissolved nitrogen (Kjeldahl).* Revision 4, 8 pages, Approval date 2006-07-12, Pretoria. South Africa.

Department of Water Affairs and Forestry (2006b) *Resource Quality Services, Macro Laboratory. Method 0015003, Automated determination of total phosphorus as phosphomolybdate.* Revision 5, 10 pages, Approval date 2006-07-12, Pretoria. South Africa.

DWA, (1986). *Management of the Water Resources of the Republic of South Africa*, Department of Water Affairs, PRETORIA.

Dillon PJ and Rigler FH (1974) A test of a simple nutrient budget model predicting the phosphorus concentration in lake water. *J. Fish. Res. Bd Can.*, **31**, 1771–1778.

Dillenberg HO and Dehnel MK (1960) Toxic water bloom in Saskatchewan 1959. *Can. Med. Assoc.J.*, **83**, 1151-1154.

Donnelly TH, Grace MR and Hart BT (1997) Algal blooms in the Darling-Barwon River, Australia. *Water, Air and Soil Pollution*, **99**, 487-496.

Downing and Van Ginkel CE (2003) *Cyanobacterial Monitoring 1990-2000: Evaluation of SA data.* WRC Report No. 1299/1/03. ISBN 1-77005-012-6. 45pp.

Dunst *et al.* (1974) Survey of lake rehabilitation techniques and experiences. Technical Bulletin No 75. Department of Natural Resources, Madison, Wisconsin.

Du Preez H and Van Baalen L (2006) *Generic incident management framework for toxic blue-green algal blooms, for application by potable water suppliers.* WRC Report No. TT 263/06. South Africa. 65pp.

Elser JJ, Marzolf ER and Goldman CR (1990) Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: a review and critique of experimental enrichments. *Can. J. Fish. Aquat. Sci.*, **47**, 1468-1477.

Eynard F, Mez K and Walther J (2000) Risk of cyanobacterial toxins in Riga waters (Latvia). *Wat. Res.*, **34**(11), 2979-2988.

Falconer IR (2004) *Cyanobacterial toxins of drinking water supplies, cylindrospermopsins and microcystins*. CRC Press. London. 279pp.

Frank CAP and Wolf H.-U. (2002) Assessment of causes, frequency and toxicological significance of toxic cyanobacterial blooms in recreational waters in south-west Germany. *Verh. Internat. Verein. Limnol.*, **28**(4), 1775-1778.

Fisher MM and Reddy KR (2001) Phosphorus flux from wetland soils affected by long-term nutrient loading. *J. Environ. Qual.*, **30**, 261-271.

Fogg GE (2002) Harmful algae – a perspective. *Harmful Algae*, **1**, 1-4.

Gerber A, Van Ginkel CE, Seloana, A and Hohls B (2004a) *Water Resource Quality Monitoring. Volume 2 Eutrophication Monitoring*. Department of Water Affairs and Forestry. Government Printers, Pretoria. ISBN: 0-621-35124-2.

Gerber A, Van Ginkel CE, Seloana, A and Hohls B (2004b) *Water Resource Quality Monitoring. Volume 3 Toxic Algae*. Department of Water Affairs and Forestry. Government Printers, Pretoria. ISBN: 0-621-35125-3.

Gillingham AG and Thorrold BS (2000) A review of New Zealand research measuring phosphorus in runoff from pasture. *J. Environ. Qual.*, **29**, 88-96.

Gleick PH, Singh A and Shi H (2001) *Emerging threats to the world's Freshwater Resources*. A report of the Pacific Institute for studies in development, Environment and security, Oakland, California. 59pp.

Goedkoop W and Törnblom E (1996) Seasonal fluctuations in benthic bacterial production in Lake Erken: The significance of major abiotic factors and sedimentation events. *Arch. Hydrobiol. Ergebn. Limnol.*, **48**, 197-205.

Goldberg DE (1989) Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, Massachusetts.

Goodwin TW (1974) Carotenoids and biliproteins. *In* W.D.P. Stewart, ed. *Algal Physiology and Biochemistry*. University of California Press. Berkeley. 176-205.

Graham PM, Dickens CWS and Mbowa S (1998) Modelling the water quality in impoundments within the Umgeni Water operational area and the consequences for potable water treatment costs. Water Research Commission Report No. WRC 615/1/98, PRETORIA, South Africa.

Greenberg HJ (2003) Mathematical programming glossary supplement: Tolerances. http://glossary.computing.society.informs.org/notes/tolerances.pdf

Grobler DC (1985) Phosphate budget models for simulating the fate of phosphorus in South African Reservoirs. *Water SA,* **11**, 219-230.

Grobler DC and Silberbauer MJ, (1984) *Impact of Eutrophication Control Measures on the Trophic Status of South African Impoundments.* Water Research Commission Report No. WRC 130/1/84, PRETORIA, South Africa.

Guerrero MG and Lara C (1987) Assimilation of inorganic nitrogen. In: Fay P and van Baalen C (eds.) The Cyanobacteria. Elsevier, Amsterdam. Pp 163-186.

Guven B and Howard A (2006) Modelling the growth and movement of cyanobacteria in river systems. *Science of the Total Environment*, **368**, 898-908.

Ha K, Jeong KS and Joo G-J (2002) Inter-annual variability of the summer *Microcystis* bloom in the lower Nakdong River (S. Korea). *Verh. Internt. Verein. Limnol.*, **28**(4), 1902-1907.

Hansson L-A, Rudstam LG, Johnson TB, Soranno P and Allen Y (1994) Patterns in algal recruitment from sediment to water in a dimictic, eutrophic lake. *Can. J. Fish. Aquat. Sci.*, **51**, 2825-2833.

Harding WR (1997) New regional emergence of cyanobacterial toxicosis. *Harmful Algae News*, **16**, 12-13.

Harding WR (2006) A research strategy for the detection and management of algal toxins in water sources. WRC TT 277/06.South Africa. ISBN 1-77005-461-8. 42pp.

Harding WR and Paxton BR (2001) *Cyanobacteria in South Africa: A review.* Report No TT153/01 to the Water Research Commission. Pretoria. ISBN 1 86845 7745.

Harding WR, Thornton JA, Steyn G, Panuska J and Morrison IR (2004a) Hartbeespoort Dam Remediation Project (Phase I). Action Plan (Volume I) Final Report. Department of Agriculture, Concervation, Environment and Tourism Project 58/2003. Content Solutions. Product Code CS 00002466. 158pp.

Harding WR, Thornton JA, Steyn G, Panuska J and Morrison IR (2004b) Hartbeespoort Dam Remediation Project (Phase II). Action Plan (Volume II) Final Report. Department of Agriculture, Concervation, Environment and Tourism Project 58/2003. Content Solutions. Product Code CS 00002472. 73pp.

Harris GP (1986) *Phytoplankton Ecology: Structure, function and fluctuation.* Chapman and Hall. London. 384pp.

Hart R and Hart RC (2006) Reservoirs and their management: A review of the literature since 1990. WRC Report no. KV 173/06. South Africa. ISBN 1-77005-418-9. 288pp.

Heathwaite L, Sharpley A and Gburek W (2000) A conceptual approach for integrating phosphorus and nitrogen management at watershed scale. *J. Environ. Qual.*, **29**, 158-166.

Hesketh N and Brookes PC (2000) Development of an indicator for risk of phosphorus leaching. *J. Environ. Qual.*, **29**, 105-110.

Hohls BC, Belcher T, Kempster PL and VAN GINKEL C. (1999). *Assessment of the Water Quality in the Roodeplaat Dam since 1990. Draft* Report No. N/A200/00/DEQ/2199. Institute for Water Quality Studies, Department of Water Affairs and Forestry, PRETORIA, South Africa.

Hohls BC, Quibell G, Du Plessis BJ and Belcher T (1998). *Assessment of the implementation of the phosphate standard at the Baviaanspoort and the Zeekoegat Water Care Works. Report No. N/A230/01/DEQ/0797.* Institute for Water Quality Studies, Department of Water Affairs and Forestry. PRETORIA, South Africa.

Horne AJ and Goldman CR (1994) *Limnology.* Second edition.McGraw-Hill Book Co. New York. ISBN 0-07-113359-3. 576pp.

Hongping, P and Jianyi M (2002) Study on the algal dynamic model for West lake Hangzhou, *Ecological Modelling*, **148**, 67 – 77.

Howarth RW, Marino R and Cole JJ (1988a) Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biochemical controls. *Limnol. Oceanogr.*, **33**, 688-701

Howarth RW, Marino R, Lane J and Cole JJ (1988b) Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 1. Rates and importance. *Limnol. Oceanogr.*, **33**, 669-687

Hutchinson GE (1975) *A treatise on Limnology Volume 2, Part 2 – Chemistry of Lakes.* John Wiley & Sons, New York. 541 – 1015.

International Environmental technology Centre (IETC) (2000) *Planning and management of lakes and reservoirs: An integrated approach to eutrophication: Students Guide.* Technical Publication Series 12. UNEP/Earthprint. ISBN 928071810x.

IPCC (1992) *1992 IPCC Supplement: Scientific assessments of climatic change, p. 24. Report preparedby Working Group I, WMO-UNEP.* Houghton JT, Callander BA and Varnay S (eds.). Cambridge University Press. Cambridge.

Ishikawa K, Tsujimura S, Vincent WF, Kumagai M and Nakahara H (2002) Growth of bloom-forming cyanobacteria under low nutrient conditions. *Verh. Internat. Verein. Limnol.*, **28**(3), 1616-1621.

Izaguirre G and Taylor WD (1995) Geosmin and 2-methylisoborneol production in a major aqueduct system. *Water Sci. Technol.* **31** (11), 41 – 48.

Izaguirre G and Taylor WD (1998) A *Pseudoanabaena* species from Castaic Lake, California, that produces 2-methylisboborneol. *Water Sci. Technol.* **32** (5), 1673 – 1677.

John DM, Whitton BA and Brook AJ (eds.) (2003) *The freshwater algal flora of the British Isles. An identification guide to freshwater and terrestrial algae.* First Edition Reprinted. Cambridge University Press. Cambridge. ISBN 0 521 77051 3. 702pp.

Jordan TE, Correll DL and Weller DE (1997) Relating nutrient discharges from watersheds to land use and streamflow variability. *Water Resources Research*, --, 2579-2590.

Karul C and Soyupak S (2006) A comparison between neural network based and multiple regression models for chlorophyll-*a* estimation. In Recknagel, F. (ed.), 2006. *Ecological Informatics. 2nd Edition.* Springer-Verlag Berlin, Heidelberg, New York, 309-323.

Kellerman TS, Coetzer JAW, Naudé TW and Botha CJ (2005) *Plant poisonings and mycotoxicoses of livestock in Southern Africa.* Seconbd Edition. Oxford University Press. Cape Town. ISBN 978 0 19 57613 4. 310pp.

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

Kohonen T. (1988). *Self-Organization and associative memory.* New York: Springer-Verlag.

Kohonen T (1995) *Self-organising maps.* Springer-Verlag. Heidelberg.

Kohonen T (2007) *Kohonen Networks.* Accessed 2007-04-03 http://www.cs.bham.ac.uk/~jlw/sem2a2/Web/Kohonen.htm

Kunz RP, Schulze RE and Scholes RJ (1995) An approach to modelling spatial changes of plant carbon:nitrogen ratios in southern Africa in relation to anticipated global climate change. *Journal of Biography*, 22, 401-408.

Lawrence I, Bormans M, Oliver R, Ransom G, Sherman B, Ford P and Wasson B (2000) *Physical and nutrient factors controlling algal succession and biomass in Burrinjuck Reservoir.* Technical Report for the Cooperative Centre for Freshwater Ecology. ISBN 0967704844. 134pp.

Lee RE (1999) *Phycology.* Third Edition. Cambridge University Press. Cambridge. ISBN 0 521 63883 6. 614pp.

Lepš J and Šmilauer P (2003) Multivariate analysis of ecological data using CANOCO. First Edition. Cambridge University Press. Cambridge. ISBN 0 521 89108 6. 269pp.

Mendelsohn L (1993) *Training neural networks.* Accessed 2008-04-01 http://www.tradertech.com/training.asp

Metcalf JS, Bell SG and Codd GA (2000) Production of novel polyclonal antibodies against the cyanobacterial toxin microcystin-LR and their application for the detection and quantification of microcystins and nodularin. *Wat. Res.*, **34**(10), 2761-2769.

Midgley DC, Pitman WV (1978) *A depth-duration-frequency diagram for point rainfall in Southern Africa.* HRU Report 2/78. University of Witwatersrand. Johannesburg.

Mitchell JFB (1991) *Simulated climate change over southern Africa in a high resolution mixed-layer model experiment.* Unpublished paper, Hadley Centre for Climate Prediction and Research. Meteorological Office. Bracknell. UK. (mimeographed)

Miwa M and Morizane K (1988) Effects of chelating agents on the growth of blue-green algae and the release of geosmin. *Water Sci. Technol.* **20** (8/9), 197 – 203.

Muttil N and Lee JHW (2005) Genetic programming for analysis and real-time prediction of coastal algal blooms. *Ecological Modelling*, **189**, 363 – 376.

Naes H and Post AF (1988) Transient states of geosmin, pigments, carbohydrates and proteins in continuous cultures of *Oscillatoria brevis* induced by changes in nitrogen supply. *Arch. Microbiol.* **150**, 333 – 337.

Naes H, Utkilen HC and Post AF (1988) Factors influencing geosmin production by the cyanobacterium *Oscillatoria brevis. Water Sci. Technol.* **20** (8/9), 125 – 131.

Naes H, Utkilen HC and Post AF (1989) Geosmin production in the cyanobacterium *Oscillatoria brevis. Arch. Microbiol.* **151**, 407 – 410.

National Rivers Authority (1990) *Toxic blue-green algae.* Water Quality Series. London.

Negoro T, Ando M and Ichikawa N (1988) Blue-green algae in Lake Biwa which produce earthy-musty odours. *Water Sci. Technol.* **20** (8/9), 117 – 123.

NIWR (1985) *The limnology of Hartbeespoort Dam.* South African National Scientific Programmes Report No. 110. 269 pp.

Ochao de Alda JAG, Tapia MI, Fabrice F, Llama MJ and Serra JL (1996) Changes in nitrogen source modify distribution of excitation energy in the cyanobacterium *Phormidium laminosum. Physiol. Plant.*, **97**, 69-78.

Oliver RL and Ganf GG (2000) *Freshwater blooms.* In: Whitton BA and Potts M (Eds.) (2000)*The Ecology of Cyanobacteria. Their diversity in time and space.*Kluwer Academic Publishers. 149-194.

Organisation for Economic Cooperation and Development (OECD) (1982) *Eutrophication of waters: monitoring, assessment and control.* Technical Report, OECD, Paris.

Orr PT and Jones GJ (1998) Relationship between microcystin production and cell division rates in nitrogen-limited *Microcystis aeruginosa* cultures. *Limnol. Oceanogr.*, **43**, 1604-1614.

Park RA, O'Neill RV, Bloomfield JA, Shugart HH, Booth RS, Goldstein RA, Mankin JB, Koonce JF, Scavia D, Adams MS, Clesceri LS, Colon EM, Settman EH, Hoopes JA, Huff DD, Katz S, Kitchell JF, Koberger RC, La Row EJ, McNaught DC, Petersohn L,

Titus JE, Weiler PR, Wilkinson JW, Zahorcak CS (1974) A generalized model for simulating lake ecosystems. *Simulation* 33-50.

Perkins RG and Underwood GJC (2000) Gradients of chlorophyll *a* and water chemistry along an eutrophic reservoir with determination of the limiting nutrient by *in situ* nutrient addition. *Wat. Res.*, **34**(3), 713-724.

Perkins RG and Underwood GJC (2001) The potential for phosphorus release across the sediment-water interface in an eutrophic reservoir dosed with ferric sulphate. *Wat. Res.*, **35**(6), 1399-1406.

Peters NE and Meybeck M (2000) Water quality degradation effects on freshwater availability: Impacts to human activities. *Water International*, **25**(2), 185-193.

Pick FR and Lean DRS (1987) The role of macronutrients (C, N, P) in controlling cyanobacterial dominance in temperate lakes. *N Z J Mar. Freshwater Res.*, **21**, 425-434.

Pillay M and Buckley CA (2001) *Detergent phosphorus in South Africa: Impact on eutrophication with specific reference to Mgeni Catchment.* WRC Report No. 465/1/01. Pretoria. ISBN 1 86845 733 8. 28pp.

Pilotto LS, Douglas RM, Burch MD, Cameron S, BeersM, Rouch GR, Robinson P, Kirk M, Cowie CT, Hardiman S, Moore C and Attewell RG (1997) Health effects of recreational exposure to cyanobacteria (blue-green algae) during recreational water-related activities. *Aust.N. Zealand J. Public Health*, **21**, 562-566.

Pouria S, De Andrade A, Barbosa J, Cavalcanti RL, Barreto VT, Ward CJ (1998) Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. *Lancet*, **352**, 21-26.

Rae B, Moollan RW and Clark RC (1999) *Algal Toxins in drinking water supplies.* Water research Commission Report No. 549/1/99, Pretoria, South Africa.

Recknagel F (1997) ANNA – Artificial neural network model for predicting species abundance and succession of blue-green algae. *Hydrobiologia*, **349**, 47-57.

Recknagel F and Benndorf J (1982) Validation of ecological simulation model SALMO. *Int. Revue Ges. Hydrobiol.*, **67**, 113-125.

Recknagel F., Hosomi, M., Fukushima, T. and Kong, D-S (1995) Short- and long-term control of external and internal phosphorus loads in lakes: a scenario analysis. *Water Research.* **29** (7), 1767-1779.

Recknagel F, Petxholdt T, Jacke O and Krusche F (1994) Hybrid expert system DELAQUA: a toolkit for water quality control of lakes and reservoirs. *Ecol. Model.*, **71**, 17-36.

Recknagel F, Welk A, Kim B and Takamura N (2006) Artificial neural network approach to unravel and forecast algal population dynamics of two lakes different in morphometry

and eutrophication. In Recknagel, F. (ed.), 2006. *Ecological Informatics. 2nd Edition.* Springer-Verlag Berlin, Heidelberg, New York, 325-345.

Reynolds CS (1984) Phytoplankton periodicity: the interactions of form, function and environmental variability. *Freshwater Biol.*, **14**, 111-142.

Reynolds CS (1992) Eutrophication and the management of planktonic algae: what Vollenweider couldn't tell us. In: Sutcliffe DW and Jones JG (eds) *Eutrophication: Research and application to water supply*, 4-29. Freshwater Biological Association, Ambleside, UK.

Robarts RD (1984) Factors controlling primary productivity in a hypertrophic lake (Hartbeespoort Dam, South Africa. *J. Plankt. Res*, **6** (1), 91-105.

Robson BJ and Hamilton DP (2004) Three-dimensional modelling of a *Microcystis* bloom event in the Swan River estuary, Western Australia. *Ecological Modelling,* **174**, 203-222.

Rossouw JN (1990) T*he development of management orientated models for eutrophication control.* WRC Report No. 174/1/90. Water Research Commission. Pretoria, South Africa.

Rossouw JN (2000) An extension of management orientated models for eutrophication control. WRC Report No. 266/1/01. Water Research Commission. Pretoria, South Africa. 69pp.

Rossouw JN and Görgens AHM (2005) Knowledge review of modelling non-point source pollution in agriculture from field to catchment scale.

Rossouw JN and Kelly H (1989) *REMDSS – A decision support system for eutrophication control: User's manual.* Draft WRC Report, Pretoria, South Africa.

Rumelhart DE and McClelland JL (Eds.) (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press.

Ruan HD and Gilkes RJ (2000) Phosphorus accumulation in farm ponds and dams in Southwestern Australia. *J. Environ. Qual.*, **29**, 1875-1881.

Sander BC and Kalff J (1993) Factors controlling bacterial production in marine and freshwater sediments. *Microbial. Ecol.*, **26**, 79-99.

Sas H (1989) *Lake restoration by reduction of nutrient loading: expectations, experiences, extrapolations.* Academia Verlag Richarz, St Augustin. 497pp.

Scheffer M, Rinaldi S, Gragnani A, Mur LR and Van Nes EH (1997) On the dominance of filamentous cyanobacteria in shallow, turbid lakes. *Ecology*, **78**, 72-282.

Sharpley AN and Rekolainen S (1997) Phosphorus in agriculture and its environmental implications. P 1-54. In H Tunney *et al.* (ed.). *Phosphorus loss from soil to water.* CAB Int., Wallingford, UK.

Schoumans OF and Groenendijk P (2000) Modeling soil phosphorus levels and phosphorus leaching from agricultural land in the Netherlands. *J. Environ. Qual.*, **29**, 111-116.

Schulze RE (1984) *Depth-duration-frequency studies in Natal based on digitised data. South African National Hydrology Symposium.* Technical Report TR 119. Department of Environment Affairs. Pretoria. RSA.

Schulze RE (1997) *South African atlas of agrohydrology and climatology.* WRC Report No. TT82/96. Pretoria, South Africa.

Schulze RE (2000) Modelling hydrological responses to land use and climate change: A Southern African Perspective. *AMBIO*, **29** (1), 12-22

Seip KL, Sas H and Vermu S (1992) Nutrient-chlorophyll trajectories across trophic gradients. *Aquatic Science*, **54**, 58-76.

Shapiro J (1990) Current beliefs regarding dominance by blue-greens: The case for the importance of CO2 and pH. *Verh. Int. Verein. Limnol.*, **24**, 38-54.

Sigwort EA (1957) Control of odor and taste in water supplies. *J.AWWA* **49**, 1507 – 1520.

Smith KA (2002) *Neural Networks: An Introduction.* Idea Group Publishing. http://www.igi-pub.com/downloads/excerpts/1930708319BookEx.pdf. 24pp.

Smithers JC, Pegram GGS and Schulze RE (2002) Design rainfall estimation in South Africa using Bartlett-Lewis rectangular pulse rainfall models. *Journal of Hydrology*, **258** (1-4), 83-99.

Smithers JC and Schulze RE (2006) *Development and evaluation of techniques for estimating short duration design rainfall in South Africa.* WRC Report No. 861/1/00. Water Research Commission, Pretoria, RSA. 356pp.

Starfield AM, Smith KA and Bleloch AL (1990) *How to model it: problem solving for the computer age.* McGraw Hill, New York. 206pp.

Steel JA (1997) Scope and limitation in algal modelling – an example from the Thames Valley reservoirs. *Hydrobiologia*, **349**, 27-37.

Steyn DJ, Scott WE, Toerien DF and Visser JH (1975) The eutrophication levels of some South African impoundments. I Rietvlei Dam. *Water SA*, **2**, 45-52.

Stofberg CE (1983) 'n Ondersoek na die rol van soplankton in Roodeplaatdam. MSc Thesis on Department of Zoology/Entomology. University of the Free State. 102pp.

Swanepoel A, Du Preez H, Schoeman C, Janse van Vuuren S and Sundram A (2007) *Condensed laboratory methods for monitoring phytoplankton, including cyanobacteria in South African freshwaters.* WRC Project K5/1533, South Africa. 107pp.

Talling JF (1957) The phytoplankton population as a compound photosynthetic system. *New Phytol*,.**56**, 133-149.

Tandeau de Marsac N (1977) Occurrence and nature of chromatic adaptation in cyanobacteria. *J. Bacteriol.*, **130**. 80-91.

Tandeau de Marsac N and Houmard J (1993) Adaptation of cyanobacteria to environmental stimuli: new steps towars molecular mechanisms. *FEMS Microbiol Rev.*, **104**, 119-190.

Tang EPY and Vincent WF (1999) Strategies of thermal adaptation by high-latitude cyanobacteria. *New Phytol.* **142**, 315-323.

Taylor FJR and Pollingher U (1987) Ecology of Dinoflagellates. In: FJR Taylor, ed. The Biology of Dinoflagellates. Blackwell Scientific Publs., Oxford. pp 398-529.

Thornton JA and Walmsly RD (1981) Applicability of phosphorus budget models to southern African man-made lakes. *Hydrobiol.*, **89** (3), 237-245.

Toerien DF, Hyman KL and Bruwer MJ (1975) A preliminary trophic state classification of some South African impoundments. *Water SA,* **1**, 15-23.

Turner PC, Gammie, AJ, Hollinrake K and Codd GA (1990) Pneumonia associated with cyanobacteria. *Br. Med. J.*, **300**, 1440-1441.

Turpin DH (1991) Effects of inorganic N availability on algal photosynthesis and carbon metabolism. *J. Phycol.*, **27**, 14-20.

Van Ginkel, C.E. (2004). *A National Survey of the incidence of cyanobacterial blooms and toxin production in major impoundments.* Internal Report No. N/0000/00/DEQ/0503. Resource Quality Services, Department of Water Affairs and Forestry. Pretoria.

Van Ginkel CE and Conradie B (2001) *Potentially toxic algal incident in the Orange River, Northern Cape, 2000.* Internal Report No. N/D801/12/DEQ/0800. Resource Quality Services (Institute for Water Quality Studies), Department of Water Affairs and Forestry, Pretoria.

Van Ginkel CE, Hohls BC, Belcher A, Vermaak E and Gerber A (2000a) *Assessment of the Trophic Status Project Main Report.* Report number N/0000/00/DEQ/1799 of the Institute for Water Quality Studies, Department of Water Affairs. Pretoria, 314pp.

Van Ginkel CE, Hohls BC, Van Niekerk H (2000b) *Pollution incident in the Roodeplaat Dam, May 2000.* Internal Report No. N/A230/01/DEQ/0600. Institute for Water Quality Studies, Department of Water Affairs and Forestry. Pretoria. 28pp.

Van Ginkel CE, Hohls BC and Vermaak E (2001) A *Ceratium hirundinella* (O.F. Müller) bloom in Hartbeespoort Dam, South Africa. Water SA, 27 (2), 269-276.

Van Ginkel CE, Silberbauer MJ, Du Plessis S, Carelsen CIC (2006) Monitoring microcystin toxin and chlorophyll in five South African impoundments. *Verh. Internat. Verein. Limnol.*, **29**, 1611 – 1616.

Van Ginkel, C.E., Silberbauer, M.J. and Vermaak, E. (2000) The seasonal and spatial distribution of cyanobacteria in South African surface waters. *Verh. Internat. Verein. Limnol.* **27**, 871 – 878.

Van Ginkel, CE and Silberbauer, MJ (2007) Temporal trends in total phosphorous, temperature, oxygen, chlorophyll *a* and phytoplankton populations in Hartbeespoort Dam and Roodeplaat Dam, South Africa. *African Journal of Aquatic Science,* **32**, In Press.

Venter A and Herold CE (2001) *Validation of the modified MINLAKE model on Roodeplaat Dam.* WRC Report No. 785/1/99. Pretoria South Africa. 57pp.

Velickov S (2004) *Nonlinear dynamics and chaos with applications to hydrodynamics and hydrological modelling.* Taylor & Francis Group. ISBN 9058096912.

Vollenweider RA (1976) Advances in defining critical loading levels for phosphorus in lake eutrophication. *Mem. Ist. Ital. Idrobiol.*, **33**, 53–83.

Vollenweider RA (1990) Eutrophication: conventional and non-conventional considerations and comments on selected topics. *Mem. Ist. Ital. Idrobiol.*, **47**, 77-134.

Walmsley RD (2003) *Development of a Strategy to Control Eutrophication in South Africa.* Department of Water Affairs and Forestry. Water Quallity Management Series U 2.1. Pretoria. South Africa. 21pp.

Walter M, Recknagel F, Carpenter C and Bormans M (2001) Predicting eutrophication in the burrinjuck Reservoir (Australia) by means of the deterministic model SALMO and the recurrent neural network model ANNA. *Ecological Modelling*, **146** (1/3), 97-113.

Ward AK and Wetzel RG (1980) Interactions of light and nitrogen source among planktonic blue-green algae. *Arch. Hydrobiol.*, **90**, 1-25

Weier TE, Stocking CR and Barbour MG (1974) *Botany: An introduction to plant biology.* Fith edition. John Wiley & Sons. New York. 693pp.

Wetzel RG (1989) Wetland and littoral interfaces of lakes: Productivity and nutrient regulation in the Lawrence lake ecosystem. *In:* Sharitz RR and Gibbons JW (eds) *Freshwater wetlands and wildlife.* U.S. Dept. Energy, Office Sci. Technical Information, New York. 227-238.

Wetzel RG (2001) *Limnology,Lake and River Ecosystems.* Third Edition. Academic Press. New York. ISBN 0-12-744760-1. 1006pp.

Wetzel RG and Likens GE (1979) *Limnological analyses.* WB Saunders Company. Philidelphia. 357pp.

Weyhenmeyer GA, Adrian R, Gaedke U, Livingston M and Maberley SC (2002) Response of phytoplankton in European lakes to a change in the North Atlantic Oscillation. *Verh. Internat. Verein. Limnol.*, **28**, 1436-1439.

Whitton BA and Potts M (Eds.) (2000) *The Ecology of Cyanobacteria. Their diversity in time and space.*Kluwer Academic Publishers. 669pp.

Wikipedia (2007) http://en.wikipedia.org/wiki/Evolutionary_algorithm. Accessed 2007-10-03.

Withers PJA, Clay SD and Breeze VG (2001) Phosphorus transfer runoff following application of fertilizer, manure, and sewage sludge. *J. Environ. Qual.*, **30**, 180-188.

Wnorowski AU (1992) *Taste and odour forming microorganisms in South African surface waters.* WRC Report No. 320/1/93. Pretoria. South Africa. ISBN 1 874858 95 0.

Wnorowski AU (1995) Taste and odours in the aquatic environment: A Review. *Water SA* **18** (3), 203 – 214.

World Health Organization (1999) *Toxic cyanobacteria in water. A guide to their public health consequences, monitoring and management.* Edited by Chorus I and Bartram J. E&FN Spon. London. ISBN 0-419-23930-8.

World Health Organization (2000) *Global water supply and sanitation assessment 2000 report.* Publication of the World Health Organisation and the United Nations Children Fund. ISBN 92 4 156202 1.

Wu J-T and Hsu Y-M (1988)Relation of algae to earthy odors of fish in Taiwan. *Bot. Bull. Acad. Sin.* **29**, 183 – 188.

Wu J-T and Juttner F (1988) Effect of environmental factors on geosmin production by *Fisherella muscicola. Water Sci. Technol.* **20** (8/9), 143 – 148.

Xie L, Xie P, Li S, Tang H and Liu H (2003) The low TN:TP ratio, a cause or a result of *Microcystis* blooms? Water Research, **37**, 2073-2080.

Yagi Am Usui T, Fujise D and Yamamoto Y (2002) Effect of copper sulfate on growth and odor production of several microorganisms. *Verh. Internat. Zverein. Limnol.*, **28**, 1425-1428.

Zimba PV, Boue S, Chatham MA and Nonneman D (2002) Variants of microcystin in south-eastern USA channel catfish (*Ictalurus punctatus* Rafinesque) production ponds. *Verh. Internat. Verein. Limnol.*, **28**(3), 1163-1166.

Zohary T and Robarts RD (1989) Diurnal mixed layers and the long-term dominance of *Microcystis aeruginosa. J. Plankt. Res*, **11**: 25-48

Personal Communications

Carelsen, C (2003) Resource Quality Services, Department of Water Affairs and Forestry

MacKay, G (2000) Hartbeespoort Water Action Group.

Recknagel, F (2006) University of Adelaide, Australia