Forecasting of dinoflagellate blooms in warm-monomictic hypertrophic reservoirs in South Africa by means of rule-based agents

Carin van Ginkel[1]*, Hongqing Cao[2], Friedrich Recknagel[2] and Sandra du Plessis[3]

[1]*Resource Quality Services, Department of Water Affairs and Forestry, P/Bag X313, Pretoria, 0001, South Africa*

[2]*School of Earth and Environmental Sciences, University of Adelaide, 5005, Australia*

[3]*School of Botany and Environmental Sciences, North West University (Potchefstroom), South Africa*

Abstract

The occurrence of large blooms of Dinoflagellates composed primarily of *Ceratium hirundinella,* in hypertrophic reservoirs in South Africa has become more pronounced since 1999. The financial and operational impact of these blooms on the drinking water industry is high as this species produces bad tastes and odours and clogs filters.

In order to enable real-time forecasting for dinoflagellate blooms a hybrid evolutionary algorithm rule set for hypertrophic reservoirs in South Africa was developed. Data from three known hypertrophic systems were used for training and developing the rule set. Testing was done on two years respectively from the same reservoirs as the training data. The rule set was also tested on unseen data from two other hypertrophic reservoirs that are known to experience dinoflagellate blooms.

The results show that the developed rule set predicts real-time dinoflagellate blooms quite accurately and the sensitivity analysis showed 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

Keywords: rule-based agents, hybrid evolutionary algorithms (HEA), Dinoflagellates, *Ceratium*, hypertrophic warm-monomictic lakes

Introduction

Numerous studies attempt to increase the understanding of the significant input variables that drive the establishment of algal blooms of different dominant species for prediction purposes (Bowden et al., 2006; Cao et al., 2006; Jeong et al., 2006; Recknagel et al., 2006) with new and emerging computational methods. The emerging discipline of ecological informatics provides a variety of computational techniques such as fuzzy logic, cellular automata, evolutionary algorithms and adaptive agents that can be applied to non-linear, dynamic and complex limnological systems (Jeong et al., 2006).

The occurrence of large blooms of Dinoflagellates composed primarily of *Ceratium hirundinella,* in hypertrophic reservoirs in South Africa, became more pronounced since 1999 (Van Ginkel et al., 2001b) and other South African reservoirs (Hart 2006). The main characteristics of *C. hirundinella* are well known. It has a mean cell length of 180 µm. It has the ability to migrate from one water-column layer to another with the flagella, to utilise nutrients; this gives this species an advantage over other phytoplankton species and may lead to bloom-forming conditions. Furthermore the ability to assimilate both organic and inorganic phosphorus gives it an advantage over other algal species that can only utilise inorganic phosphorus. Its ability to undergo seasonal polymorphism or cyclomorphosis has been studied by numerous researchers (Hutchinson, 1957; Cranwell, 1976; Heaney, 1976; Harris et al., 1979; Heaney and Furness, 1980; Moore, 1981; Frempong, 1984; Sommer et al., 1984; Muellerelser and Smith, 1985; Padisak, 1985; Rybak, 1985; Heaney et al., 1986; Buck, 1989; James et al., 1992; Buck and Zurek, 1994; Péres-Martínez and Sánchez-Castillo, 2001; Wetzel, 2001; Grigorszky et al., 2003). All these features give *C. hirundinella* the potential to form algal blooms.

The main external growth-controlling factors are also well known, namely the impact of the light intensity in the water (Buck, 1989; Muellerelser and Smith, 1985), the optimum growth temperature (ranging from 5°C to 30°C according to Buck (1989)) and the need for sufficient nutrients within the water body (Reynolds, 1978; Buck, 1989).

According to Moore (1981), *C. hirundinella* is primarily a midsummer species and is scarce or absent in winter, thus meroplanktonic. However, in South Africa the species was also found during winter (Van Ginkel et al., 2001b). The winter water temperatures in many South Africa impoundments seldom go below 10°C (Van Ginkel et al., 2001a), which is within the optimum growth temperature range suggested by Buck (1989). In Spain the species occurs during all four seasons (Péres-Martínez and Sánchez-Castillo, 2001).

C. hirundinella is a dinoflagellate species that has been implicated in filter clogging and taste and odour problems associated with the drinking water industry. It has a financial implication and although it is not a health risk to consumers, numerous complaints showed it to be a problem for the purification industry.

This study was done in order to improve our understanding of the cause of these extreme dinoflagellate blooms that are currently experienced in a number of South African reservoirs and to be able to predict future dinoflagellate blooms.

The hypertrophic reservoirs, Bon Accord, Hartbeespoort, Klipvoor and Rietvlei Reservoirs in South Africa are warm,

* To whom all correspondence should be addressed.
☎ +2712 808-9509; fax: +2712 808-0338;
e-mail: VGinkelC@dwaf.gov.za
Received 26 February 2007; accepted in revised form 26 June 2007.

monomictic impoundments, known to experience dinoflagellate blooms, primarily *Ceratium hirundinella* (Van Ginkel et al., 2001b) and Roodeplaat Reservoir show occasional presence of Dinoflagellates. Only Hartbeespoort, Rietvlei and Roodeplaat Reservoirs have water treatment works that withdraw water for potable use, but all four these reservoirs are used extensively for recreation.

The real time rule-based agents for the dinoflagellate group have been developed and validated by means of limnological time-series data and the hybrid evolutionary algorithm HEA according to Cao et al. (2006). The HEA was designed to assemble and optimise both the structure and parameters of real-time predictive rules using genetic programming and genetic algorithms. 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 Reservoirs 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 to 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. This rule set was then also tested on Bon Accord and Klipvoor Reservoir data that had not been used in the training data set. These reservoirs are of the same trophic status and most are known to experience severe dinoflagellate blooms (Van Ginkel, personal observation).

The agent proved to be generic for five warm, temperate and hypertrophic lakes of which four were monomictic and one dimictic. It can be implemented for forecasting outbreaks of the dinoflagellate (*Ceratium*) blooms, which affect the drinking water supply and water treatment costs of many South African reservoirs.

Materials and methods

Study sites and data collection

The Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs are situated in the Crocodile West/Marico Water Management Area, downstream of the most populated centres of South Africa. These study sites are situated in a summer rainfall area known for afternoon thunderstorms usually causing flash floods with high flows of short duration. The mean annual precipitation, across the catchments, is between 500 mm and 600 mm. The mean air temperatures at the 5 reservoirs varied between –3.5°C and 36.5°C. Water surface temperatures ranged between 10.1°C and 30.7°C in the reservoirs.

A number of limnological variables have been collected over a period of 14 years (1991 to 2004) for Hartbeespoort, Rietvlei and Roodeplaat Reservoirs. Klipvoor has been monitored regularly since 1999. All these reservoirs are monitored regularly as part of the National Eutrophication Monitoring Programme (NEMP), the National Chemical Monitoring Programme (NCMP) and the National Hydrological Monitoring Programme of the Department of Water Affairs and Forestry (DWAF) of South Africa. DWAF provided the chemical and biological data for the reservoirs. The South African Weather Service (SAWS) and the Institute for Soil Climate and Water of the Agricultural Research Council (ARC) provided the meteorological information. 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 information for the Bon Accord and the Roodeplaat Reservoir was extracted from four sites (Pretoria, Roodeplaat (30093); Pretoria National Botanical Inst. (30687); Roodeplaat–AGR (17254) and Irene (0153385)).

Bi-weekly sampling was done. *In situ* temperature profiles, oxygen profiles and Secchi depths were determined at all four reservoirs. Integrated water samples (0 to 5m) were collected close to the dam wall and the following water quality parameters were measured by Resource Quality Services Laboratories at Roodeplaat: pH, dissolved inorganic nitrogen (DIN= ΣNO_3, NO_2 & NH_4), dissolved inorganic phosphorus (DIP as PO_4-P); total phosphorus (TP); total nitrogen (TN as ΣKN, NO_3 & NO_2-N); Chl *a* and dinoflagellate dominance (DinoD) as a percentage of the phytoplankton population. The dinoflagellate biomass (DinoB) was calculated as:

TABLE 1
General characteristics of Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat Reservoirs

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) (10^6 m^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)	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
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

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 33 No. 4 July 2007
ISSN 1816-7950 = Water SA (on-line)

TABLE 2
Limnological properties of the Bon Accord, Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat reservoirs

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)	21.6±47.8	0	1.0	163.1	1.4±0.8	0.02	1.4	3.6
Total nitrogen (mg/ℓ)	23.9±48.0	0.9	3.0	167.5	2.7±0.7	0.02	2.7	5.85
Dissolved inorganic phosphorous (µg/ℓ)	163±155	12	114	1190	54.5±53.2	3.0	35.0	493
Total phosphorous (µg/ℓ)	383±322	96	285	1978	112.7±59.3	26	93	493
Chlorophyll <u>a</u> (µg/ℓ)	304±334	7.0	196.3	2281	46.3±59.6	0.58	26.1	617.4
Dinoflagellate biomass (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)	0.27±0.52	0.04	0.09	4.35	1.1±1.0	0.0	0.9	4.7
Total nitrogen (mg/ℓ)	1.18±0.81	0.33	0.97	4.82	2.5±1.5	0.0	2.2	8.8
Dissolved inorganic phosphorous (µg/ℓ)	482±192	34	490	1451	578±705	6.0	219.6	4084.0
Total phosphorous (µg/ℓ)	699±291	56	699	2447	920±986	0.0	370.1	5582.0
Chlorophyll <u>a</u> (µg/ℓ)	110.6±131.9	2.8	66.0	966.9	54.8±105.9	1.0	26.2	1290.0
Dinoflagellate biomass (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)	0.8±0.7	0.0	0.6	4.5				
Total nitrogen (mg/ℓ)	1.8±0.6	0.6	1.7	4.5				
Dissolved inorganic phosphorous (µg/ℓ)	156.1±120.9	5.0	132.8	620.0				
Total phosphorous (µg/ℓ)	238.3±139.6	12.0	208.0	712.0				
Chlorophyll <u>a</u> (µg/ℓ)	37.3±27.8	1.7	32.9	197.2				
Dinoflagellate biomass (cm^3/m^3)	2.6±16.1	0.0	0.0	143.5				

$$DinoB = (DinoD/100)*(Chla/2.5) \quad (1)$$

where:
DinoB is the dinoflagellate biomass measured as cm^3/m^3
DinoD is the dinoflagellate dominance (measured as %)
Chl *a* is the measured chlorophyll *a* concentration (measured in µg/ℓ)

A simple linear interpolation was used to fill in missing values to produce a complete daily time series for the assessment period.

Rule development by hybrid evolutionary algorithms

The data sets for the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs were merged because of their similarity in hypertrophic and climatic conditions. For this study the data training was done with data from 1991 to 2004, excluding 1993 and 2004 that had been used for testing the rule set. This gave a total of 36 years (12 years from each reservoir) for training and 6 years (2 years from each reservoir) for testing.

The adaptive evolutionary algorithm (EA) methods mimic biological evolution processes, 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 merits of self-organisation, self-learning, intrinsic parallelism and generality, enable EA to be applied to recognise patterns, predict outcomes, optimise control and do parallel processing (Goldberg, 1989; Bäck et al., 1997).

The basic framework of the rule discovery for real-time forecasting of dinoflagellate biomass in hypertrophic South African reservoirs is represented in Fig. 1. The detailed algorithm for the rule discovery and parameter optimisation by HEA is the same as used by Cao et al. (2006) (Fig. 2).

Genetic programming (GP) is used in the HEA to generate and optimise the structure of rule sets. The genetic algorithm (GA) is used to optimise 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) (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

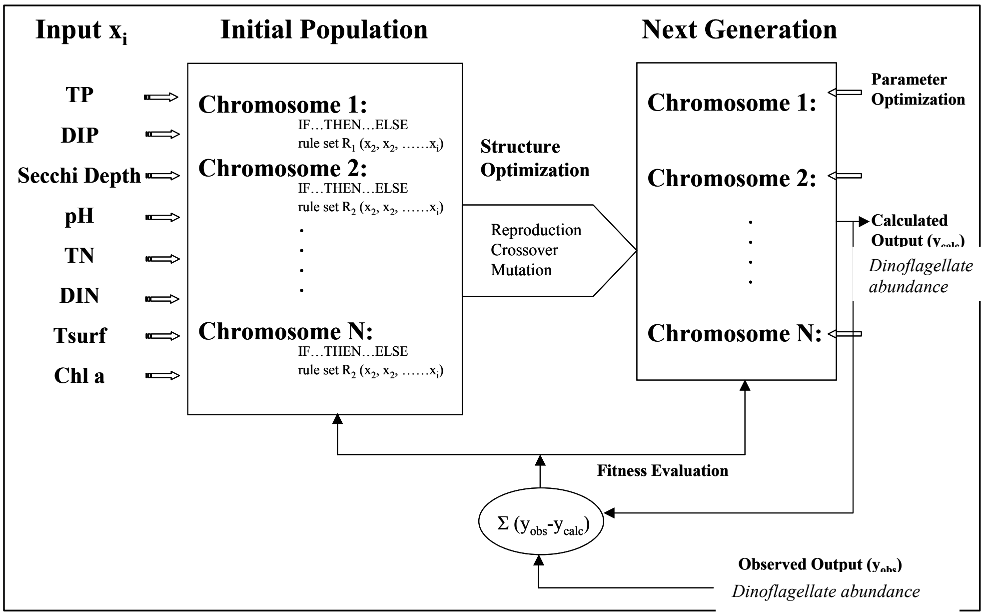

Figure 1
Conceptual diagram of HEA for the discovery of a predictive rule set for Dinoflagellate biomass in three hypertrophic South African reservoirs

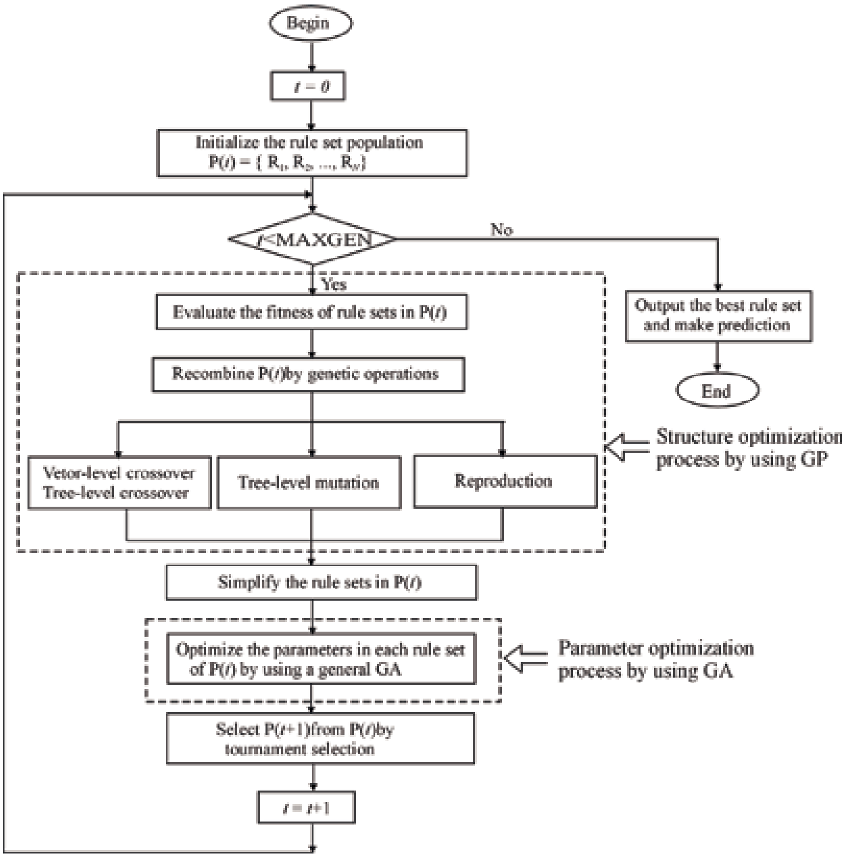

Figure 2
Flow chart showing the hybrid evolutionary algorithm (HEA) (from Cao et al. 2006)

algorithm is then used to optimise the random parameters in the rule set.

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.

Using Hartbeespoort, Rietvlei and Roodeplaat Reservoir data for the experimental programming using HEA, seasonal succession of Dinoflagellates, the daily input data of the following variables: TP, DIP, Secchi depth, pH, TN, DIN, SO_4, Tsurf, Chl *a* and dinoflagellate biomass were used respectively. For the application of the HEA during this study, an initial population of 200, a maximum of 100 generations and 20 runs, were preset for 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 developed by Cao et al. (2006). The parameter settings are listed in Table 3.

To validate the results of the rule set the correlation coefficient of the measured and fitted data were determined and the root mean square error (RMSE) of the training error and the testing error were calculated as follows:

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

where:

m is the number of testing data points

y_i and $\hat{y}_i$ are the ith observation and the ith predicted value of the output variable, namely the dinoflagellate biomass

Testing of the rule set

The developed rule set was tested for 5 reservoirs, namely the 3 reservoirs that it had been trained on (Hartbeespoort, Rietvlei and Roodeplaat) and 2 unseen reservoirs (Bon Accord and Klipvoor). Both the latter reservoirs annually experienced large

TABLE 3 Parameter settings of the hybrid evolutionary algorithm rule set discovery for the Hartbeespoort, Rietvlei and Roodeplaat reservoirs	
Structure Optimisation (GP)	$N = 200$ F_L = {AND, OR} F_C = { >, <, ≥, ≤} F_A = { +, −, ∗, /, sin, cos, exp, ln} MAXK = 1 $D_{IF} = D_{THEN/ELSE} = 4$ MAXGEN = 100
Parameter Optimisation (GA)	popsize = 50 $a = -0.5$ $b = 1.5$ $M = 8$ $MAX = 500$

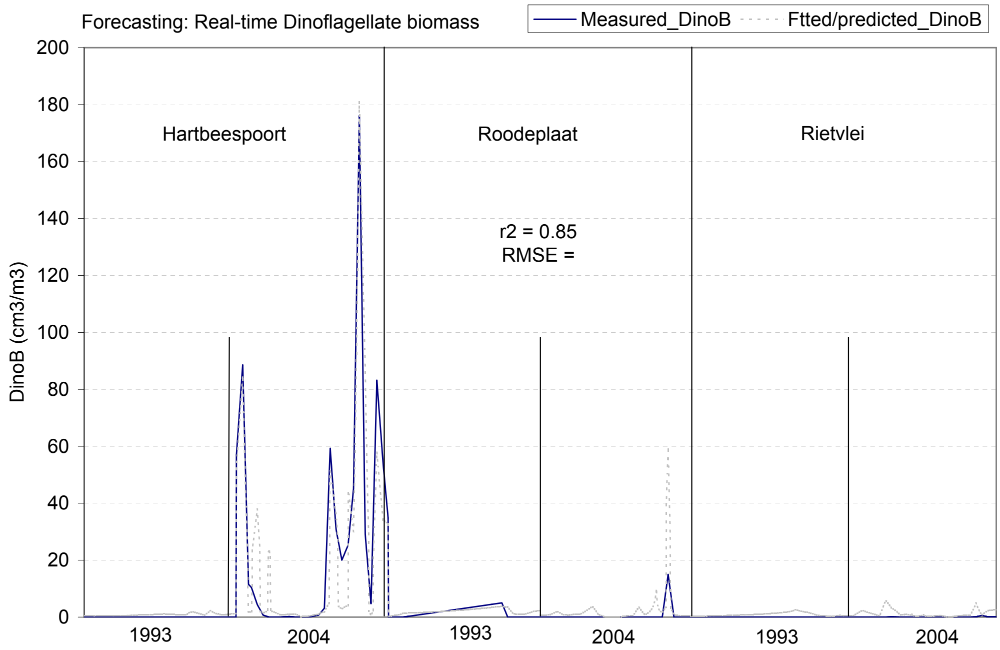

Figure 3
Testing results of the real-time forecasting of Dinoflagellate biomass in the Hartbeespoort, Roodeplaat and Rietvlei reservoirs using the rule set as shown in Figure 4

dinoflagellate blooms during the period 2000 to 2005. The rule set was applied to 1993 and 2004 on the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs. The rule set was applied to the 2000 to 2005 period for the unseen reservoirs, Bon Accord and Klipvoor Reservoirs.

To validate the results of the rule set the correlation coefficient of the measured and fitted data was determined. The root mean square error (RMSE) of the testing error was also calculated for the different reservoirs using Eq. (2) 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.

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 median values were used for the other variables. The starting point of the variable was the minimum and the maximum values measured. This gives an indication of the importance of each variable in driving the changes (increases or decreases) of the outcome, namely dinoflagellate biomass.

Results

All 5 reservoirs are situated in a temperate climate and have been classified as hypertrophic systems with excessive nutrients available. Four of the 5 reservoirs are warm monomictic (Hartbeespoort, Klipvoor, Rietvlei and Roodeplaat), while Bon Accord Reservoir is warm dimictic. Hartbeespoort is the largest of the 5 reservoirs and Bon Accord the smallest. Hartbeespoort and Roodeplaat are deep reservoirs with maximum depths of 32.5 m and 43 m and mean depths of 9.6 m and 10.6 m respectively. The other three reservoirs, Bon Accord, Klipvoor and Rietvlei had maximum depths of 20.9 m, 19 m and 7.4 m respectively and mean depths of below 6.2 m, and can thus be classified as shallow reservoirs. Of the 5 reservoirs the mean annual residence time of only three was available. These 3 reservoirs had a mean annual residence time of less than one year.

Rule development by hybrid evolutionary algorithms

Real-time rule-set discovery for dinoflagellate biomass in Hartbeespoort, Roodeplaat and Rietvlei Reservoirs

The 36 years of real-time training on the Hartbeespoort, Rietvlei and Roodeplaat Reservoirs daily data produced a rule set for the prediction of dinoflagellate biomass as shown in Fig. 4. 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 significant. Figure 3 shows the results of the tested data on the 3 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 dinoflagellate biomass peaks were predicted quite well in all 3 the reservoirs.

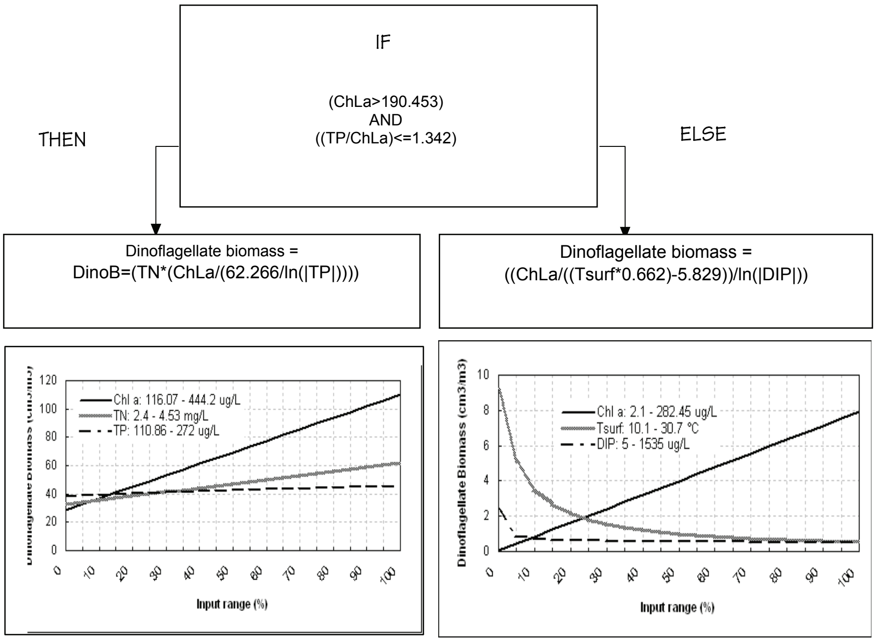

Figure 4
Sensitivity analysis of the input data for the THEN-branch (left) and the ELSE-branch (right) of the rule set for real-time Dinoflagellate biomass forecasting

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 dinoflagellate biomass. The THEN branch is applicable to situations when the Chl *a* concentrations and the dinoflagellate biomass are very high (Fig. 4). The sensitivity analysis that was done on the real-time testing data for the 3 reservoirs to determine the dinoflagellate biomass shows that with the THEN branch of the rule set Chl *a*, TN and TP are the variables that the dinoflagellate biomass is most sensitive to. The TN concentrations vary between 2.1 mg/ℓ and 4.53 mg/ℓ and TP concentrations vary between 110.86 ug/ℓ and 272 ug/ℓ. Under these high nutrient conditions, other variables were not important in the forecasting of the dinoflagellate biomass.

The ELSE branch of the rule set is used when Chl *a* concentrations varied between 2.1 ug/ℓ and 282.45 ug/ℓ and the dinoflagellate biomass was below 10 cm^3/m^3. The temperature change of the surface water is important in determining the dinoflagellate biomass. Increases in the temperature input were important in the decrease of the dinoflagellate biomass (Fig. 4). 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 biomass. Higher dinoflagellate biomass 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 82 ug/ℓ after which the effect of the change in DIP input on the determination of the dinoflagellate biomass was insignificant.

Real-time rule-set for dinoflagellate biomass 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 dinoflagellate biomass in the Bon Accord Reservoir (Fig. 5) 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 was under-predicted or over-predicted in certain instances, but all the peaks were predicted.

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 biomass of the Dinoflagellates in the Bon Accord Reservoir. The maximum peak of dinoflagellate biomass in the spring of 2004 was predicted very well.

In the Klipvoor Reservoir (Fig. 6) the correlation of the measured and predicted dinoflagellate biomass 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 extent of the peaks was 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.

These results indicated 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 be investigated further 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.

Conclusions

The developed hybrid evolutionary algorithm (HEA) for forecasting dinoflagellate biomass proved to be 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 5 reservoirs used for the testing of the rule set are all within the same temperate climatic region of South Africa and all had the presence and dominance of Dinoflagellates during the study period, even though the reservoir sizes and other limnological characteristics differed.

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 33 No. 4 July 2007
ISSN 1816-7950 = Water SA (on-line)

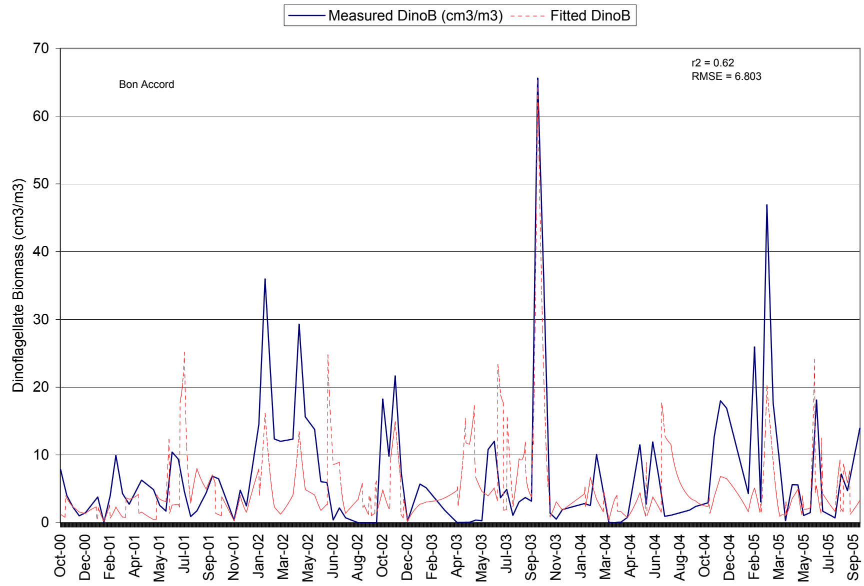

Figure 5
Testing of the developed rule set on Bon Accord Dam for real-time Dinoflagellate biomass forecasting

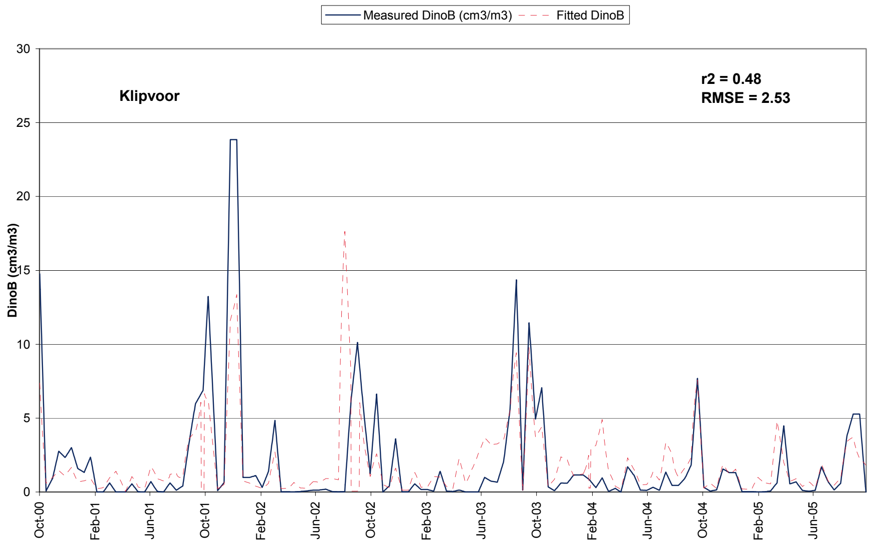

Figure 6
Testing of the developed rule set on Klipvoor Dam for real-time Dinoflagellate biomass forecasting

The developed rule set indicates that a temperature range of 5°C to 17°C and DIP concentrations below 82 ug/ℓ are important during the development of the initial low dinoflagellate biomass. In the development of excessive dinoflagellate blooms TN is the most important variable in determining the extent of the dinoflagellate bloom.

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 biomass. This study is promising for the application of machine learning to complex ecosystems such as man-made reservoirs.

References

BÄCK T, HAMMEL U and SCHWEFEL H-P (1997) Evolutionary computation: comments on the history and current state. *IEEE Trans. Evol. Comput.* 1 (1) 5-16.

BOWDEN GJ, DANDY GC and MAIER HR (2006) *An Evaluation of Methods for the Selection of Inputs for an Artificial Neural Network Based River Model.* Springer-Verlag Berlin, Heidelberg, New York. 275-292.

BUCK H (1989) Ecology of selected planktonic algae causing water blooms. *Acta Hydrobiol.* **31** 207-258.

BUCK H and ZUREK R (1994) Trophic relations between phyto- and zooplankton in a field experiment in the aspect of the formation and decline of water blooms. *Acta Hydrobiol.* **34** 139-155.

CAO H, RECKNAGEL F, WELK A, KIM B and TAKAMURA N (2006) Hybrid evolutionary algorithm for rule set discovery in time-series data to forecast and explain algal population dynamics in two lakes different in morphometry and eutrophication. In: Recknagel F (ed.) *Ecological Informatics* (2nd edn.). Springer-Verlag Berlin, Heidelberg, New York. 347-367.

CRANWELL PA (1976) Decomposition of aquatic biota and sediment formation: organic compounds in detritus resulting from microbial attack on the alga *Ceratium hirundinella*. *Freshwater Biol. (Oxford England)* **6** (1) 41-48.

FREMPONG E (1984) A seasonal sequence of diel distribution patterns for the planktonic Dinoflagellate *Ceratium hirundinella* in a eutrophic lake. *Freshwater Biol. (Oxford England)* **14** (4) 401-421.

GOLDBERG DE (1989) Genetic Algorithms in Search, Optimization and Machine Learning, Reading, Mass.: Addison-Weseley Pub. ISBN 0201157675.

GRIGORSZKY I, PADISÁK J, BORICS G, SCHITCHEN C and BORBÉLY G (2003) Deep chlorophyll maximum by *Ceratium*

hirundinella (O.F. Müller) Bergh in a shallow oxbow in Hungary. *Hydrobiol.* 506-509 209-212.
HEANEY SI, SMYLY WJ and TALLING JF (1986) Interactions of physical, chemical and biological processes in depth and time within a productive English Lake during summer stratification. *Int. Rev. der gesamten Hydrobiol.* **71** (4) 441-494.
HARRIS GP, HEANEY SI and TALLING JF (1979) Physiological and environmental constraints in the ecology of the planktonic Dinoflagellate *Ceratium hirundinella*. *Freshwater Biol. (Oxford England)* **9** (5) 413-428.
HART RC (2006) Phytoplankton dynamics and periodicity in two cascading warm-water reservoirs from 1989 to 1997 – taxonomic and functional (C-S-R) patterns, and determining factors. *Water SA* **32** (1) 81-92. http://www.wrc.org.za/downloads/watersa/2006/Jan%2006/1912.pdf.
HEANEY SI (1976) Temporal and spatial distribution of the Dinoflagellate *Ceratium hirundinella* O.F. Muller within a small productive lake. *Freshwater Biol. (Oxford England)* **6** (6) 531-542.
HEANEY SI and FURNASS TI (1980) Laboratory models of diel vertical migration in the Dinoflagellate *Ceratium hirundinella*. *Freshwater Biol. (Oxford England)* **10** (2) 163-170.
HUTCHINSON (1957) *A Treatise on Limnology* 3. John Wiley, New York.
JAMES WF, TAYLOR WD and BARKO JW (1992) Production and vertical migration of *Ceratium hirundinella* in relation to phosphorus availability in Eau Galle Reservoir, Wisconsin. *Can. J. Fish. Aquat. Sci.* **49** (4) 694-700.
JEONG K-S, RECKNAGEL F and JOO G-J (2006) Prediction and elucidation of population dynamics of the blue-green algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzschii* in the Nakdong River-Reservoir system (South Korea) by a recurrent artificial neural network. In: Recknagel F (ed.) *Ecological Informatics* (2nd edn.) Springer-Verlag Berlin, Heidelberg, New York. 255-273.
MOORE JW (1981) Seasonal abundance of *Ceratium hirundinella* (O.F. Muller) Schrank in lakes of different trophy. *Arch. Hydrobiol. (Stuttgart)* **92** (4) 535-548.
MUELLERELSER M and SMITH WO (1985) Phased cell division and growth rate of planktonic Dinoflagellate, *Ceratium hirundinella*, in relation to environmental variables. *Arch. Hydrobiol. (Stuttgart)* **104** (4) 447-491.
PADISAK J (1985) Population dynamics of the freshwater Dinoflagellate *Ceratium hirundinella* in the largest shallow lake of Central Europe, Lake Balaton, Hungary. *Freshwater Biol. (Oxford England)* **15** (1) 43-52.
PÉREZ-MARTÍNEZ C and SÁNCHEZ-CASTILLO P (2001) Temporal occurrence of *Ceratium hirundinella* in Spanish reservoirs. *Hydrobiol.* **452** 101-107.
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.) *Ecological Informatics* (2nd edn.). Springer-Verlag Berlin, Heidelberg, New York. 325-345.
REYNOLDS CS (1978) The plankton of the northwest Midland meres. Caradoc and Severn Valley. *Field Club* (Occasional paper) **2** 1-36.
RYBAK M (1985) Some ecological effects of artificial circulation on the phytoplankton. *Hydrobiol. (The Hague)* **122** (1) 89-96.
SOMMER U, WEDEMEYER C and LOWSKY B (1984) Comparison of potential growth rates of *Ceratium hirundinella* with observed density changes. *Hydrobiol. (The Hague)* **109** (2) 159-164.
VAN GINKEL CE, HOHLS BC, BELCHER A, VERMAAK E and GERBER A (2001a) Assessment of the Trophic Status Project. Internal Report No. N/0000/00/DEQ/1799. Institute for Water Quality Studies. Department of Water Affairs and Forestry. Pretoria, South Africa.
VAN GINKEL CE, BC HOHLS and VERMAAK E (2001b) A *Ceratium hirundinella* (OF Müller) bloom in the Hartbeespoort Dam, South Africa. *Water SA* **27** (2) 269-276. http://www.wrc.org.za/archives/watersa%20archive/2001/April/1365.pdf (Accessed on?)
WETZEL RG (2001) *Limnology. Lake and River Ecosystems* (3rd edn.). Academic Press, San Diego.