Evolving Difference Equations to Model Freshwater Phytoplankton

P.A. Whigham
Dept. of Information Science
University of Otago, Dunedin
New Zealand
pwhigham@infoscience.otago.ac.nz

F. Recknagel
Dept. of Soil and Water
University of Adelaide
Waite Campus, Glen Osmond
Adelaide, Australia
friedrich.recknagel@adelaide.edu.au

Abstract- The underlying dynamics of algal species in freshwater systems are a complex non-linear problem. Process-based models have been previously developed to describe the time varying behaviour of chlorophyll-a, a measure of algal concentration, for these systems. This paper describes the application of a genetic programming equation discovery system to study various generalisations of a process-based model based on a time series difference equation.

1 Introduction

Developing models for predicting time-varying aspects of freshwater systems is important both for understanding the dynamics of these natural systems and as one component of decision support and management software. The ability to explore the underlying processes driving a natural system help the development of approaches to combat unwanted conditions, such as algal blooms, and to better predict when the system is moving towards conditions that are undesirable. Previous work has produced differential equations for describing dynamic systems [1, 2]. The differential components of the equations were produced by creating additional variables as part of the input based on calculating differences between variables over a time window. The equations were discovered by a systematic and exhaustive search of a limited part of the equation space. The approach described here differs from this work by allowing a larger search space to be investigated using an evolutionary search technique. Additionally, the equations produced allow difference expressions to be created that have not been directly coded as part of the input.

The dynamics of freshwater systems have been studied for many years, leading to the development of many forms of process-based models to describe their behaviour. These models are stated as partial differential equations describing how population dynamics change over time. Solutions to these process-based models are often expressed as difference equations of the form $F(t+1) = G(F(t), v_1, \ldots, v_n)$, where $F(t)$ represents the magnitude of the variable of interest at time t, G is a function developed from the differential equation, and $v_1, \ldots, v_n$ are exogenous and endogenous variables.

This paper describes the application of a genetic programming(GP)[3] system to study various generalisations of a standard process-based model for predicting the time varying composition of chlorophyll-a. The language representing the various models is denoted by a context-free grammar, which imposes certain structure on the possible programs that may be evolved. A set of daily data for a five year period, which describes water quality dynamics at Lake Kasumigaura in the south-eastern part of Japan, is used for this study. This data has been previously modeled using artificial neural networks [4, 5] and a context-free grammar version of GP[6].

The paper is structured as follows: section 2 describes the basics of phytoplankton, the chlorophyll-a data is described in section 3, the process-based model for predicting chlorophyll-a is described in section 3.2, the genetic programming system is described in section 4, the resulting evolved models are described in section 4.4, modifications to the process-based model are described in section 5 and conclusions are drawn in section 6.

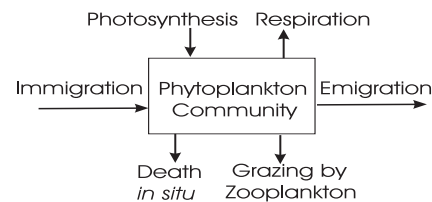## 2 The Ecology of Phytoplankton

Figure 1: A Simple Model of Phytoplankton Dynamics

Phytoplankton are a community of microscopic plants which exist in suspension in aquatic environments [7]. They include representatives of several groups of algae and bacteria and are usually distinguished by being freely floating and dependent on water movement for maintenance and transport. All algae species rely on light for photosynthesis and require nutrients such as nitrogen and phosphorus for growth and reproduction[8]. Other factors which determine their growth include water temperature, thermal stratification of the water column, grazing from zooplankton and competition. The reliance on light for growth produces a seasonal pattern to the phytoplankton community, although this signal may be altered by changes in any of the previously mentioned factors.

Variable	Av $\pm$ Std.Dev	Units
Ortho Phosphate (p)	14.14 ± 25.71	Mg/l
Solar Radiation (l)	1281 ± 671	MJ/m^2
Water Temperature (t)	16.36 ± 7.79	$^\circ$C
Cladocera (cl)	169.87 ± 156.37	inds/l
Copepoda (co)	221.69 ± 83.67	inds/l
Chlorophyll-a (chla)	74.43 ± 42.51	Mg/l

Table 1: Lake Kasumigaura Water Quality Variables

The system displays a highly non-linear behaviour which depends on climatic conditions, nutrient loads and the dynamics of the aquatic food chain. A simple model of the dynamics for the growth of phytoplankton is shown in Fig. 1. Photosynthesis and respiration, which control the growth of the community, are driven by light, temperature and nutrients. The population also increases by immigration (especially in river systems). Decreases in the population occur from death at the end of the life cycle, grazing from zooplankton, such as Copepoda and Cladocera, and emigration from the neighborhood.

The most commonly used biomass estimators for a community of phytoplankton are chlorophyll and carbon [9]. Chlorophyll-a represents the photosynthetic pigment, which is directly correlated with the amount of phytoplankton in a waterbody.

3 Lake Kasumigaura Data

Lake Kasumigaura is situated in the south-eastern part of Japan. It is a large, shallow water body where no thermal stratification occurs. Water temperatures vary widely, ranging from 4° in winter to over 30° in summer. The lake has high nutrient loadings and therefore phytoplankton abundance is high for the majority of the year. Given the reliance on light and temperature for growth there are clear seasonal patterns in the data.

A description of the variables used for this study are shown in Table 1. The data set is composed of five years of daily data from 1989-1993 inclusive.

3.1 Training and Test Data

The data set has been divided into a 2 year training set ($1989 - 1990$) and a 3 year test set ($1991 - 1993$), as shown in Fig. 2. This division was selected based on the fact that the later years exhibit a greater range of chlorophyll-a and are therefore more difficult to model. A previous study using this data [6] developed models by using non-consecutive years for testing. Because this study will evolve difference equations that rely on the previous predicted value to predict the current value it is necessary to use consecutive days when training and testing the evolved models.
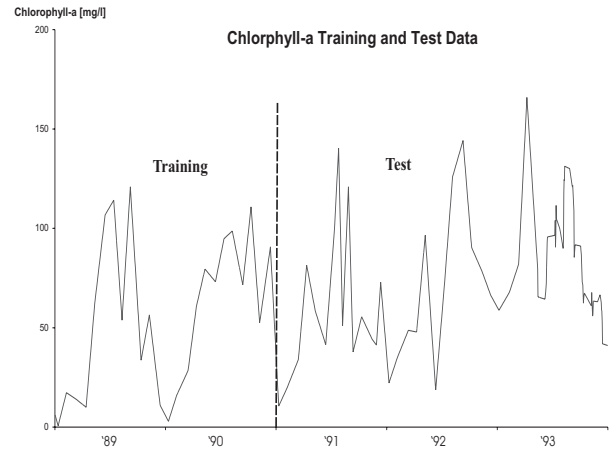

Figure 2: Training and Test Chlorophyll-a Data

3.2 A Process-Based Model for Chlorophyll-a

$$chla_{t+1} = chla_t + chla_t * (Phot - Resp) \qquad (1)$$
$$- chla_t * (co + cl) * 0.00008$$
$$Phot = (0.0667 * t) * \frac{0.0206 * l}{22.4 + 0.0299 * l} *$$
$$\left(\frac{\frac{p}{chla_t}}{\frac{1.36}{X} + \frac{p}{X} + \frac{1.36}{chla_t} + \frac{p}{chla_t}} \right)$$
$$X = 4.613 * chla_t^{0.328}$$
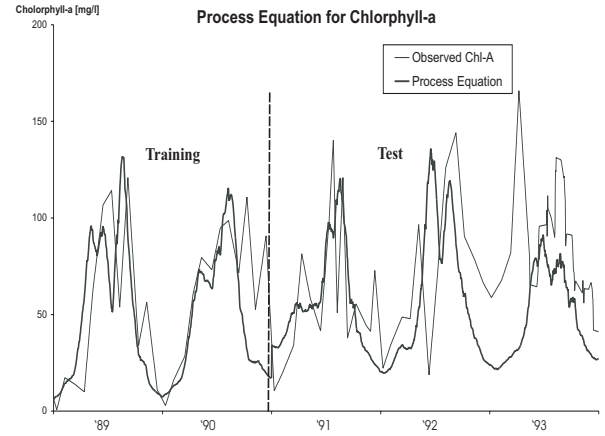$$Resp = (0.0018 * t) + 0.2638 * Phot$$

Figure 3: Chlorophyll-a Prediction using Eqn. 1

Equation (1) was developed based on process understanding of phytoplankton dynamics [10]. This equation includes a growth component based on phytosynthesis and a decay term based on grazing by the zooplankton species Copepoda and Cladocera. Note that immigration and emigration are not included in this equation, however for a lake environment these terms may be assumed to be negligible. The constants of the

original equation were determined from laboratory measurements in an idealised situation. For a particular freshwater system these constants need to be calibrated. For this study the constants of Equation (1) have been calibrated based on the 2 years of training data, using a hill-climbing mutation and constraining the original constant values to within $\pm 20\%$ of their experimentally determined values. The resulting difference equation had a root mean square error (RMSE) of 28.05 for the training data and 37.8 based on the unseen test data for the years 1991-1993. The predicted chlorophyll-a pattern is shown in Figure 3.

4 The Grammatical GP System

The context-free grammar GP system CFG-GP [11, 12] was used to explore possible equations for this study. CFG-GP extends the standard genetic programming approach by using a grammar to define the language of possible sentences that may be expressed by the system. The grammar is used to generate the initial random population of equations and to control the structure of evolved equations in later generations. Crossover and mutation are performed on the derivation trees representing the equations, which ensure that these operators do not produce strings outside the language defined by the grammar. The grammar allows a declarative definition of the language bias for the problem under consideration, and therefore is suitable for studying various structural constraints on the form of evolved expressions.

A context-free grammar was used to define the space of possible expressions that can be evolved to model chlorophyll-a concentration. All grammars were allowed to express equations using the input variables from Table 1. Equation (1) uses the previous predicted value of chlorophyll-a to predict the current value of chlorophyll-a. To allow the comparison between the evolved and process equations the evolved equations could also use the previous predicted value as part of the current time step prediction. For all experiments the population was set to 1000 and allowed to evolve for 50 generations. The fitness measure used was the root mean square error, which the evolving population attempted to minimise over the 2 years of training data. Each setup was run 30 times and the program with the lowest RMSE based on the training data was selected as the best program. Further details regarding the setup for each experiment are given in the relevant section.

4.1 The Most General Grammar $G_{general}$

$$G_{general} =$$
$$\{S,$$
$$\quad N = \{RES, T\},$$
$$\quad \sum = \{+, -, \times, /, pow, p, l, t, cl, co, chla_t, \Re\},$$
$$\quad P =$$
$$\quad\quad \{S \to abs\ RES$$
$$\quad\quad RES \to T$$
$$\quad\quad T \to +\ T\ T\ |-T\ T\ |\times T\ T\ |/T\ T$$

$$\quad\quad T \to pow\ T\ \Re[0.0 - 1.0]$$
$$\quad\quad T \to p\ |\ l\ |\ t\ |\ cl\ |\ co\ |\ chla_t\ |\ \Re[-50.0 - 50.0]$$
$$\quad\quad \}$$
$$\}$$

The grammar $G_{general}$ does not impose any structure to the possible form of the evolved equations, and therefore it is possible for Equation (1) to be expressed as part of the language. Note that for all examples the absolute value of the result is taken to ensure positive values for the prediction. This bias is valid since chlorophyll-a measures are always ≥ 0. The pow function allows arbitrary expressions to be raised to a power: x^y, where y ranges between zero and one. Note that real numbers in the inclusive range [x .. y] are represented in the grammar as $\Re[x - y]$. For example, random real numbers in the range $-50.0 - 50.0$ are defined as part of the basic terminals for the language, derived from the nonterminal T. Crossover and mutation were applied to the nonterminal set $\{T\}$ with a probability of 90% and 5% respectively.

4.2 A Growth-Decay Grammar

$$G_{growth-decay} =$$
$$\{S,$$
$$\quad N = \{RES, G, D, GT, DT\},$$
$$\quad \sum = \{+, -, \times, /, abs, pow, p, l, t, cl, co, chla_t, \Re\},$$
$$\quad P =$$
$$\quad\quad \{S \to abs\ RES$$
$$\quad\quad S \to -\ G\ D$$
$$\quad\quad G \to abs\ GT$$
$$\quad\quad GT \to +\ GT\ GT\ |-GT\ GT$$
$$\quad\quad GT \to \times\ GT\ GT\ |/GT\ GT$$
$$\quad\quad GT \to pow\ GT\ \Re[0.0 - 1.0]$$
$$\quad\quad GT \to p\ |\ l\ |\ t\ |\ chla_t\ |\ \Re[-50.0 - 50.0]$$
$$\quad\quad D \to abs\ DT$$
$$\quad\quad DT \to +\ DT\ DT\ |-DT\ DT$$
$$\quad\quad DT \to \times\ DT\ DT\ |/DT\ DT$$
$$\quad\quad DT \to pow\ DT\ \Re[0.0 - 1.0]$$
$$\quad\quad DT \to co\ |\ cl\ |\ chla_t\ |\ \Re[-50.0 - 50.0]$$
$$\quad\quad \}$$
$$\}$$

The grammar $G_{growth-decay}$ defines equations where the resulting chlorophyll-a prediction is of the form $growth(p, l, t, chla_t) - decay(cl, co, chla_t)$. The growth component is a function of phosphorus, light, temperature and the previous predicted chlorophyll-a value, whereas the decay component is a function of cladocera, copepoda and the previous predicted chlorophyll-a value. This follows the basic structure of Equation (1) however no internal structure is imposed on the form of either of these functions. Note that the absolute value of both the growth and decay terms are used so that the decay term is guaranteed to be subtracted from the growth term. Crossover was applied to the nonterminal set $\{G, D, GT, DT\}$ with a probability of 90% and mutation to the set $\{GT, DT\}$ with a probability of 5%.

Technique	Training (RMSE)	Test (RMSE)
Process Equation	28.05	37.88
$G_{general}$	19.44	32.79
$G_{growth-decay}$	17.86	33.0
G_{diff}	14.13	32.18

Table 2: Resulting Errors for Training and Test Predictions

4.3 Generalised Difference Equation

$$G_{diff} =$$
$$\{S,$$
$$N = \{RES, GD, ABSG, ABSD, G, D, GT, DT\},$$
$$\sum = \{+, -, \times, /, abs, pow, p, l, t, cl, co, chla_t, \Re\},$$
$$P =$$
$$\{S \to abs\ RES$$
$$RES \to +\ chla_t\ GD$$
$$GD \to -\ ABSG\ ABSD$$
$$ABSG \to abs\ G$$
$$ABSD \to abs\ D$$
$$G \to *\ chla_t\ GT$$
$$GT \to +\ GT\ GT\ |\ -\ GT\ GT$$
$$GT \to \times\ GT\ GT\ |\ /\ GT\ GT$$
$$GT \to pow\ GT\ \Re[0.0 - 1.0]$$
$$GT \to p\ |\ l\ |\ t\ |\ chla_t\ |\ \Re[-50.0 - 50.0]$$
$$D \to *\ chla_t\ DT$$
$$DT \to +\ DT\ DT\ |\ -\ DT\ DT$$
$$DT \to \times\ DT\ DT\ |\ /\ DT\ DT$$
$$DT \to pow\ DT\ \Re[0.0 - 1.0]$$
$$DT \to co\ |\ cl\ |\ \Re[-50.0 - 50.0]$$
$$\}$$
$$\}$$

The grammar G_{diff} generalises Equation (1) by constraining equations to the structure:

$$chla_{t+1} = chla_t + chla_t * \{\phi(chla_t, p, l, t) - \varphi(cl, co)\} \quad (2)$$

where ϕ and φ represent functions based on the given arguments. Note that φ is not a function of $chla_t$, to be consistent with Equation (1). Crossover was applied to the nonterminal set $\{G, D, GT, DT\}$ with a probability of 90% and mutation to the set $\{GT, DT\}$ with a probability of 5%.

4.4 Results

The resulting errors for each technique are summarised in Table 2. The three evolved equations are significantly better at the 90% level (based on RMSE) than the original difference equation, however they are all similar in performance to each other. By examining Figs. 4,5 and 6 it is clear that each equation performs worse for the test years. Interestingly, although G_{diff} has the lowest test RMSE, when Figs.4,5 and 6 are compared, G_{diff} does not appear to predict the test years as well as $G_{general}$ or $G_{growth-decay}$. This is partially a product of the fact that the RMSE is not always the best measure

of similarity for a time series, given that similarity is a subjective concept [13]. In fact, this is a good example where our intuitive notion of a good fit to a time series and the measure produced using RMSE do not coincide. Note that all of the equations described in this paper failed to predict the peak value at the end of 1992. Further work is required to determine whether the variables used in this study are adequate for representing this peak in the data. Equation 3 was the best
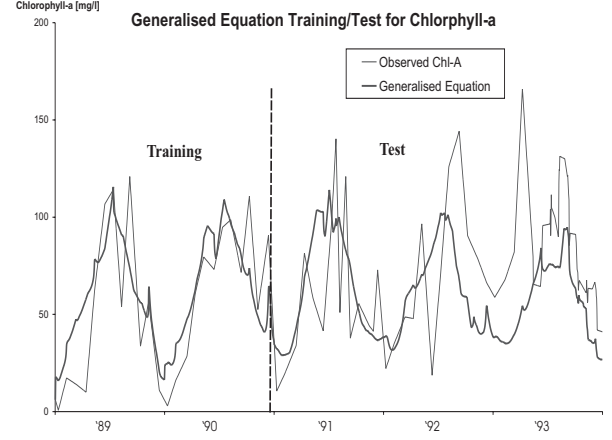

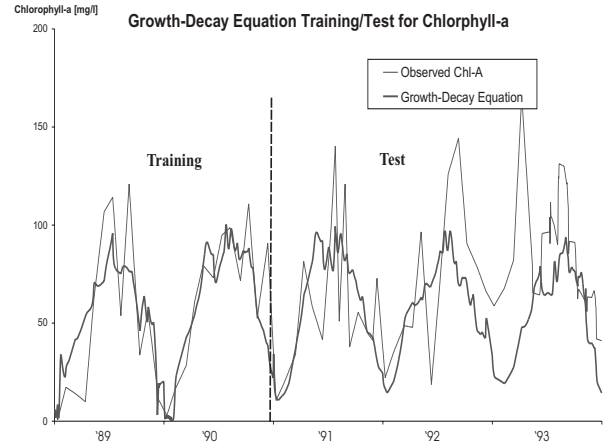Figure 4: Chlorophyll-a Prediction using $G_{general}$

Figure 5: Chlorophyll-a Prediction using $G_{growth-decay}$

evolved solution using $G_{general}$, based on the training data. The equation is a function of water temperature, phosphorus, copepoda and the previous chlorophyll-a prediction value.

$$chla_{t+1} = abs[chla_t * \{\frac{t^{0.32} + 4t}{t^2 p}\} + \frac{co^{0.789}}{t^{0.0081}} + t^{1.358}] \quad (3)$$

The attraction of this solution lies in its simplicity. The surprising result for this solution is that the previous value for chlorophyll-a has been used as one of the variables in the equation. Since it is very easy to construct equations that

create very poor predictions when using $chla_t$, due to positive feedback, most equations which use $chla_t$ are not passed to future generations and become scarce in the population. It would be interesting to study the lineage of this type of program creation to see how these terms survive. It is likely that the mutation operator has introduced the term later in the evolution of the population although this would have to be tested. Equation 4 was the best evolved solution using $G_{growth-decay}$, based on the training data. The equation is a function of all variables except for solar radiation, and is dominated by water temperature.

$$chla_{t+1} = abs[t + 2chla_t^{0.719} - \frac{chla_t^{0.517}}{t^{0.517}} \quad (4)$$
$$+\{\frac{34.92}{(p+17.19)} + \frac{t}{p} + 2t\}^{0.719}]$$
$$-abs[\frac{chla_t + 6cl - 12co - 61.477}{co}]$$

The exclusion of solar radiation may be due to the fact that water temperature is largely determined by radiation and water depth. Equation 4 is more complex than the generalised equation, due to the language bias that was imposed.
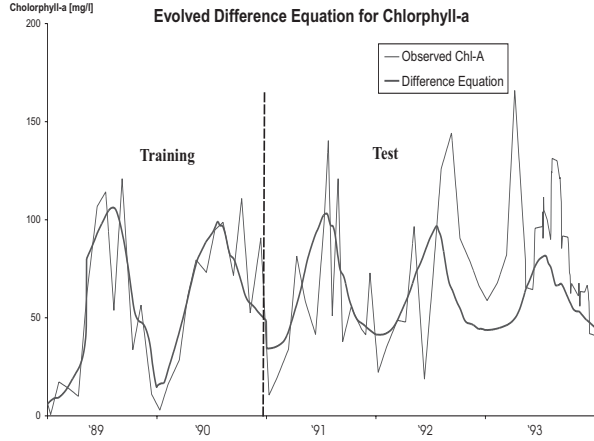

Figure 6: Chlorophyll-a Prediction using G_{diff}

Equation 5 was the best evolved solution from the 30 test runs using G_{diff}, based on the training data.

$$chla_{t+1} = abs[chla_t + chla_t * \{\phi(chla_t, p, l, t) - \varphi(cl, co)\}] \quad (5)$$
$$\phi(chla_t, p, t) = abs[\frac{t^2}{chla_t^2} * \frac{(p - chla_t)}{31.77} *$$
$$\frac{1}{(chla_t - 2t - 5.31)}]$$
$$\varphi(cl, co) = abs[0.0056 * \frac{cl}{co}]$$

Although Equation 5 had the lowest RMSE compared with all other equations, the shape of the predicted chlorophyll-a does not appear to capture the variation in concentration as well as the previous equations.

5 Learning Modifications to the Process-based Model

The previous sections have evolved difference equations to predict chlorophyll-a concentration by following various constraints based on the process-based model described in section 3.2. Although the functional relationships for the process model were followed by various generalisations of the grammar defining the program, the models that have been produced have not helped to produce a better theory of chlorophyll-a modelling. This section will consider other approaches to extending the process-based model by searching for new representations of one component of the process model, a single component at a time. For example, the photosynthesis, respiration or grazing terms could be evolved within the original process-based model framework, while keeping the other factors in the model constant.

To demonstrate this concept, the grazing term from Equation 1 will be evolved within the original model. The purpose is to search for a better representation of the grazing term in relation to the experimentally determined process model. The original grazing term was $-chla_t * (co + cl) * 0.00008$. Two approaches will be described; the first will evolve a grazing term of the form $chla_t * \phi(co, cl)$; the second will allow the $chla_t$ variable to be used more then once in the grazing term, in other words, the grazing term is a function of all three variables: $\phi(chla_t, co, cl)$. Note that both equations do not allow the use of the $-$ operator, to ensure that the grazing term is always positive and therefore, when substracted, always lowers the predicted $chla_{t+1}$ concentration. The following grammars represent the two grazing equations that will be evolved.

$G_{grazing1} =$
$\quad \{S,$
$\quad\quad N = \{GT\},$
$\quad\quad \sum = \{chla_t, +, -, \times, /, sinh, cosh, exp, cl, co, \Re\},$
$\quad\quad P =$
$\quad\quad\quad \{S \rightarrow * chla_t \ GT$
$\quad\quad\quad\quad GT \rightarrow + GT \ GT \ | \ * GT \ GT$
$\quad\quad\quad\quad GT \rightarrow / GT \ GT \ | \ exp \ GT$
$\quad\quad\quad\quad GT \rightarrow sinh \ GT \ | \ cosh \ GT$
$\quad\quad\quad\quad GT \rightarrow \Re[0.0 - 0.0001] \ | \ \Re[0.0 - 10.0]$
$\quad\quad\quad\quad GT \rightarrow co \ | \ cl$
$\quad\quad\quad \}$
$\quad \}$

$G_{grazing2} =$
$\quad \{S,$
$\quad\quad N = \{GT\},$
$\quad\quad \sum = \{chla_t, +, -, \times, /, sinh, cosh, exp, cl, co, \Re\},$
$\quad\quad P =$
$\quad\quad\quad \{S \rightarrow GT$
$\quad\quad\quad\quad GT \rightarrow + GT \ GT \ | \ * GT \ GT$
$\quad\quad\quad\quad GT \rightarrow / GT \ GT \ | \ exp \ GT$
$\quad\quad\quad\quad GT \rightarrow sinh \ GT \ | \ cosh \ GT$
$\quad\quad\quad\quad GT \rightarrow \Re[0.0 - 0.0001] \ | \ \Re[0.0 - 10.0]$
$\quad\quad\quad\quad GT \rightarrow co \ | \ cl \ | \ chla_t$

}
}
Both equations were evolved using a population size of 1000 and evolved for 50 generations. Crossover was set to 90% and mutation 5% over the nonterminal $\{GT\}$. Each setup was run 20 times, and the best result based on the training example selected as the solution.

5.1 Resulting Grazing Equations

The best solution for $G_{grazing1}$ had a training error of 28.5 and a test error of 35.42. Note from Table 2 that these results are comparable with the original difference equation. Since it was not possible to significantly improve on the original equation, when constrained to only use co and cl as variables, it is possible to conclude that this form of equation is an appropriate formalisation of the model. The grazing term produced using $G_{grazing1}$ was:

$$chla_t * 0.000048 * [2co + cl + 0.000069 + \tag{6}$$
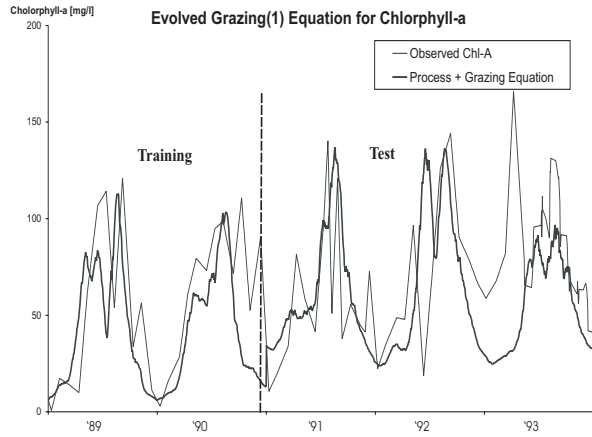$$(0.000069cl + 0.000069co)\cosh(\frac{cl}{co})]$$

Figure 7: Chlorophyll-a Prediction using $G_{grazing1}$

Equation 6 is basically the same form as the original grazing term and can be simplified to $chla_t * 0.000048 * (2co + cl)$, since the $\cosh$ terms are multiplied by a factor of 10^{-10} and therefore can be ignored. The resulting prediction is shown in Figure 7. The main difference compared with the original definition is the increased dominance of the Copepoda term. Note that the constant 0.000069 appears a number of times in Equation 6. This is a result of the crossover operator propagating useful sub-expressions, which contain this constant, through a number of generations to construct the solution. Because the constant was associated with good partial solutions it has spread throughout the population. In terms of producing process equations this is a useful side effect of the technique, since it allows the possibility of general constants in an equation to be discovered.
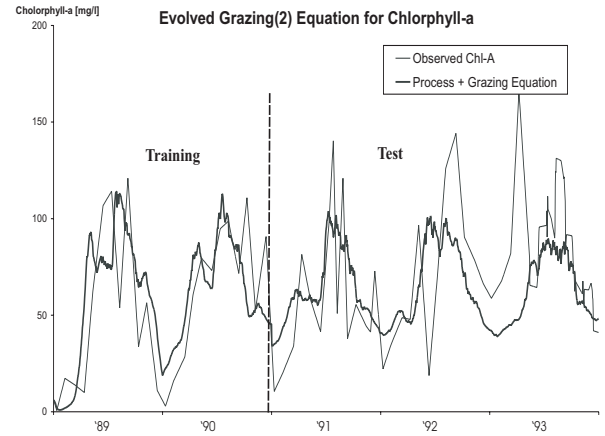

Figure 8: Chlorophyll-a Prediction using $G_{grazing2}$

The best solution for $G_{grazing2}$ had a training error of 22.26 and a test error of 29.63. The resulting predicted curve is shown in Figure 8. These results were a significant improvement over the original process model at the 90% level. The grazing term produced using $G_{grazing2}$ was:

$$\frac{exp(\frac{chla_t}{18.0312})}{co} \times (\frac{\frac{18.0312}{chla_t} + \frac{cl}{co} + cl + co}{\frac{28.8591}{chla_t} + chla_t}) \tag{7}$$

Note that the constant 18.0312 is repeated in this equation, and an interesting question appears as to whether 28.8591 should also be replaced by 18.0312. This would indicate a general constant in the grazing term that could be given some meaning related to how the system is functioning. These considerations, however, are beyond the scope of this preliminary work. The form of Equation 7 is reminiscent of hyperbolic/inverse hyperbolic relationships that are quite common in resource dependent dynamics. This is a promising outcome of the work and shows that it may be possible to use this data-driven approach to reconstruct and extend theories regarding freshwater system dynamics.

6 Discussion and Conclusion

Each of the evolved difference equations have produced significantly better predictors of chlorophyll-a, based on a RMSE measure, compared with the original process-based model. Future work will use this approach to explore and develop new physical interpretations of chlorophyll-a models and to determine whether they represent any underlying process knowledge. The main contribution of this paper is in demonstrating that difference equations may be evolved using the genetic programming framework. The use of a grammar to declaratively represent language bias has also been demonstrated as a useful tool when exploring forms of language and the structure of solutions.

One important conclusion of this work is that for each type

of equation the previous predicted value of chlorophyll-a was found to be useful in predicting the current value. This is particularly interesting in relation to the equation created using $G_{general}$, given that this language did not require $chla_t$ to be included as part of the solution. This form of equation follows the standard approach to developing process-based models and therefore supports the general framework currently used by ecologists when describing population changes. The use of previous values in time series to predict the current value for a model also suggests that this approach can be extended to explore the time dependent nature of a system. By allowing all independent variables to express previous as well as current values the time lags for a system should be able to be explored.

The preliminary work described in Section 5 shows some promise. Equation 6 supports the current linear representation of grazing based on the original process-based equation. The more general expression of Equation 7, with the subsequent improvement in prediction, suggests that the grazing term is not a linear function of the current chlorophyll-a concentration. The similarities with resource dependent dynamics indicates that the work has the possibility of extending current process understanding, which will agree with the theory of general non-linear systems. However, further work is required before any conclusive results can be stated.

Reducing the overall search space is a desirable goal for any problem domain when applying an evolutionary learning system. Research is currently underway to consider how knowledge of the units (see Table 3) required by an equation can be used to reduce the possible legal combinations of variables, and therefore reduce the search space of the problem. For example, the form of Equation 7 should produce a resulting value in terms of Mg/l, however the resulting grazing term has units $\frac{Mg.l}{l^2 + Mg^2}$. Logically this is not a valid expression, even though the resulting grazing equation does produce significant results. Limiting equations that satisfy the constraint imposed by the units definition is a requirement that will be explored in the near future.

The use of RMSE for measuring the fitness of individuals has been mentioned in Section 4.4. Since this measure does not directly incorporate any shape information, in the form of slope or shape measures between data points, it is possible for an equation with a low RMSE to be selected which does not satisfy our subjective concept of similarity. Work is currently underway to produce a set of classifications to assist in the selection of appropriate fitness measures when using time series data.

Bibliography

[1] S. Dzeroski and L. Todorovski, "Discovering dynamics," in *Proceedings of the 10th International Conference on Machine Learning*, pp. 97–103, Morgan Kaufmann, San Mateo, CA, 1993.

[2] L. Todorovski and S. Dzeroski, "Declarative bias in equation discovery," in *Proceedings of the 14th International Conference on Machine Learning*, pp. 376–384, Morgan Kaufmann, San Mateo, CA, 1997.

[3] J. Koza, *Genetic Programming:on the programming of computers by means of natural selection*. A Bradford Book, The MIT Press, 1992.

[4] F. Recknagel, "Anna - artificial neural network model for predicting species abundance and succession of blue-green algae," *Hydrobiologia*, vol. 394, pp. 47–57, 1997.

[5] F. Recknagel, M. French, P. Harkonen, and K. Yabunaka, "Aritificial neural network approach for modelling and prediction of algal blooms," *Ecological Modelling*, vol. 96, pp. 11–28, 1997.

[6] P. Whigham and F. Recknagel, "Predictive modelling of plankton dynamics in freshwater lakes using genetic programming," in *Proceedings of the International Congress on Modelling and Simulation, MODSIM'99* (L. Oxley, F. Scrimgeour, and A. Jakeman, eds.), vol. 3, pp. 679–684, Modelling and Simulation Society of Australia and New Zealand Inc., 1999.

[7] C. Reynolds, *The Ecology of Freshwater Plankton*. Cambridge:Cambridge University Press, 1984.

[8] C. Reynolds, "Phytoplankton periodicity: the interactions of form, function and environmental variability," *Freshwater Biology*, vol. 14, pp. 111–142, 1984.

[9] G. Harris, *Phytoplankton Ecology:Structure, Function and Fluctuation*, ch. 6. Chapman and Hall Ltd., 1986.

[10] F. Recknagel and J. Benndorf, "Validation of the ecological simulation model salmo," *Int. Revue ges..Hydrobiol.*, vol. 394, pp. 47–57, 1982.

[11] P. Whigham, "Inductive bias and genetic programming," in *First International Conference on Genetic Algorithms in Engineering Systems:Innovations and Applications*, pp. 461–466, UK:IEE, Sept. 1995.

[12] P. Whigham and P. Crapper, "Time series modelling using genetic programming: An application to rainfall-runoff models," in *Advances in Genetic Programming 3* (L. Spector, W. Langdon, U. O'Reilly, and P. Angeline, eds.), ch. 5, pp. 89–104, USA:MIT Press, 1999.

[13] E. Keogh and M. Pazzani, "Relevance feedback retrieval of time series data," in *The Twenty-Second Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, 1999.