Ecological Application of Evolutionary Computation: Improving Water Quality Forecasts for the Nakdong River, Korea

Dong-Kyun Kim, Bob Mckay, *Senior Member, IEEE,* Haisoo Shin,
Yun-Geun Lee and Xuan Hoai Nguyen, *Member, IEEE*

***Abstract*— Water quality is an important global issue, requiring effective management, which needs good predictive tools. While good methods for lake water quality prediction have previously been developed, accurate prediction of river water quality has hitherto been difficult. This project combines process-model and data mining approaches through evolutionary methods, resulting in tools for more effective water management. Although the work is still in its preliminary stages, error rates of the predictive models are already around half those resulting from representative applications of either pure process-based or pure data mining approaches.**

I. Introduction

The Nakdong River system in South Korea is one of the major river systems of North East Asia. It drains an area home to around ten million, including the million-strong city of Daegu, and at its mouth forms the primary water supply for Busan, a city of five million. Such intensive use inevitably leads to conflicting requirements, a key issue being the problem of algal blooms, fueled by the nutrients injected upstream, which periodically blight the river in the vicinity of Busan. So important is the management of this river that the Korean government is preparing to invest in the vicinity of $US30 billion in a scheme to improve its water management.

But management requires information. It is impossible to manage a river effectively without an understanding of the effects of management decisions. In the case of algal blooms, this requires a model of the effects of both management changes – e.g. decisions on water release from dams, or controls on nutrient release – and of exogenous changes – e.g. changes in rainfall levels and timing as a result of climate change. Currently available models are unsatisfactory.

This project is using evolutionary methods to generate better models of the algal dynamics of the catchment. The work is currently in progress, but has already halved the predictive error of previously-published models. We are reporting on it now, to document what has been achieved, and also to illustrate how the flexibility of evolutionary methods supports an interplay between algorithms and expert knowledge that would be difficult to duplicate with competing methods.

Algal blooms are not unique to Korea. Rivers around the world are subject to increasing development, and algal blooms have become a major concern in many countries. The models we are developing for the Nakdong are designed around specific knowledge of the ecological setting of that river, and the data available for it, but the overall techniques are general, and can be adapted to river systems worldwide.

In the remainder of this paper, we will first (section II) describe what has previously been done in modelling algal blooms in lakes and rivers, and discuss some of the available techniques. Section III follows with a more detailed description of the modelling problem and of the data sources available, with section IV outlining the methods we are using and how we plan to extend them. Section V presents the results of the modelling so far, comparing it with previous work. We discuss our future plans for the project in section VI, concluding in section VII with a summary both of the work itself, and of the role of evolutionary computation.

II. Background

A. Ecological Modelling

An ecosystem can be compared to a living system in complexity. It is subject to internal feedback mechanisms, often unknown and sometimes chaotic; it changes over time, with the influence of different components waxing and waning. A large open system like a river not only contains numerous internal components, but is also affected by unpredictable external forces, both natural (e.g. weather variations) and anthropogenic (land use changes, dams and barrages etc.). The complexity is far beyond what we can hope to model, so models can only approximate the most important influences.

On the other hand, data for building good models is often scarce. Some kinds of ecological data can be cheaply gathered, e.g. from satellites or automated stations. But much essential data can only be gathered by painstaking hand collection, so that ecological datasets are often of very small size, perhaps a few hundred instances. Moreover the data is subject to noise, arising from measurement error, external influences, and often the difficulty of reproducing the measurement circumstances accurately.

One common approach to ecological modelling builds the models by hand, relying on expert knowledge and experiments to define both the structure of the model and its

Dong-Kyun Kim, Bob Mckay, Haisoo Shin, and Yun-Geun Lee are with the Department of Computer Science and Engineering, Seoul National University, Seoul, Korea (phone: +82 2 8801481; emails: [dkkim1004,rimsnucse,dasony,ey9ey9]@gmail.com).

Xuan Hoai Nguyen is with the Faculties of IT at Le Quy Don and Hanoi Universities, Hanoi, Vietnam (phone: +84 982626900; email: nxhoai@gmail.com).

parameters. Unfortunately such approaches often reveal difficulties in practice. The expert's knowledge may be wrong, experimentally-derived parameter values may be subject to unknown variation "in the wild", and so on.

An alternative approach uses machine learning methods to build models from data. But in many circumstances – such as our water modelling domain – this also suffers from problems. The available data is simply not sufficient to justify a model of the complexity required for accurate prediction.

1) Model Revision: This dilemma, that process models based on prior knowledge may be too inaccurate, yet there may be insufficient data for machine learning, can be resolved in another way: through model revision. In this method, the problem is not to generate a model de novo, but to learn revisions to an existing model. Thus the data may be used to build a model more complex than could be justified by it alone, yet more accurate than one based on an a-priori model. In a sense, this is a generalisation of a widely used approach, using learning methods to fit parameters in a pre-built model. it has had limited application in ecological modelling; Todorovski and Dzeroski [1], [2] implemented equation revision using the equation discovery system LAGRAMGE, and illustrated its effectiveness by improving a pre-exisisting model of the net production of carbon in the Earth ecosystem. However LAGRAMGE uses complete search of the space of possible modifications; evolutionary methods may be better suited to this process.

B. Water Quality Modelling

1) Modelling Water Quality in Lakes: Lakes are the simplest environments for modelling water quality processes. While many water resource agencies still rely on manually-constructed process models, machine learning methods have demonstrated somewhat enhanced performance. Whigham and Recknagel [3] used Context Free Grammar Genetic Programming (CFG-GP) to adapt the lake model SALMO, while Cao et al. [4] optimised the parameters of the closely-related SALMO-OO using an evolutionary algorithm. Liu and Yao [5] directly evolved neural network models predicting algal growth, and Welk et al. [6] forecast algal population dynamics in freshwater lakes using an evolutionary algorithm. Recknagel et al. [7] compared time-series algal predictive models developed by artificial neural networks and genetic algorithm in freshwater lakes.

2) Modelling Water Quality in Rivers: A river ecosystem introduces much greater complexity for modelling water quality. Most notable is the importance of flow. While water quality may be modelled locally in a lake, in a river the determinants of water quality are heavily affected by flow. Taking our example of algal growth, both the algae themselves, and the nutrients on which they depend, are carried with the flow. But we can't concentrate purely on the flow either, because the algae grow as they are transported. In effect, we have two extremes: in the zero-flow case (i.e. a lake) we may pay attention simply to an algal growth model; at the opposite extreme, in a high flow river, we may ignore growth (in most such cases, there is no algal problem, so no need for a model). In the cases of interest, such as the Nakdong River, the situation lies between these extremes.

Data collection is also more difficult. Monitoring and sample measurement are needed in widely separated locations (by comparison with typical lakes), so that good-quality data is less available. Determining the flow may itself be problematic. Fitting the mass balance may face unpredictable water loss and discharge (e.g. illegal intake and dumping, ground water loss, evaporation, etc). As a result, river models are much less developed at this stage than lake models.

The most general and best-known process model in river ecosystems is QUAL2E [8], which has been widely used to simulate dissolved oxygen under steady flow conditions. However, application has revealed many inaccurate simulations in different river ecosystems. Extensions have been made to handle specific conditions, but were mostly designed to deal with specific physicochemical issues. When the complexities arose from biological interactions, in many cases it was not accurate enough to use. Conversely, such models could be extended with more reasonable simulation results by considering additional biological components [9], [10].

At the opposite extreme, it is possible to take a pure data mining approach similar to those successfully applied in lake modelling, in the hope that variables available to the model may supply surrogates for the unavailable information about flow and upstream conditions. This was the approach taken by Kim et al. [11], [12] and Cao et al. [13] in previous attempts to model algal growth in the Lower Nakdong River, by learning respectively nonlinear mathematical models and decision rules predicting the chlorophyll a level.

3) Modelling Water Flow: River flow modelling is important because timely information may allow better management, whether it be for better control of water flows in a regulated rivers, or more timely evacuation in the event of flooding. There have been many applications of learning methods to these problems, emphasising time-series prediction in rainfall-runoff modelling [14], [15], and station-to-station velocities and level relationships in relation to flooding [16]. Solomatine [11] presents a range of other applications of learning methods to flow-related prediction, mainly in flood control.

III. The Modelling Problem

A. Description of the Study Sites

The Nakdong is the longest river in South Korea (ca. 525 km), with approximately ten million people living in and using water from the basin. The annual rainfall is nearly 1200 mm per year, over 60% concentrated in the Summer monsoon (June to September) [17]. Four large multi-purpose dams are sited in the head streams to control flow, which is highly regulated. Near the mouth, an estuarine barrage protects the fresh water from salt intrusion, thus increasing the residence time of the water body, creating what, in some flow states can be lake-like conditions [18]. The large population of Busan metropolitan city draws a huge intake from the river. The

combination of these circumstances has led to a deterioration of water quality in the lower Nakdong River, resulting in a proliferation of recurrent algal blooms.

We have data from nine measuring stations throughout the catchment (see Figure 1). They were originally selected based on the availability of data and geographical importance. Six stations (S1 to S6 from lower to upper) are located in the main channel of the river, while the other three stations (T1 to T3, see Figure 1) are situated in major tributaries. Of those stations, algal concentration in the lowest (Mulgeum, S1) is most important because the high population (ca. five million) of Busan draws its water nearby. To predict the algal biomass at this station is the underlying task of our model.

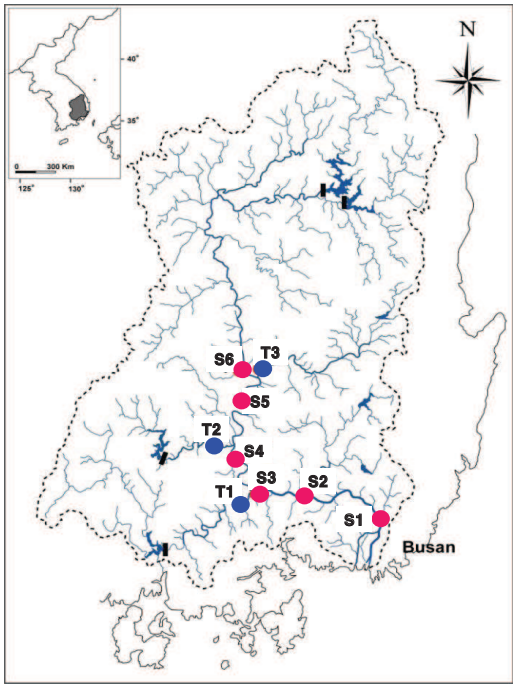

Fig. 1. Nakdong River Basin

B. Data and Sources

For construction of the model, we have a total of twelve different types of variables and parameters (Table I). Most hydrological and meteorological data are derived from the website of the Korean Water Management Information System (WAMIS [19]). Water levels are continuously logged in each station. Flow rates are derived through a regression formula based on the river height provided by WAMIS. The formula generally has the form given in equation 1:

$$\text{Flow} = \alpha * (\text{Height} + \beta)^{\gamma} \quad (1)$$

where α, β and γ are site-specific (and time-specific) parameters that depend on the riverbed contour.

The flow data are particularly problematic, because the three parameters α, β and γ change wildly when the shape of the river bed changes; this change is generally slow, but punctuated at irregular intervals by the silt carried in the extreme flows from typhoons (for which we know the dates), and more rarely, by dredging (for which we do not). The fitting parameters are tuned by WAMIS at irregular intervals, not coincident with events affecting the river contours, by measuring actual flows over a period of a few months, then calibrating the regression constants. In addition, WAMIS' main purpose is biased toward flood prevention (i.e. high flow) rather than water quality (low flow); thus flows are most often measured during high flow, and the models fitted to minimise error at these times. Disparity between these data, especially in low flow seasons, might have a large impact on algal concentration in the water model.

Irradiance and rainfall data were provided courtesy of the Korea Meteorological Administration (also available from the KMA website [20]). Catchment areas were derived from electronic map data. Water temperature and dissolved oxygen were manually measured at each station. Nutrient concentrations (nitrate, phosphate and silica) and chlorophyll a concentration were analysed and measures in the laboratory from field-collected water samples. Generally, algal biomass was directly related to chlorophyll a concentration, hence it was predicted using chlorophyll a in this paper.

TABLE I
VARIABLES

	Variable	Unit
Geographical	catchment area	km^2
Hydrological	water level	M
	flow velocity	m/s
	flow rate	m^3/s
Meteorological	irradiance	MJ/m^2
	rainfall	mm/day
Physicochemical	water temperature	°C
	dissolved oxygen O_2	mg/l
	nitrate NO_3	mg/l
	phosphate PO_4	mg/l
	silica SiO_2	mg/l
Biological	chlorophyll a	μg/l

In building a time-based model from data, it is important to take data intervals into account. Data from 1996 to 2008 (13 years) were used. The hydrological and meteorological data were collected daily, while other (field measured) data were sampled weekly (at Mulgeum over most of the period) or biweekly (elsewhere). To provide consistency, longer-interval data were linearly interpolated to a daily scale.

IV. Models and Methods

A. The Process Model

The overall process model we use as a starting point consists of two parts: a river flow model (hydrological processes), and an algal growth model (biological processes).

1) The River Flow Model: uses a simple flow mass balance between stations. This is used to estimate the flow time between stations, and thus to provide time information to the biological process (algal growth) model. Properties of

the water flowing through each station are predicted from earlier data for its upstream stations. Hydrological properties like flow rate are estimated based on hydrological processes. Assuming water flows from station A to station B, flow at station B can be estimated from equation 2.

$$F_{B,t+d} = (1 - r_A)F_{A,t} + r_B F_{B,t} + R \tag{2}$$

$F_{X,t}$ denotes the flow at station X at time t, while d is the time it takes for water from A to reach B. It can be calculated from the flow velocity, which in turn depends on the flow rate. River flow is not laminar; we have to take into account the lower rate of flow at the edges, water that may be trapped in side pools etc. We do this by considering the water to be divided into two portions: one portion subject to laminar flow, and the other which is retained in the river reach till the next time period. r_A is the retention ratio of station A, meaning the proportion of water that does not flow out of the station in a given time period. This value also depends on the flow velocity. R is the amount of water added by rainfall – of course, it depends on the rainfall and the catchment area of each station. To summarise, the flow rate at a certain time is sum of flow rates of water from the upper station, from previously retained water, and from runoff from rainfall.

2) The Algal Growth Model: models the simultaneous process of algal growth in the flowing water. The algal biomass changes according to the algal growth model, which can be approximated by equations 3, 4 and 5.

$$\frac{\mathrm{d}\,Chl}{\mathrm{d}\,t} = Chl \cdot (\mu - \gamma - \delta) \tag{3}$$

$$\mu = f_1(L, T, N) \tag{4}$$

$$\gamma = f_2(T) \tag{5}$$

Equation 3 relates the rate of change of algal biomass (*Chl*) to the growth rate (μ), respiration rate (γ) and mortality rate (δ). The growth rate is determined by the average light level *L*, temperature *T* and nutrient concentrations *N*. The respiration rate depends on water temperature, and mortality is assumed constant. In lake models, there is also a sink, loss due to settling; flow turbulence in rivers renders this process unimportant. To simplify the model at this initial stage, we also omitted any mechanism for predation.

3) Integrating the two Models: The simulation must take into account several water flows merging into one. The water arriving at station S5 is the result of merging water flows from S6 and T3. To simplify the computational model, virtual stations have been added at each junction. Thus instead of water flowing from S6 and T3 directly to S5, it first flows to a virtual station J3, where the two flows are merged, then the merged flow propagates to station S5.

Merging water flows is straightforward. Some properties, such as flow, can be simply added together, while others, such as water temperature or algal mass per liter, are calculated as averages weighted by flow rate.

4) Validating the Flow Data: The flow model has another role: providing a cross-check on the flow data we obtained from WAMIS model through mass balance: since the flow into a reach should roughly equal the flow out. There may be additional effects, especially due to the abstraction and return of water for the city of Daegu; by our calculations these effects should not affect the lowest flows by more than 10%. Using the original flow formulae supplied by WAMIS, the mass balance over the whole period, and over all reaches, was out by an average of 73.1%! There is no possibility that such a huge volume of water could be lost from a river like the Nakdong. The flow figures must be wrong by at least this amount. Working backward with the WAMIS data and formulae, using the known occurrences of typhoons, our domain expert constructed regression formulae reducing this imbalance to 55.3% – still unacceptable, but sufficient that we should see some improvement in our algal growth predictions if the flow errors substantially affect the outcome.

5) The Overall Model: To summarise, all measured data from the four highest stations (one main channel: S6 and three tributaries: T1, T2 and T3) were used as sources to calculate values in downstream stations. In estimating flow rates, they were recalculated at the confluence where the tributaries joined the main channel, then propagated to the next reach. We paid particular attention to the retention ratio for the flow in each reach, calibrating it to changes in flow velocity, because water retention – and the consequent longer residence times – may play a key role in algal growth, especially in highly regulated rivers like the Nakdong [18], [21]. We anticipated that increases in retention time might accelerate algal blooms during the peak (dry, Winter) periods.

Although the primary process functions can be expressed simply as equations 2 and 3, the secondary processes incorporate a variety of combinations of variables and constant parameters. These parameters were derived from both river [9], [22] and lake [23], [24] process models. However the parameters of the model can be fitted to the system's environmental characteristics. Thus we applied a genetic algorithm to the Nakdong River model to find a well-fitted process revised by parameter optimisation.

Testing the model requires fixing a simulation interval.We used a constant interval of 36 hours. Properties of the water at each station were estimated every 36 hours over the 13 year time span. This is a necessary trade-off. Assuming the model is reasonably good, a shorter time interval may give a more accurate result, but the simulation will also be slower. When it is incorporated into the evaluation loop, this is a critical issue. So far, 36 hours seems to give reasonable results.

B. Parameter Optimisation by Genetic Algorithm

We optimised the model parameters using a canonical genetic algorithm (GA). The gene structure is an 18-dimension real vector, representing the 18 model parameters. Whenever the fitness is required, parameters are substituted into the river model and the model is run; the fitness is the overall error of the model over the period. The genetic operators

were tournament selection (size 4), uniform crossover and gaussian mutation. Preliminary testing was used to find suitable parameter settings. In particular, we found the best results with a high mutation rate. We used an elite of one. Table II shows the evolutionary parameters in detail.

TABLE II
EVOLUTIONARY PARAMETERS FOR GENETIC ALGORITHM

GA Type	canonical genetic algorithm
Max Generation G_{max}	500
Population Size	100
Elite Size	1
Selection	Tournament, size 4
Xover	Uniform Crossover (p_c = 0.6)
Mutation	Gaussian Mutation (p_m = 1.0)

V. MODEL RESULTS

A. Predefined vs. evolved process model

Our first experiment compared the results we could expect from a traditional process-based approach, with parameters based on expert opinion combined with lab experiments, with what could be achieved by GA parameter optimisation. The results are shown in figures 2 and 3.

Before we discuss them in detail, we note some important issues regarding Chlorophyll a concentration. The peaks are of most interest, since low algal levels pose no problem. There is clear evidence of seasonality and recurrent variation in time scale, but magnitude and timing of onset were irregular (Figure 2). As a long term trend, algal blooms were more severe during the 1990s, with peaks moderating since 2002. Algal blooms occurred most commonly in the Winter; but on the few occasions when they did occur in Summer, peak concentrations were much higher.

Figure 2 shows chlorophyll a predictions based on expert/lab parameters. Prediction error was 38.76 (RMSE) from 1996 to 2008. Overall, predicted values underestimated blooms, and frequent fluctuations inconsistent with the data were observed. Figure 3 depicts the process model after tuning by GA parameter optimisation. This model gives a prediction error roughly half (RMSE, 21.34) that of the predefined model. More important in practice, though difficult to quantify, it tracked the real peaks more accurately, even when it did not precisely predict their scale. Parameter fitting by GA, in this instance, gave much better model performance than expert opinion and lab-determined parameters.

B. Model Performance and Flow Data

One potential source of error in the overall model is error in the flow data. As we discussed in section III, our cross-checking revealed serious inconsistency in the original WAMIS flow data. Since the peaks in algal growth are highly correlated with the troughs in flow, errors in these flow rates could potentially seriously affect the model predictions.

To test the importance of flow error, we assessed the effect of different flow data on the model. We fed both the original WAMIS data, and our expert-corrected data, into the models obtained from the GAs. We are reasonably confident that the expert corrections resulted in more accurate flow data, because of the reduction in mass balance inconsistencies. The resulting prediction errors were 21.38 vs 21.34. Figure 4 shows the resulting time-series algal dynamics, while figure 5 shows the time-series of relative error for flow rate (in terms of mass balance) and for chlorophyll a concentration (in this context, the relative error is more relevant than the absolute error). We can see that there is little correlation: the times when the flow predictions are poor are essentially unrelated to the times of poor chlorophyll a prediction. From all these, it appears that flow errors caused by river bed alteration do not have a large effect on the errors of the algal growth model (of course, this could change if subsequent work eliminates other sources of error).

C. Cross-Validation for generality of the model

We used eight sets of training/test data (five years training, with the subsequent year for testing) to estimate model predictability. Table III shows the resulting accuracy on training and test data. Despite large year-to-year variations (some years are just more predictable than others), we see generally reasonable results; it is particularly notable that the first few years in our series – with large Summer blooms – were particularly difficult for our models to fit. Overall, there is no apparent evidence of overfitting. It is worth noting, though, that lower RMSE was generally observed on test than on training data. This may arise partly from the effect already noted: that the earlier data may have been inherently less predictable than the later. Unfortunately the usual methods for handling such problems – 'leave one out' cross validation or bootstrapping – are not available to us because the data must be input into the model process in consecutive order.

TABLE III
CROSS-VALIDATION

	Year	**RMSE**	r^2	**AME**
Train	1996 - 2001	28.3	0.47	14.8
Test	2002	17.2	0.70	12.5
Train	1997 - 2002	25.5	0.54	12.7
Test	2003	21.4	0.56	11.9
Train	1998 - 2003	18.0	0.66	12.0
Test	2004	13.5	0.90	8.5
Train	1999 - 2004	16.0	0.78	10.7
Test	2005	17.4	0.75	11.3
Train	2000 - 2005	10.5	0.84	6.6
Test	2006	15.9	0.79	10.1
Train	2001 - 2006	13.4	0.89	7.3
Test	2007	13.4	0.89	7.3
Train	2002 - 2007	14.9	0.82	9.3
Test	2008	5.3	0.73	3.9

D. Performance Relative to Previous Nakdong River Modelling

Our study is not the first to attempt prediction of algal growth in the lower Nakdong. Two previous studies used Recurrent Artificial Neural Networks [25] and Genetic

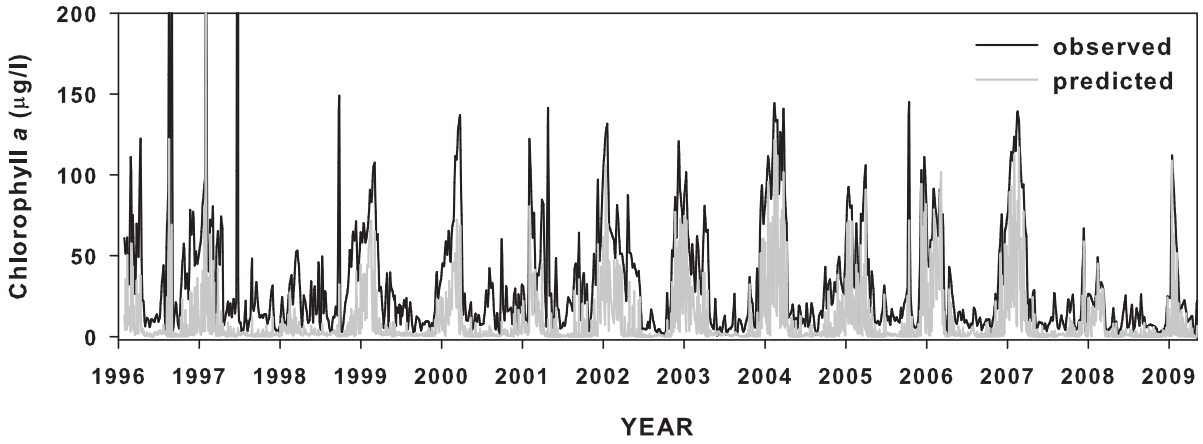

Fig. 2. Chlorophyll *a*, Actual vs Predicted (Traditional Process-Based Model)

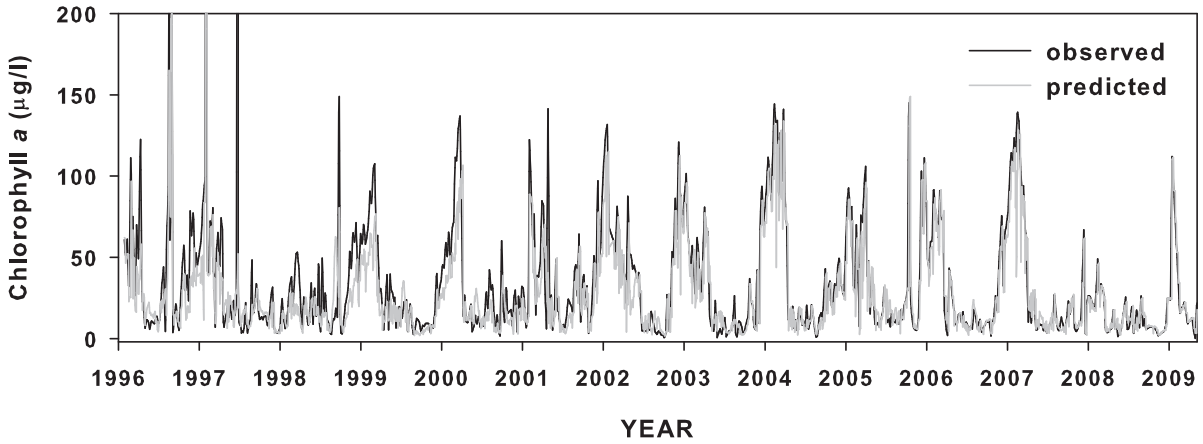

Fig. 3. Chlorophyll *a*, Actual vs Predicted (with GA Parameter Fitting)

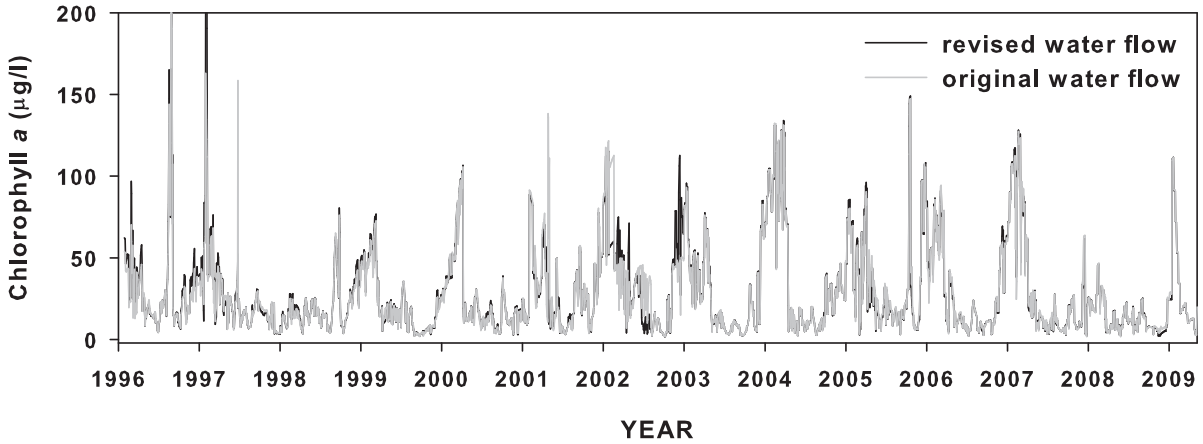

Fig. 4. Chlorophyll *a* Predictions, before and after Flow Modification

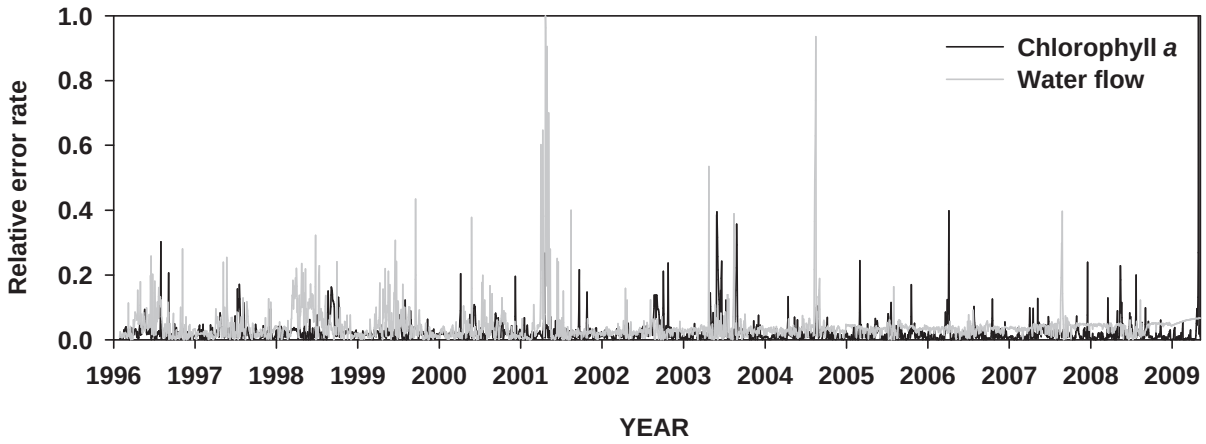

Fig. 5. Relative Error for Flow Rate and for Chlorophyll a Concentration

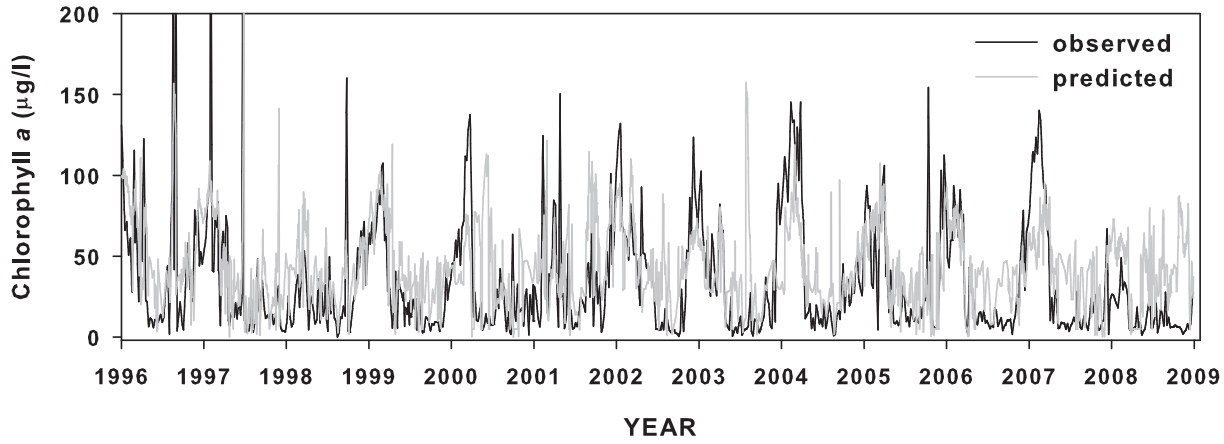

Fig. 7. Decision Rules learnt by Genetic Programming, 1996-2008

Programming [13] to forecast chlorophyll a concentration. Although the latter was developed over a different time period (1996-1998), it used the same environmental variables, and is sufficiently simple that it can readily be re-applied to our data. When it was extended to the period (1996-2008) consistent with the new model, its RMSE was 32.41 (compared to 21.34). Figure 7 shows the performance over this period. Lest it seem unfair to compare the model's RMSE outside the range for which it was developed, we note that its RMSE over 1996-1998 was 37.46; figure 6 shows the detail.

VI. Future Directions

A. Relative Importance of Flow and Growth Model Error

A key issue for our future work, is where to concentrate our effort to further reduce the error in our predictions. There are two possible areas: the flow and growth models. Initially, we thought the large errors in the flow model would be important, so we should invest effort in that area. Substantial improvements in flow data have yielded no improvement in algal growth accuracy; other indications also suggest that these errors may not be important. Why not, when the primary difference between our model and others is the incorporation of flow? We suspect this may be because large errors in flow are tied to high flows; but when these flows occur, the algal model correctly predicts near-zero values (because of flushing effects), thus resetting the error. Thus although large errors remain in the flow data, it seems this should not be the main focus of our work. However it remains an option, since improvements in the algal growth model could reinstate the flow model as a major source of error.

B. Model Revision of the Growth Model

Although we have described the work as using a simple GA to fit parameters of the model, this is only part of the story. However, the parameters are just one part of a larger-scale model, to which we plan to apply model revision

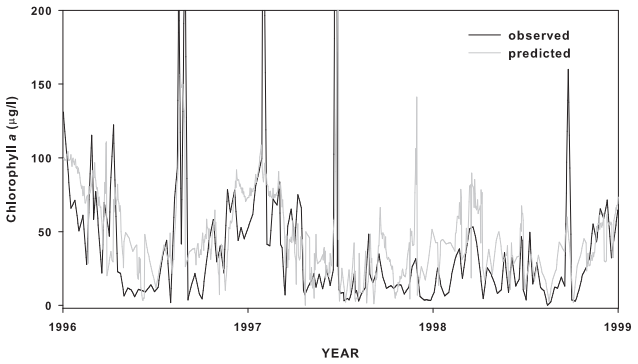

Fig. 6. Decision Rules learnt by Genetic Programming, 1996-1998

through a grammar-based GP system [26]. The grammar describes not only the process model, but also a space of possible modifications to the model, deeply embedding our environmental expert's knowledge of the most likely places where the process model might need adjustment. This will form the next stage of our research, aiming to optimise not merely the parameters, but also the structure, of the process model, and to do so in ways which allow for the paucity of data at our disposal.

VII. Conclusions

The results of the project so far are:

- Confirmation that combining process-based methods and evolutionary methods can improve prediction.
- A substantially better predictive model than was previously available from pure data mining approaches – not only more accurate, but more usefully accurate, in accurately predicting the timing and scale of algal blooms in almost all cases.
- An overall modelling approach that extensible beyond the Nakdong catchment to a wide range of rivers.

Water quality problems are a major global issue. Effective use of water resources in the face of conflicting demand is crucial to future economic, social and environmental wellbeing. A flexible and predictive river model could support water resource management across the globe. However designing and developing accurate models is generally difficult and expensive – so difficult that it is not often achieved. The combination of evolutionary methods with expert knowledge has enabled us to explore the search space much more rapidly than either could alone.

Acknowledgment

This work was partially supported by the Brain Korea 21 Project. Seoul National University Institute for Computer Science and Technology provided facilities for the research.

References

[1] J. L. Giraudel and S. Lek, "A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination," *Ecological Modelling*, vol. 146, no. 1-3, pp. 329–339, 2001.

[2] H.-M. Oh, C.-Y. Ahn, J.-W. Lee, T.-S. Chon, K. H. Choi, and Y.-S. Park, "Community patterning and identification of predominant factors in algal bloom in daechung reservoir (korea) using artificial neural networks," *Ecological Modelling*, vol. 203, no. 1-2, pp. 109–118, 2007.

[3] P. A. Whigham and F. Recknagel, "Predicting chlorophyll-a in freshwater lakes by hybridising process-based models and genetic algorithms," *Ecological Modelling*, vol. 146, no. 1-3, pp. 243–251, 2001.

[4] H. Cao, F. Recknagel, L. Cetin, and B. Zhang, "Process-based simulation library salmo-oo for lake ecosystems: Part 2: Multi-objective parameter optimization by evolutionary algorithms," *Ecological Informatics*, vol. 3, pp. 181–190, 2008.

[5] Y. Liu and X. Yao, "Evolving Neural Networks for Chlorophyll-a Prediction," in *Proceedings of the Fourth International Conference on Computational Intelligence and Multimedia Applications*. IEEE Computer Society, 2001, p. 185.

[6] A. Welk, F. Recknagel, H. Cao, W.-S. Chan, and A. Talib, "Rule-based agents for forecasting algal population dynamics in freshwater lakes discovered by hybrid evolutionary algorithms," *Ecological Informatics*, vol. 3, no. 1, pp. 46–54, 2008.

[7] F. Recknagel, J. Bobbin, P. Whigham, and H. Wilson, "Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes," *Journal of Hydroinformatics*, vol. 4, no. 2, pp. 125–133, 2002.

[8] C. L. Brown and T. O. Barnwell, "The enhanced stream water quality models qual2e and qual2e-uncas: documentation and user manual," Environmental Research Laboratory, Environmental Protection Agency, Athens, Georgia, Tech. Rep. EPA/600/3-85/040, 1987.

[9] S. S. Park and Y. S. Lee, "A water quality modeling study of the nakdong river, korea," *Ecological Modelling*, vol. 152, pp. 65–75, 2002.

[10] M. D. Yang, R. M. Sykes, and C. J. Merry, "Estimation of algal biological parameters using water quality modeling and spot satellite data," *Ecological Modelling*, vol. 125, no. 1, pp. 1 – 13, 2000.

[11] D. Solomatine, "Applications of data-driven modelling and machine learning in control of water resources," *Computational intelligence in control*, pp. 55–78, 2002.

[12] D.-K. Kim, K.-S. Jeong, P. A. Whigham, and G.-J. Joo, "Winter diatom blooms in a regulated river in south korea: explanations based on evolutionary computation," *Freshwater Biology*, vol. 52, no. 10, pp. 2021–2041, 2007.

[13] H. Cao, F. Recknagel, G.-J. Joo, and D.-K. Kim, "Discovery of predictive rule sets for chlorophyll-a dynamics in the nakdong river (korea) by means of the hybrid evolutionary algorithm hea," *Ecological Informatics*, vol. 1, no. 1, pp. 43–53, 2006.

[14] D. Savic, G. Walters, and J. Davidson, "A genetic programming approach to rainfall-runoff modelling," *Water Resources Management*, vol. 13, no. 3, pp. 219–231, 1999.

[15] A. S. Tokar and P. A. Johnson, "Rainfall-runoff modeling using artificial neural networks," *Journal of Hydrologic Engineering*, vol. 4, no. 3, pp. 232–239, 1999.

[16] K. Chau, "River stage forecasting with particle swarm optimization," in *Innovations in Applied Artificial Intelligence, 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE*, ser. Lecture Notes in Artificial Intgelligence, vol. 3029. Springer, 2004, pp. 1166–1173.

[17] K.-S. Jeong, D.-K. Kim, and G.-J. Joo, "Delayed influence of dam storage and discharge on the determination of seasonal proliferations of microcystis aeruginosa and stephanodiscus hantzschii in a regulated river system of the lower nakdong river (south korea)," *Water Research*, vol. 41, no. 6, pp. 1269 – 1279, 2007.

[18] G.-J. Joo, H.-W. Kim, K. Ha, and J.-K. Kim, "Long-term trend of the eutrophication of the lower nakdong river," *Korean Journal of Limnology*, vol. 30, no. supplement, pp. 472–480, 1997.

[19] Korean water management information system website. [Online]. Available: http://www.wamis.go.kr

[20] Korean meterological agency website. [Online]. Available: http://www.kma.go.kr/

[21] H. W. Paerl, J. L. Pinckney, J. M. Fear, and B. L. Peierls, "Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying neuse river estuary, north carolina, usa," *Marine Ecology Progress Series*, vol. 166, pp. 17–25, 1998.

[22] E. Everbecq, V. Gosselain, L. Viroux, and J. P. Descy, "Potamon: a dynamic model for predicting phytoplankton composition and biomass in lowland rivers," *Water Research*, vol. 35, no. 4, pp. 901–912, 2001.

[23] H. Pei and J. Ma, “Study on the algal dynamic model for west lake, hangzhou,” *Ecological Modelling*, vol. 148, no. 1, pp. 67–77, 2002.
[24] G. B. Arhonditsis and M. T. Brett, “Eutrophication model for lake washington (usa): Part i. model description and sensitivity analysis,” *Ecological Modelling*, vol. 187, no. 2-3, pp. 140–178, 2005.
[25] K.-S. Jeong, G.-J. Joo, H.-W. Kim, K. Ha, and F. Recknagel, “Prediction and elucidation of phytoplankton dynamics in the nakdong river (korea) by means of a recurrent artificial neural network,” *Ecological Modelling*, vol. 146, pp. 115–129, 2001.
[26] X. H. Nguyen, R. I. B. McKay, and D. L. Essam, “Representation and structural difficulty in genetic programming,” *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 2, pp. 157–166, April 2006.