

## 비선형 시계열 하천생태모형 개발과정 중 시간지연단계와 입력변수, 모형 예측성 간 관계평가

정광석 · 김동균<sup>1</sup> · 윤주덕 · 라금환<sup>2</sup> · 김현우<sup>2</sup> · 주기재\*

(부산대학교 자연과학대학 생명과학과, <sup>1</sup>서울대학교 공과대학 컴퓨터공학부,  
<sup>2</sup>순천대학교 사범대학 환경교육과)

**Relationship among Degree of Time-delay, Input Variables, and Model Predictability in the Development Process of Non-linear Ecological Model in a River Ecosystem. Jeong, Kwang-Seuk, Dong-Kyun Kim<sup>1</sup>, Ju-Duk Yoon, Geung-Hwan La<sup>2</sup>, Hyun-Woo Kim<sup>2</sup> and Gea-Jae Joo\***  
(Department of Biological Sciences, Pusan National University, Busan 609-735, Korea; <sup>1</sup>School of Computer Science & Engineering, Seoul National University, Seoul 151-742, Korea; <sup>2</sup>Department of Environmental Education, Suncheon National University, Suncheon 540-742, Korea)

In this study, we implemented an experimental approach of ecological model development in order to emphasize the importance of input variable selection with respect to time-delayed arrangement between input and output variables. Time-series modeling requires relevant input variable selection for the prediction of a specific output variable (e.g. density of a species). Inadequate variable utility for input often causes increase of model construction time and low efficiency of developed model when applied to real world representation. Therefore, for future prediction, researchers have to decide number of time-delay (e.g. months, weeks or days;  $t-n$ ) to predict a certain phenomenon at current time  $t$ . We prepared a total of 3,900 equation models produced by Time-Series Optimized Genetic Programming (TSOGP) algorithm, for the prediction of monthly averaged density of a potamic phytoplankton species *Stephanodiscus hantzschii*, considering future prediction from 0- (no future prediction) to 12-months ahead (interval by 1 month; 300 equations per each month-delay). From the investigation of model structure, input variable selectivity was obviously affected by the time-delay arrangement, and the model predictability was related with the type of input variables. From the results, we can conclude that, although Machine Learning (ML) algorithms which have popularly been used in Ecological Informatics (EI) provide high performance in future prediction of ecological entities, the efficiency of models would be lowered unless relevant input variables are selectively used.

**Key words :** input variable selectivity, ecological modeling, *Stephanodiscus hantzschii*, machine learning, ecological informatics, river ecosystems

생태모형은 생태계의 구조 혹은 기능의 변화 양상을 모의할 수 있는 개체를 의미한다. 생태모형을 개발하는 목적은 현상 파악이 힘들거나 매우 복잡한 관계에 놓여 있는 관심 대상 항목을 모의하여 그 현상의 개연성을 설명

하거나 앞으로의 변화 양상을 예측하기 위함이며 (Jeong et al., 2008), 생태계와 같이 실험적인 접근을 통하여 대상의 현상규명이 힘든 경우에 사용할 수 있다. 생태모형이 다루는 요소는 생물 및 무생물적 요소를 망라하며, 이

\* Corresponding author: Tel: 051) 510-2258, Fax: 051) 583-0172, E-mail: gjjoo@pusan.ac.kr

들 간의 관계를 수학적 혹은 경험적 관계로 재구성하여 구축된다(Recknagel, 2006). 따라서 효과적인 생태모형의 개발은 이러한 관계를 얼마나 잘 반영할 수 있는지, 그리고 관계가 반영된 모형이 얼마나 실 세계를 잘 대표할 수 있는지에 의하여 평가된다(Gevrey *et al.*, 2003).

하천의 식물플랑크톤 동태를 예측하는 생태모형은 여러 가지 생태모형들 중 가장 빈번히 개발되고 보완되어 온 주제 중 하나이다(Gallardo *et al.*, 2009; Ghermandi *et al.*, 2009; Kim *et al.*, 2009). 연중 식물플랑크톤 종의 동태는 뚜렷한 계절성이 존재하는 경우가 많으며, 발생하지 않는 시기에는 0으로 기록되는 경우가 대부분이어서 모형을 개발하는 데 많은 어려움을 야기한다(Joo and Jeong, 2005; Jeong *et al.*, 2006). 또한 수질의 한 요소로서 식물플랑크톤을 다룰 경우 먹이연쇄 관점에서의 접근(특히 섭식 효과)이 간과되는 경우 역시 많고, 하천의 특성상 흐름의 다양성이 크게 작용하여 예측이 힘든 요소 중 하나이다(Kim *et al.*, 2007a).

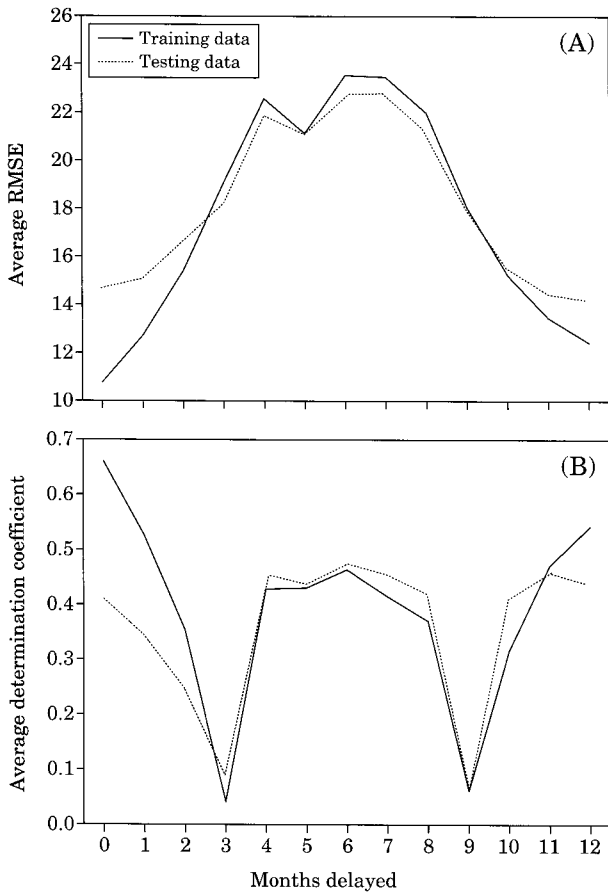
복잡한 생태학적 일계를 수학적으로 해석하는 것은 상당히 힘든 작업이므로, 지난 10여 년간 생태모형 개발 알고리즘으로 관심을 받게 된 생태정보학적 기법은 이러한 관점에서 상당히 많은 장점을 갖고 있다(Gevrey *et al.*, 2003). 생태정보학에서 주로 다루는 인공지능망, 유전자 프로그래밍 등의 알고리즘은 구조가 매우 유연하여 블랙박스(내부의 프로세스 혹은 메커니즘이 완전히 밝혀지지 않은 구조) 모형임에도 불구하고 모형 구축이 수월하며, 모형의 성능 역시 통계 모형 등에 비하여 양호한 것으로 알려져 있다(Jeong *et al.*, 2008). 하지만, 생태정보학적 기법이 직면하고 있는 문제점 중의 하나가 관심의 대상이 되는 생물분류군의 동태를 설명할 수 있는 최적의 입력 변수를 선택하는 작업이다. 이 과정은 입력변수와 출력변수 간의 관계를 해석해야 하는 과정이 포함되어 있으며, 특히 시계열 예측의 경우 시간차이에 따라 발생하는 인과관계까지 동시에 고려하여야 하기 때문에 상당히 많은 시간과 노력이 필요하다.

본 연구에서는 이러한 작업을 효율적으로 활용할 수 있는 기법을 낙동강 하류의 규조류 *Stephanodiscus hantzschii* 대변성 데이터에 적용하여 효율적인 입력변수 선택 방식을 검토하고자 한다. Time-Series Optimized Genetic Programming (TSOGP)은 알고리즘의 구조가 간단하면서 시계열 데이터를 이용한 예측 모형 개발에 효율적인 것으로 알려져 있으며(Jeong *et al.*, 2003; Kim *et al.*, 2007b), 본 연구에서는 TSOGP를 이용하여 *S. hantzschii*의 대변성을 모의할 수 있는 모형을 구축하고, 각 모형에서 활용된 입력 변수의 양상을 파악하여, 시계열 관계에서 생태

모형 개발을 위한 입력변수 선택의 중요성을 논하였다.

낙동강에서 매년 겨울 대변성을 일으키는 *S. hantzschii*는 전 세계의 담수생태계에서 빈번히 관찰되는 규조류 종으로, 대체로 봄에 호수에서 주로 나타나는 것으로 알려져 있다(Harris, 1986; Reynolds, 1992). 하지만 우리나라에서는 특이하게 겨울 기간에 성장률이 최대에 달하며, 발생 정도 역시 대변성을 일으키는 수준으로(전체 식물플랑크톤 군집의 90% 이상 점유), 우리나라의 겨울철 담수계 부영양화와 관련된 수질 문제에 있어서 대표적인 종이다(Kim *et al.*, 2008). 낙동강 하류에서 이 종은 매우 뚜렷한 계절성을 가지고 있어서 주로 수온이 낮은 겨울철에 발생하는 것으로 파악되며(Jeong *et al.*, 2008), 수온 및 유량 감소로 인한 포식압의 저하, 그리고 자원에 대한 경쟁 관계의 개선에 의해서 대변성이 결정되는 것으로 예상되고 있다(Kim *et al.*, 2007b). 이 종이 대변성을 일으키는 낙동강 하류역은 조절강 생태계의 대표적인 예로, 상류에 위치한 대형 다목적댐(안동, 임하, 합천, 남강댐)과 더불어 하구둑에 의하여 흐름이 매우 강하게 조절된다(Jeong *et al.*, 2007). 특히 하구둑 축조 이후 진행된 부영양화 상태는 여름 남조류 대변성을 촉발시킨 것으로 추정되며, *S. hantzschii*의 겨울 대변성 역시 이러한 영향에 의한 것으로 판단된다(Ha *et al.*, 2002).

규조류 *S. hantzschii*의 예측을 위해 본 연구에서 활용된 TSOGP 알고리즘은 유전자 프로그래밍(Genetic Programming; GP)을 이용하여 예측 모형식을 진화시키는 방식으로 모형을 구축한다(Whigham and Keukelaar, 2001). 이 알고리즘은 생물학적인 메커니즘에 기원한 학습알고리즘으로, 지정된 구조 안에서 무작위적으로 생성된 여러 개의 모형식들을 진화시켜 나간다. 진화의 과정은 교차(crossover), 돌연변이(mutation)와 같은 생물의 진화과정에서 볼 수 있는 형태의 변환과정을 통해 이루어진다. 그리하여 진화된 모형들을 학습 데이터에 대하여 평가하고, 예측성이 우수한 모형식들을 선별한다. 보정된 모형식들은 다시 입력 데이터에 대해서 평가되고, 평가 결과에 따라 새로이 진화되며, 이러한 과정을 주어진 진화의 길이(세대 수)만큼 반복하여 학습해 나간다. 또한 GP 알고리즘의 특징인 전방위 탐색(global search) 특성에 의해서 특정 데이터에 치중되어 학습되는 경우를 상대적으로 수월하게 배제할 수 있어, 최종적으로 확보된 모형식은 비교적 국소적 최적값(local optima)으로부터 비교적 자유로우면서 해당 생태계의 특성을 잘 반영할 수 있는 모형이 된다. 최종적으로 학습이 완료된 모형식은 별도의 평가 데이터에 대해 적용되며, 이를 통하여 최적의 모형을 선별하게 된다. TSOGP를 담수생태계에 적



**Fig. 1.** Changes of average model fitness ( $n=300$  per each of months delayed; RMSE and  $r^2$ ) in terms of months delayed during the model development process. A, RMSE changes; B, determination coefficient ( $r^2$ ) changes.

용하는 사례는 Jeong *et al.* (2003)과 Kim *et al.* (2007b)에 잘 설명되어 있다.

본 연구에서는 *S. hantzschii*의 대변성을 예측할 수 있는 생태모형을 TSOGP로 구축하기 위하여 1995년 1월부터 2004년 12월까지 총 10년간 27개의 환경 변수에 대해서 관측된 자료를 활용하였다(Jeong *et al.*, 2008로부터 데이터 인용). 이들 변수는 기상, 수문, 물리화학 및 생물학적 요소 등 크게 4개의 그룹으로 나뉜다(Appendix 1 참조). TSOGP 알고리즘은 월평균으로 전환된 입력변수 값을 이용하여 *S. hantzschii*의 동태를 예측할 수 있는 수식 모형을 추출하였으며, 각 모형에 활용된 입력변수를 검토하여 어떠한 변수가 식물플랑크톤 예측에 효과적인지 검토하였다. 또한 시계열 관계에 대한 평가를 위하여, 당월 예측부터 12개월 후 예측까지 총 13개의 경우에 대해서 각각 300개의 모형식을 구축하여, 시간 지연 상황

에 따라서 모형의 예측성과 입력변수 선택 양상이 어떻게 바뀌는지를 동시에 검토하였다. 교차율과 돌연변이율은 각각 95% 및 5%으로 고정하였으며, 모형의 성능은 실측 및 예측값에 대한 결정계수(determination coefficient,  $r^2$ )와 Root Mean Squared Error (RMSE) 값을 이용하여 평가하였다. 변수의 선택성은 하나의 식에 특정 변수가 몇 회 사용되었는지를 파악하였으며, 1개의 식에 동일 변수가 2회 이상 사용되면 변수가 식에 나타난 횟수 그대로 평가하였다. 따라서 각 변수의 최대 출현 빈도는 300회를 넘을 수 있다. TSOGP 알고리즘을 활용한 낙동강 플랑크톤 모형 예측 사례를 Jeong *et al.* (2003)과 Kim *et al.* (2007b)에서 찾아볼 수 있으며, 모형의 구축 과정은 이들 연구에서 제시된 바를 따라 수행되었다.

Fig. 1에 각 월별 300개 모형들이 학습 및 평가 데이터에 대해서 산출한 결정계수와 RMSE의 평균 양상이 월별 지연상황에 따라 나타나 있다. RMSE의 경우 학습 및 평가 데이터에 대해서 모두 3~8개월 후를 예측하도록 하였을 때 RMSE가 높아졌으며, 0~2개월 그리고 9~12개월 후를 예측하는 경우에서 상대적으로 낮은 RMSE 값이 계산되었다. 데이터 구분에 따라서는 RMSE의 값이 다소 차이가 발생하였으며, 대체로 평가 데이터에 대한 RMSE 값이 학습데이터보다 높아지는 경향이 존재하였다. 하지만 RMSE가 비교적 높은 경우(즉, 3~8개월 후 예측)에서는 이러한 상황이 역전되었으나 큰 차이는 보이지 않았다. 결정계수는 다소 다른 양상을 보였으며, 학습 데이터에는 가장 높은 값은 미래 예측을 실시하지 않은 경우에서 나타났다. 하지만 평가 데이터에서는 비교적 결정계수가 높게 계산된 경우들 모두에서 유사한 수준으로 산출되었다. 3개월과 9개월 후를 예측시킨 경우에 결정계수가 가장 낮게 나타났고, 그 외의 기간에서는 대체로 유사한 수준에 있었다. 따라서 미래 예측을 실시할 경우 입력 데이터에 어느 수준의 시간 지연을 두는가에 따라서 모형의 예측성이 달라지는 것을 알 수 있다. RMSE와 결정계수는 시간 지연입력 상황에 따라서 서로 다른 형태의 결과를 보였는데, 이는 4~8개월 미래 예측에서는 상대적으로 계절성이 뚜렷한 요소(예, 수온)가 원래와는 반대의 형태로 출력 데이터에 관계지어짐에 따라 주기성을 잘 파악하는지를 판단할 수 있는 결정계수에서는 상대적으로 높은 값이 나타났으나 출력 데이터의 관측값과 예측값 사이의 차이를 이용하여 평가하는 RMSE 값은 상대적으로 불량해진 것으로 해석된다.

미래 예측을 위한 각 월별 지연상황에 따른 입력변수 선택성을 평가한 결과(Table 1), 각 입력변수별 최다 선택횟수가 나타난 지연상황은 서로 다른 것으로 나타났다.

**Table 1.** Number of variable selection by TSOGP algorithm in responsible to the different time-lags for the prediction of *S. hantzschii*. Acronyms or abbreviations of the variables can be found in Appendix 1 (list of variables used in modelling).

Categories	Variables	Months delayed												
		No	1	2	3	4	5	6	7	8	9	10	11	12
Climate	SOI	20	18	18	30	97	88	74	17	5	15	18	27	35
	PDO	29	26	17	14	13	20	67	193	279	118	79	33	13
	MEI	47	22	17	9	27	59	29	16	8	21	22	7	22
	Nino3	53	35	27	12	23	47	17	13	14	13	22	18	29
	Nino34	36	15	12	11	30	51	24	13	7	9	39	27	16
	AO	53	28	24	18	69	45	61	10	8	26	85	24	37
Hydrological	Rainfall	36	135	26	5	7	4	0	26	21	14	13	34	82
	Dam W.L.	23	17	13	20	0	6	2	5	5	10	1	6	19
	Dam Stor.	14	22	148	140	1	8	9	35	39	145	3	3	7
	Dam Inflow	45	139	189	14	4	20	1	16	1	0	0	14	70
	Dam Disch.	5	12	36	181	4	15	6	10	2	8	2	5	31
	River Flow	41	2	8	25	0	12	0	1	16	19	223	187	108
	River W.Vel.	51	15	17	45	37	54	27	16	23	19	18	46	38
	River W.L.	21	12	19	81	15	9	7	5	14	22	12	12	16
Physico-chemical	Water Temp.	291	125	16	37	306	332	320	304	266	2	16	224	256
	DO	111	39	103	25	9	35	17	47	26	28	15	43	29
	%DO Sat.	87	8	11	26	6	17	1	5	4	17	2	1	6
	pH	27	8	23	53	3	16	16	17	8	11	12	14	16
	Secchi Depth	26	31	16	4	0	31	2	4	5	13	2	4	11
	Conductivity	92	30	5	2	1	22	0	1	2	26	69	81	99
	Alkalinity	93	67	31	74	2	31	1	5	3	66	23	85	94
	Turbidity	18	8	29	38	2	11	2	8	8	30	28	6	6
	Nitrate	25	29	78	51	0	92	22	5	4	8	22	55	48
	Silica	67	18	40	30	0	14	8	94	26	114	5	9	26
Phosphate	152	144	131	84	0	37	1	3	4	3	1	21	16	
Biological	CFR Phytopl.	1	4	0	0	0	0	0	0	0	0	0	2	0
	CFR Bacteria	4	3	2	1	0	2	0	10	26	3	0	4	2

이러한 상황은 같은 그룹 내에 속한 변수들에 대해서도 서로 다르게 나타났다. 모든 변수들 중 수온이 가장 빈번하게 활용되었으며(총 2,495회), 다음으로 PDO가 자주 선택되었다(총 901회; 정식 명칭은 Appendix 1 참조). 생물요소를 제외한 모든 입력변수가 최소 100여 회 이상 활용되었으며, 식물플랑크톤 및 박테리아에 대한 섭식률 변수는 매우 낮은 빈도로 활용되었다. 각 그룹별로는 기상요소에서 PDO, 수문요소에서 하천유량, 물리화학적 요소에서 수온, 생물학적 요소에서 박테리아 섭식률 변수가 빈번히 활용되었다. PDO는 대체로 7~9개월 예측에서 많이 활용되었으며, 나머지 경우에는 다른 기상요소의 변수들과 유사한 상황에 있었다. 수문요소에서는 변수들이 서로 다른 양상을 명확하게 보여주고 있었으며, 강우량, 댐저수량 등 수문요소는 대체로 0~3개월 예측에서 빈번히 활용되었다. 물리화학적 요소에서 수온은 2~3개월, 9~10개월을 제외한 모든 지연상황에서 빈번

히 활용되었으며, 그 외의 변수들은 0~3개월, 9~12개월 지연에서 비교적 많이 선택되었다. 마지막으로 생물요소의 경우 식물플랑크톤 섭식률은 거의 선택되지 않았으나 박테리아 섭식률은 7~8개월 예측에서 낮은 빈도로 선택되는 경향을 보였다.

모형의 성능이 시간지연 상황에 따라서 서로 다르게 나타나는 것은 입력 및 출력변수 간의 시계열 관계에 의한 것으로 사료된다. 대부분의 입력변수는 계절성을 가지고 있으며, 특히 강우량, 유량과 같은 변수들은 매우 뚜렷한 계절성이 존재한다. 따라서 입력 및 출력변수 간의 계절성이 엇갈리는 경우(본 연구에서는 3~8개월 미래예측의 경우) 오히려 모형의 예측성이 현저히 저하되는 것으로 나타났으며, 이와 반대로 12개월 예측에서 상대적으로 예측성이 높아지는 것은 *S. hantzschii*의 동태와 계절성이 겹치는 변수(수온, 강우량, 유량 등)들이 존재하기 때문으로 여겨진다. 따라서 미래 예측 시간 간격 정도에 따

라서 활용할 입력변수를 달리 고려하여야 한다.

시계열 요소를 반영한 생태모형의 구축은 크게 두 가지로 나뉠 수 있으며, 자기회귀형(AR, MA, ARMA 및 ARIMA 등), 혹은 다변량 회귀식 형태로 구축할 수 있다. 자기회귀형의 경우 특정변수의 과거 변동을 이용하여 미래를 예측하는 과정으로, 보통 주가 예측, 인구 추이 등에서 많이 활용된다. 이에 반해 생태모형 개발의 경우 환경변수의 제어 목적을 위해서 다변량 회귀식 형태의 모형을 선호하며, 이를 위해서 과거의 환경 상태(유량, 수온, 영양염류 농도 등)를 당월 혹은 당일의 특정변수(본 연구에서는 *S. hantzschii* 밀도)에 연관시켜 분석한다(Jeong *et al.*, 2008). 따라서 특정변수의 변동을 잘 설명할 수 있는 입력변수를 활용하는 것이 관건이며, 대체로 다양한 종류의 입력변수를 활용하려는 경향이 강해진다. 하지만 생태모형 개발에서 입력변수의 수가 늘어나는 것과 모형의 성능 간에는 큰 관계가 없으며, 오히려 입력변수가 많아질수록 모형의 구조가 복잡해지고 성능이 저하되는 경우가 존재한다고 알려져 있다(Jørgensen, 1997). 따라서 Jeong *et al.* (2006)은 예측성을 높여줄 수 있으면서 비용/편익관점에서 효율적인 조사가 가능한 환경변수를 모형 개발에서 이용하는 것이 현명하다고 주장하였다. 본 연구의 결과도 이러한 상황을 잘 반영하고 있으며, 많은 변수를 활용한다 하더라도 모형에서 빈번히 활용되는 변수는 국한되어 있음을 알 수 있다. 특히 시간지연 상황에 따라서 활용되는 변수의 종류가 서로 다른 점은 시계열 통계 방법 중 교차상관(Cross-correlation) 분석에서 볼 수 있는 변수 종류에 따라 서로 다른 시간 관계를 갖는다는 부분과 거의 동일한 개념으로 볼 수 있다. 따라서 적절한 수준의 미래예측 단계를 우선 설정하고, 이에 맞추어 입력변수를 활용하는 것이 효율적인 생태모형 개발상에서 중요하다 할 수 있다. 적절한 변수 선택이 우선적으로 전제되면 생태정보학 알고리즘을 이용하여 보다 장기간의 미래 예측(6개월~1년 이상) 역시 어느 정도 가능할 것으로 판단된다.

본 연구에서는 이와 같이 TSOGP와 같은 생태정보학적 기법이 변수 선택성을 어떻게 반영하여 주는지를 잘 보여 준다는 결론을 얻을 수 있었으며, 따라서 통계적인 모형 혹은 프로세스 기반의 모형(예, 미분방정식 모형 등) 등을 개발할 때 이와 같은 접근방법을 통하여 보다 효율적인 입력-출력변수 간 관계를 시계열 관점을 통하여 파악하여 정보를 반영할 수 있을 것으로 예상된다. 여기에 덧붙여 앞으로 유전자 프로그래밍 개발상의 여러가지 조건(본 연구에서는 일정 수준으로 결정되어 있던 교차율, 돌연변이율 등)을 변화시킴에 따라 보다 효율적인 입력변수 선택성

이 파악될 수 있는지에 대해 심도있는 연구가 필요하다.

## 사 사

본 연구를 수행하기 위해서 TSOGP 알고리즘이 탑재된 모델링 셸(TimeSeriesToolbox)을 제공하여 주신 University of Otago의 Peter Whigham 교수님께 감사 말씀드립니다. 본 연구는 부산대학교 연구지원사업에 의하여 이루어졌습니다.

## 인 용 문 헌

- Gallardo, B., S. Gascon, M. Gonzalez-Sanchis, A. Cabezas and F.A. Comin. 2009. Modelling the response of flood-plain aquatic assemblages across the lateral hydrological connectivity gradient. *Mar. Freshw. Res.* **60**: 924-935.
- Gevrey, M., I. Dimopoulos and S. Lek. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modelling* **160**: 249-264.
- Ghermandi, A., V. Vandenbergh, L. Benedetti, W. Bauwens and P.A. Vanrolleghem. 2009. Model-based assessment of shading effect by riparian vegetation on river water quality. *Ecol. Eng.* **35**: 92-104.
- Ha, K., M.-H. Jang and G.-J. Joo. 2002. Spatial and temporal dynamics of phytoplankton communities along a regulated river system, the Nakdong River, Korea. *Hydrobiologia* **470**: 235-245.
- Harris, G.P. 1986. *Phytoplankton Ecology: Structure, Function and Fluctuation*. Chapman and Hall, New York.
- Jørgensen, S.E. 1997. *Integration of Ecosystem Theories: a Pattern*. Kluwer Academic Publishers, Dordrecht.
- Jeong, K.-S., D.-K. Kim and G.-J. Joo. 2006. River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system. *Ecol. Inform.* **1**: 235-245.
- Jeong, K.-S., D.-K. Kim and G.-J. Joo. 2007. 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 Res.* **41**: 1269-1279.
- Jeong, K.S., D.K. Kim, J.M. Jung, M.C. Kim and G.J. Joo. 2008. Non-linear autoregressive modelling by Temporal

- Recurrent Neural Networks for the prediction of freshwater phytoplankton dynamics. *Ecol. Modelling* **211**: 292-300.
- Jeong, K.-S., D.-K. Kim, P. Whigham and G.-J. Joo. 2003. Modelling *Microcystis aeruginosa* bloom dynamics in the Nakdong River by means of evolutionary computation and statistical approach. *Ecol. Modelling* **161**: 67-78.
- Joo, G.-J. and K.-S. Jeong. 2005. Modelling community changes of cyanobacteria in a flow regulated river (the lower Nakdong River, S. Korea) by means of a Self-Organizing Map (SOM), p. 273-287. *In: Modelling Community Structure in Freshwater Ecosystems* (Lek, S., M. Scardi, P.F.M. Verdonshot, J.-P. Descy and Y.-S. Park, eds.). Springer, Berlin.
- Kim, D.-K., H. Cao, K.-S. Jeong, F. Recknagel and G.-J. Joo. 2007a. Predictive function and rules for population dynamics of *Microcystis aeruginosa* in the regulated Nakdong River (South Korea), discovered by evolutionary algorithms. *Ecol. Modeling* **203**: 147-156.
- Kim, D.-K., K.-S. Jeong, P.A. Whigham and G.-J. Joo. 2007b. Winter diatom blooms in a regulated river in South Korea: explanations based on evolutionary computation. *Freshwater Biol.* **52**: 2021-2041.
- Kim, M.-C., G.-H. La, H.-W. Kim, K.-S. Jeong, D.-K. Kim and G.-J. Joo. 2008. The effect of water temperature on proliferation of *Stephanodiscus* sp. in vitro from the Nakdong River, South Korea. *Korean Journal of Limnology* **41**: 26-33.
- Kim, M.-C., K.-S. Jeong, D.-K. Kang, D.-K. Kim, H.-S. Shin and G.-J. Joo. 2009. Time-series relationship between hydrology and phytoplankton biomass in a regulated river (Nakdong River) using time-series statistics. *J. Ecol. Field Biol.* (in press).
- Recknagel, F. (Editor). 2006. *Ecological Informatics: Scope, Techniques and Applications*. Springer, Berlin.
- Reynolds, C.S. 1992. Algae, p. 195-215. *In: The River Handbook: Hydrological and Ecological Principles* (Calow, P. and G.E. Petts, eds.). Blackwell Scientific Publication, Oxford.
- Whigham, P.A. and J. Keukelaar. 2001. Evolving structure-optimizing content. *Proceedings of the Congress on Evolutionary Computational Biology and Chemistry*, p. 1228-1235.

(Manuscript received 2 January 2010,  
Revision accepted 18 March 2010)

**Appendix 1.** List of input variables and their acronyms used in the model development.

Categories	Variables names	Acronyms /abbreviations	Units
Climatological	Southern Oscillation Index	SOI	
	Pacific Decadal Oscillation	PDO	
	Multivariate ENSO Index	MEI	
	Nino 3	Nino3	
	Nino 3.4	Nino34	
	Arctic Oscillation	AO	
Hydrological	Rainfall	Rainfall	mm
	Dam water level	Dam W.L.	m
	Dam Storage	Dam Stor.	ton
	Dam Inflow	Dam Inflow	ton sec <sup>-1</sup>
	Dam Discharge	Dam Disch.	ton sec <sup>-1</sup>
	River Flow	River Flow	ton sec <sup>-1</sup>
	River water velocity	River W.Vel.	m sec <sup>-1</sup>
	River water level	River W.L.	m
Physico-chemical	Water Temperature	Water Temp.	°C
	DO concentration	DO	mg L <sup>-1</sup>
	%DO Saturation	%DO Sat.	%
	pH	pH	
	Secchi Depth	Secchi Depth	cm
	Conductivity	Conductivity	μS cm <sup>-1</sup>
	Alkalinity	Alkalinity	mg L <sup>-1</sup>
	Turbidity	Turbidity	NTU
	Nitrate	Nitrate	mg L <sup>-1</sup>
	Silica	Silica	mg L <sup>-1</sup>
Phosphate	Phosphate	mg L <sup>-1</sup>	
Biological	Community filtration rate for phytoplankton	CFR Phytopl.	μg C L <sup>-1</sup> day <sup>-1</sup>
	Community filtration rate for bacteria	CFR Bacteria	μg C L <sup>-1</sup> day <sup>-1</sup>