

호소내 Chl-a의 일단위 예측을 위한 신경망 모형의 적정 파라미터 평가

연인성 · 홍지영 · 문현생[†]

국립환경과학원 수질통합관리센터

Estimating Optimal Parameters of Artificial Neural Networks for the Daily Forecasting of the Chlorophyll-a in a Reservoir

Insung Yeon · Jiyoung Hong · Hyunsaing Mun[†]

Water Quality Control Center, National Institute of Environmental Research

(Received 6 May 2011, Revised 7 July 2011, Accepted 7 July 2011)

Abstract

Algal blooms have caused problems for drinking water as well as eutrophication. However it is difficult to control algal blooms by current warning manual in rainy season because the algal blooms happen in a few days. The water quality data, which have high correlations with Chlorophyll-a on Daecheongho station, were analyzed and chosen as input data of Artificial Neural Networks (ANN) for training pattern changes. ANN was applied to early forecasting of algal blooms, and ANN was assessed by forecasting errors. Water temperature, pH and Dissolved oxygen were important factors in the cross correlation analysis. Some water quality items like Total phosphorus and Total nitrogen showed similar pattern to the Chlorophyll-a changes with time lag. ANN model (No. 3), which was calibrated by water temperature, pH and DO data, showed lowest error. The combination of 1 day, 3 days, 7 days forecasting makes outputs more stable. When automatic monitoring data were used for algal bloom forecasting in Daecheong reservoir, ANN model must be trained by just input data which have high correlation with Chlorophyll-a concentration. Modular type model, which is combined with the output of each model, can be effectively used for stable forecasting.

keywords : Algal blooms, Artificial neural networks, Chlorophyll-a, Cross correlation, Daily forecasting, Parameter

1. 서론

최근 기후변화의 관심이 증대되면서 수온과 강우에 영향을 받는 수질 및 조류발생에 대하여 정체수역에서의 실시간 감시가 더욱 중요해지고 있는 시점이다. 특히 음용수를 취수하는 호소에서 수질관리의 일차적인 목표 중 하나가 조류발생을 조기에 감지하고 조류확산에 적절하게 대응하는 것이다.

조류대발생의 문제를 해결하기 위해서는 유역의 오염원을 줄이는 것이 근본적인 해결방법이므로, 오염배출량의 상류 유입에 따른 하류의 수질변화를 분석하는 데에 수질모델이 많이 사용된다. 조류발생 등으로 매년 문제가 되고 있는 대청호의 경우에도 CE-QUAL-W2, Water quality analysis simulation program (WASP) 등의 수질모델을 이용한 연구(서동일, 1997; 심순보 등, 1995; 유병로, 2001; 유순주 등, 2006; 정세용 등, 2007)들이 수행된 바 있다. 한편, 조류발생에 대한 근본적인 저감대책이 어려운 상황에서는 조류대발생에 대한 시기, 규모에 대한 조기예측이 매우 중요하며, 호소 및 취·정수장에서의 수질관리를 위해서 실

시간 및 단기예측 기법이 활용될 수 있다.

여름철에는 조류가 불과 수일 만에 평상 시 농도에서 대발생까지 급성장하기 때문에 취·정수장에서 현재의 조류정보에 따라 대응하는 것이 쉽지 않다. 세포수 확인 및 실험실 분석 절차를 만족하기 위해서는 장시간이 소요되며, 목적에 따라 실시간 관리 및 사전대응이 요구되는 현장에서는 사전 예측이 필수적이라 할 수 있다. 이러한 환경 속에서 조류대발생을 조기에 예측하고자 하는 다양한 연구가 국내·외에서 시도되고 있다. 조기예측의 방법 중 하나는 인공신경망을 이용한 예측방법으로 입·출력이 간편할 뿐 아니라, 물리, 화학, 생물학적인 수질모형과는 달리 유역내의 미래 유량이나 부하량 등이 없어도 과거의 축적된 자료를 통해 미래의 수질을 예측할 수 있다.

신경망 모형의 장점은 (i) 구성이 용이하고, (ii) 응답이 빨라서 실시간 운영에 적합할 뿐만 아니라 (iii) 동적(dynamic)이고, 비선형(non-linear)적이며, 잡음(noisy) 섞인 자료들, 특히 물리적/생물학적 관계가 명확하지 않은 상태에서도 적용할 수 있다. 매우 복잡하고, 대단히 비선형적인 조류 변동(algal dynamics)을 모델링하는데 효과적이다(Lee et al., 2003).

호수의 영양상태는 오염물질의 유입형태, 유입시기, 체류시간, 순환률 등과 같은 여러가지 요인들에 의해 영향을 받게 되는데, 대청호는 호수길이가 길고, 폭이 좁은 형태의

[†] To whom correspondence should be addressed.
mungis@korea.kr

특성과 기상 등의 환경요인 및 인위적 수량조절 등에 의한 호수환경 변화에 따라 상류, 중류, 하류 및 만입수역 등 수역구간별로 영양상태가 다르게 나타나며 조류의 발생 양상 역시 수역별로 다양하게 나타나는 것으로 알려져 있다(국립환경과학원, 2009). 신경망 기법을 통한 접근 방법은 실시간으로 측정되는 방대한 양의 수질자동측정 자료에 기반하여 매년마다 다양한 원인과 양상을 보이는 대청호 Chlorophyll-a (Chl-a)의 농도 변화를 해석하기 위한 효과적인 도구가 될 것으로 판단된다.

호주에서는 하구관리 프로그램에서 조류발생 조기경보를 위해 신경망 기법을 적용(Hornsby shire council: HSC, 2008) 했으며, 조류(algae)의 종 분류 및 세포 수, Chl-a 농도예측을 위해 유사 연구들이 수행되었다. Recknagel 등(1997)에 의한 종별 세포수의 예측 이후, Maier 등(1998)에 의한 남조류 세포수의 주단위 예측, Karul 등(2000)에 의한 Chl-a 농도와 조류의 종 예측, Lee 등(2003)에 의한 홍콩 해안의 조류 세포수 주단위 예측, Velo-Suarez and Gutierrez-Estrada (2007)에 의한 해로운 남조류의 주단위 예측, Chen 등(2008)에 의한 실측치와 위성영상간 Chl-a 농도 보정(유전자 알고리즘), Palani 등(2008)에 의한 싱가포르 해안 Chl-a 농도의 주간별 예측이 이어졌다. 그러나 하천이나 호소가 아닌 연안에서의 적조 예측이 대부분이며, 주(weekly) 단위 자료를 사용하였으므로 호소내에서 수일 사이에 급격히 증가하는 조류대발생을 예측하기 위해서는 보다 짧은 측정주기와 측정위치를 고려해야 효과적일 것으로 판단된다. 국내에서도 인공신경망을 이용한 조류예측 연구가 팔당호(박해경 등, 2005)와 대청호(국립환경과학원, 2008; 정건희, 2001; Oh 등, 2007), 낙동강(김동균, 2008; 정광석, 2004)에 적용된 바 있으며, 조류와 인공신경망을 활용한 독성감시(정종혁 등, 2008) 등의 연구가 수행된 바 있다.

본 연구에서는 대청호를 대상으로 일별 Chl-a를 예측하기 위해서 연속 측정된 자료를 학습한 신경망 모형을 적용하고, 예측에 효과적인 변수 및 적절한 모형의 구조를 검토하였다. 상관성 분석을 통해서 수질자동측정망의 Chl-a와 상관성이 높은 수질항목을 분석하고, 변화패턴을 학습하기 위한 입력 자료들을 선정하였다. 1일, 3일, 7일의 단기간 선행예측을 통해서 조류발생을 조기예측하기 위하여 은닉층 노드수 및 학습회수를 변경해 가며 신경망 모형을 적용하고, 입력 자료의 변경에 따른 적용성을 평가하였다.

2. 연구방법

2.1. 신경망 기법

신경망 모형에서 가장 많이 이용되는 다층신경망 모형은 주로 Back Propagation (BP, 오차역전파) 알고리즘을 이용하여 학습을 수행하는데 입력패턴을 이용해 신경망 네트워크가 계산한 출력값과 목표값과의 오차를 최소화시키는 방향으로 연결강도를 조절하는 방법이다. 신경망 네트워크는 입력층, 은닉층, 출력층 방향으로 연결되어 있으며, 출력값을 생성하는 전달함수는 시그모이드(sigmoid) 함수가 많이

사용된다. 학습하는 동안 목표값과 실제 출력값 사이의 계산된 오차 정보 네트워크를 통하여 역방향으로 전달되며 연결 가중치에 영향을 준다. 입력에서 출력과정을 다음과 같이 나타낼 수 있다.

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

$$H = [h_1, h_2, \dots, h_m] \quad (2)$$

$$Y = [y_1, y_2, \dots, y_k] \quad (3)$$

여기서 X 는 입력층 벡터, H 는 은닉층 벡터, Y 는 출력층 벡터이다. 목표치 T 와 최종출력 Y 를 비교하여 식 (4)와 같이 제곱오차 E 를 계산한다.

$$E = \frac{1}{2}(T - Y)^2 \quad (4)$$

Levenberg-marquadt 알고리즘은 최소제곱오차를 이용하는 과정에서 자코비안 행렬과 몇 개의 파라미터를 이용함으로써 학습속도가 빠른 것으로 알려져 있다. 우선, 모든 입력값에 대한 에러들의 제곱의 합, $e_i^2(W)$ 을 계산한다.

$$E(W) = \sum_{i=1}^m e_i^2(W) = \|f(W)\|^2 \quad (5)$$

여기서 $e_i^2(W) = (y_{di} - y_i)^2$, y_{di} 는 i 노드에서의 목표값, y_i 는 i 노드에서의 출력값이다. J_k 는 W_k 에서의 자코비안 행렬, λ 는 Marquadt 파라미터, I 는 단위행렬일 때, 연결강도의 조정량 ΔW_k 는 자코비안 행렬을 통한 식 (6)에 의해 계산한다.

$$\Delta W_k = -(J_k^T f(W_k))(J_k^T J_k + \lambda I)^{-1} \quad (6)$$

λ 를 작은 값부터 시작하며, 계산 결과에 따라 식 (7)에 의해 W_{k+1} 를 계산한다.

$$W_{k+1} = W_k + \Delta W_k \quad (7)$$

이 때 $E(W_k + \Delta W_k) \geq E(W_k)$ 이면, λ 를 증가시키고 식 (6)부터 다시 계산하며, $E(W_k + \Delta W_k) < E(W_k)$ 이면 λ 를 감소시키고, $W_k = W_k + \Delta W_k$ 로 업데이트한다.

오차 E 가 감소하도록 은닉층과 출력층간의 연결강도 변화량 ΔW , 입력층과 은닉층 간의 연결강도 변화량 ΔV 를 계산하여 다음 학습단계의 조정된 연결강도 W 와 V 를 구한다.

2.2. 자료 및 방법

본 연구의 시범대상 지역인 대청호의 조류발생 예측에는

2007년부터 2009년까지의 수질자료 중 조류 발생기간 동안의 Chl-a, 수온, pH, Dissolved oxygen (DO), 물벼룩 임펄스(Impulse), Total nitrogen (TN), Total phosphorus (TP)와 수문인자 자료인 댐 수위, 댐 유입량, 강우량 자료를 이용하였다. 수질자료는 환경부에서 실시간으로 운영하고 있는 국가수질자동측정망으로부터 일평균 자료를 산출하여 사용하였으며, 수문자료는 WAMIS (Water management information system, 국가수자원관리 종합정보시스템)에서 일자료를 수집하여 활용하였다. 예측 모형을 평가하기 위해서 2007년 140개 쌍의 일자료를 모형의 학습자료로, 2008년 136개 쌍의 일자료를 모형의 선정을 위한 검증자료로, 2009년 190개 쌍의 일자료를 실제모의를 가정한 테스트 자료로 사용하였다. 적용 자료의 특성을 Fig. 1에 나타내었다.

수온과 Impulse를 비롯한 몇 개의 수질항목들은 매년 유사한 범위 내에서 평균을 기준으로 일정하게 분포된 양상을 보이며, 수리·수문 인자인 유입량, 방류량, 강우량은 여름철 호우로 인한 극치의 자료들이 평균에 큰 영향을 미치는 것으로 나타났다. 연도별 변화양상을 보면, 2007년부터 2009년까지 수리·수문 인자의 상한값과 중앙값, 표준편차는 작아졌으며, 반대로 Chl-a의 상한값, 중앙값, 평균 농도는 점차 증가하여 큰 강우보다는 작은 강우가 Chl-a의 증가에 큰 역할을 하고 있음을 추정할 수 있다.

그러나 이상의 분석으로는 시간에 따른 변화를 고려할 수가 없으므로 먼저 Chl-a와 연관성 있는 수질 항목을 분석하기 위하여 그래프를 통한 변동추이를 비교하고, 자료간의 교차상관분석을 수행하였다. 이를 통해 상관성 큰 자료

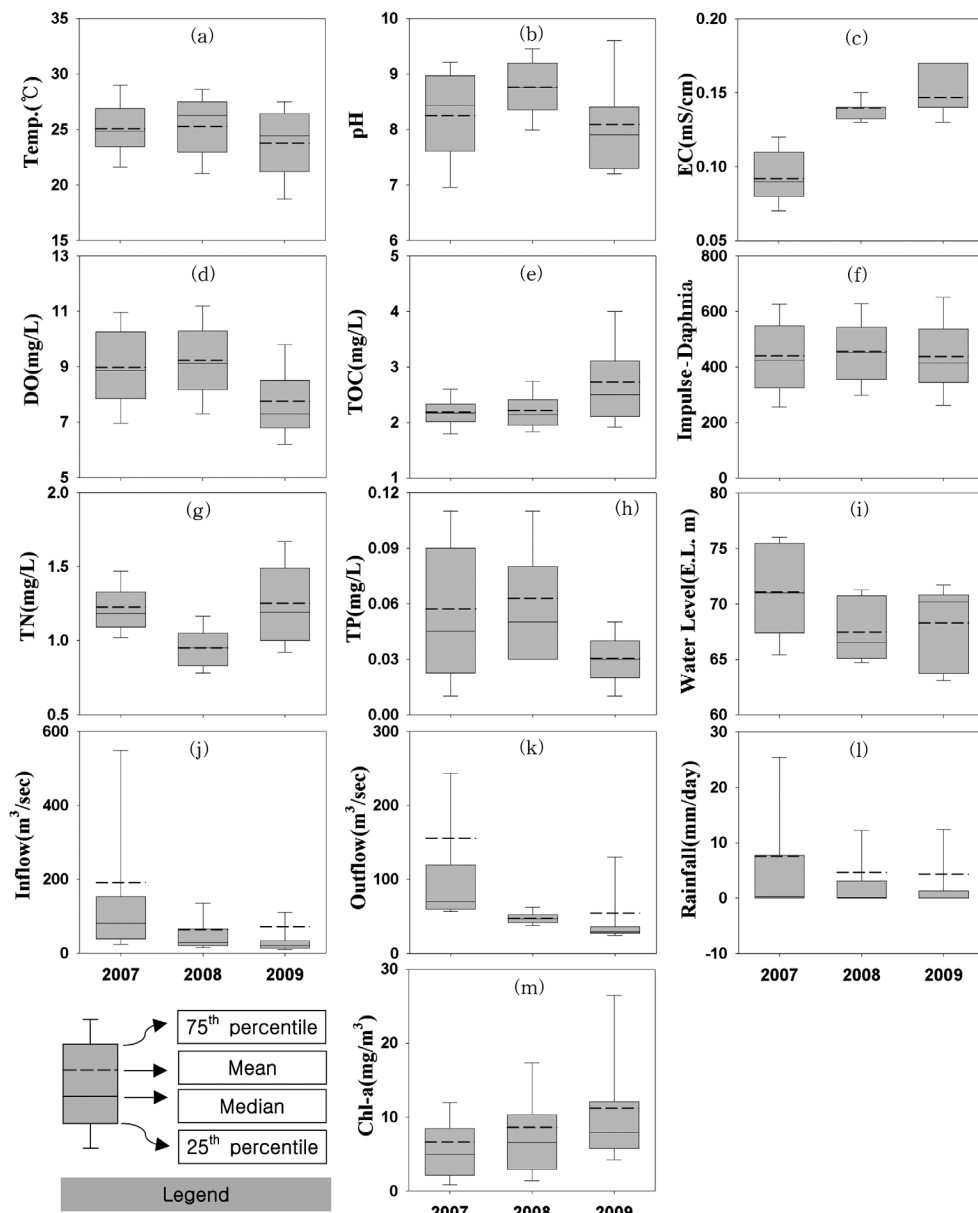


Fig. 1. Automatic water quality data characteristics for 3 years. (a) Water temperature, (b) pH, (c) Electrical conductivity, (d) Dissolved oxygen, (e) Total organic carbon, (f) Impulse, (g) Total nitrogen, (h) Total phosphorus, (i) Water level, (j) Inflow, (k) Outflow, (l) Rainfall, (m) Chl-a

들을 중심으로 입력 자료를 조합하고, Chl-a를 예측하기 위해서 다층신경망 기법을 적용하였다. 다층신경망은 k 일 전인 과거($t-k$ 시점)부터 현재(t 시점)까지의 기지값을 통해 변화 상태를 학습하고, i 일 후인 미래($t+i$ 시점)의 값을 예측하도록 구성되었다.

예측모형을 입력자료에 따라 8개의 모형으로 구분하였는데, No. 1부터 No. 5 모형까지는 입력자료로서 수온을 시작으로 DO, pH, 과거 Chl-a, 댐 유입량을 순차적으로 추가하여 구성하였으며, No. 6부터 No. 8 모형까지는 일반적으로 조류 분석에 주요 인자로 알려진 수온과 TP를 포함한 후, Impulse, 수위를 각각 추가한 모형과 모두 포함되지 않은 모형으로 구성하였다. 각 모형은 입력층 노드가 n 개 일 때, 은닉층 노드에 따라 $n, 2n, 2n+1, 3n, 4n$ 개까지 구분되었으며, 학습회수는 2000, 5000, 10000회로 충분한 학습이 고려되었다. 또한 예측 간격에 따라 1일, 3일, 7일 모형으로 나누며, 예측 간격별 값을 조합한 조합모형 1, 2를 추가하여 비교 분석 하였다. Table 1에 적용된 파라미터를 조건별로 구분하여 나타내었으며, 총 600개($8 \times 5 \times 3 \times 5$)의 케이스가 적용되었다.

7일 예측 모형은 7일 후의 Chl-a를 예측하도록 구성되었는데, 예를 들어 수온, pH, DO를 입력변수로 하여 Chl-a를 예측하는 모형의 경우에 7일전 수온[수온($t-7$)], 7일전 pH [pH($t-7$)], 7일전 DO [DO($t-7$)]를 입력자료로 사용하여 현재(오늘)의 Chl-a(t)를 계산하도록 학습되었으며, 학습된 신경

망 모형을 통해서 6일전 수온(수온($t-6$)), 6일전 pH [pH($t-6$)], 6일전 DO [DO($t-6$)]를 입력 자료로 하여 그로부터 7일 후인 내일의 Chl-a($t+1$)를 계산(예측)하였다. 3일 예측, 1일 예측 모형도 입력 자료를 달리하여 같은 과정으로 진행되었다.

예측 간격에 따라 1일, 3일, 7일로 구분된 다층신경망 모형은 입력 자료별, 은닉층 노드수 및 학습회수별 변동에 따라 평가되었다. 모형 내의 파라미터를 조정해 가면서 예측 오차를 분석하고, 가장 적합한 구조의 모형을 선정하게 된다. 적절한 구조는 예측 간격에 따라 다를 수도 있다. 1일 예측 모형의 적절한 구조가 반드시 3일, 7일에도 적용되는 것은 아니다. 보다 안정적인 값을 얻기 위해서 실험적인 방법을 활용하였는데, 각 예측 간격별로 선정된 적정 모형들의 예측값을 평균한 것이 그 방법이다. 1일 예측 모형에 의해 1일전에 예측되어진 Chl-a (Output 1)와 3일 예측 모형에 의해 3일전에 예측되어진 Chl-a (Output 3), 7일 예측 모형에 의해 7일전에 예측되어진 Chl-a (Output 7)를 모두 평균하여 조합(Fig. 2)된 값을 조합모형(combination)으로 표현하였다. 공통된 입력자료가 적용된 모형들 중, 1, 3, 7일 예측에서 각각 최소오차를 갖는 파라미터들로 구성된 3개의 모형들의 출력값을 평균하여 조합모형1로 구분하였으며, 1, 3, 7일 예측에서 상대적으로 낮은 오차를 보여 모든 경우에 적용 가능하다고 판단된 1개 모형의 1, 3, 7일 예측에 대한 각각의 출력값을 평균하여 조합모형2로 구분하였다.

Table 1. The tested parameters of the Neural networks for the Chl-a forecasting

No.	Input factor	Input nodes (ea)	Hidden nodes (ea)	Epoch (number)	Time step (day)
No. 1	W.T.	1			
No. 2	W.T., DO	2			
No. 3	W.T., DO, pH	3	n	2000	1
No. 4	W.T., DO, pH, Lagged Chl-a	4	$2n$	5000	3
No. 5	W.T., DO, pH, Lagged Chl-a, Inflow	5	$2n+1$	10000	7
No. 6	W.T., Impulse, TP	3	$3n$		1 (combination1)
No. 7	W.T., W.L., TP	3	$4n$		1 (combination2)
No. 8	W.T., TP	2			

* W.T. : water temperature, DO : dissolved oxygen, Inflow : dam inflow, Impulse : Daphnia impulse, W.L. : water level, TP : total phosphorus

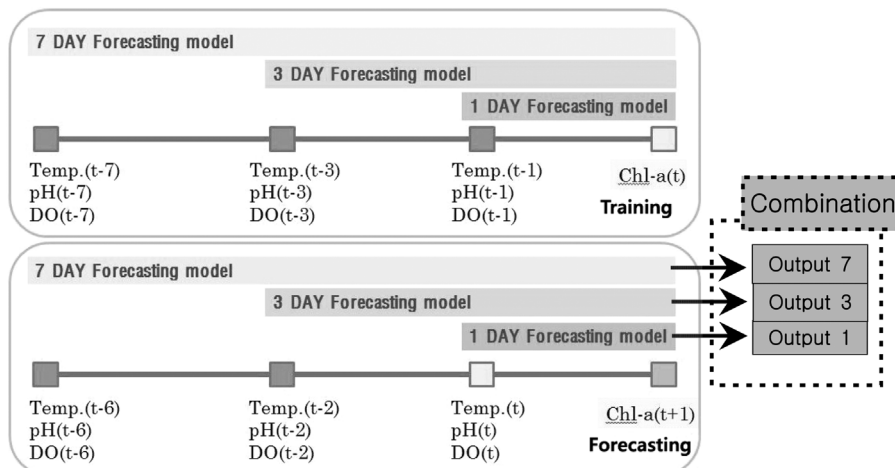


Fig. 2. Process for the Chl-a forecasting.

3. 결과 및 고찰

3.1. 상관분석

대청호의 조류는 강우의 영향과 수문조작에 큰 영향을 받는다. 강우로 인한 하천 유입량의 증가로 흐름이 발생하며, 이때 수온이 급감하게 된다. 비교적 큰 농도의 조류발생이 빈번한 여름철을 중심으로 자료에 대한 분석을 수행하였다.

대청호 측정소의 연간 Chl-a 농도변동에 영향을 미치는 수질항목들(수온, pH, DO, Impulse, TN, TP), 수리·수문인자들(댐 수위, 댐 유입량, 강우량)을 대상으로 변동성향을 비교해 본 결과, 대청호 Chl-a와는 수온, pH, DO가 시차를 두고 서로 같은 변동을 보이며, 댐 유입량, TN, TP가 유사한 변동 패턴을 보였다(Fig. 3).

교차상관분석에서는 Chl-a의 농도는 수온, pH, DO와 상

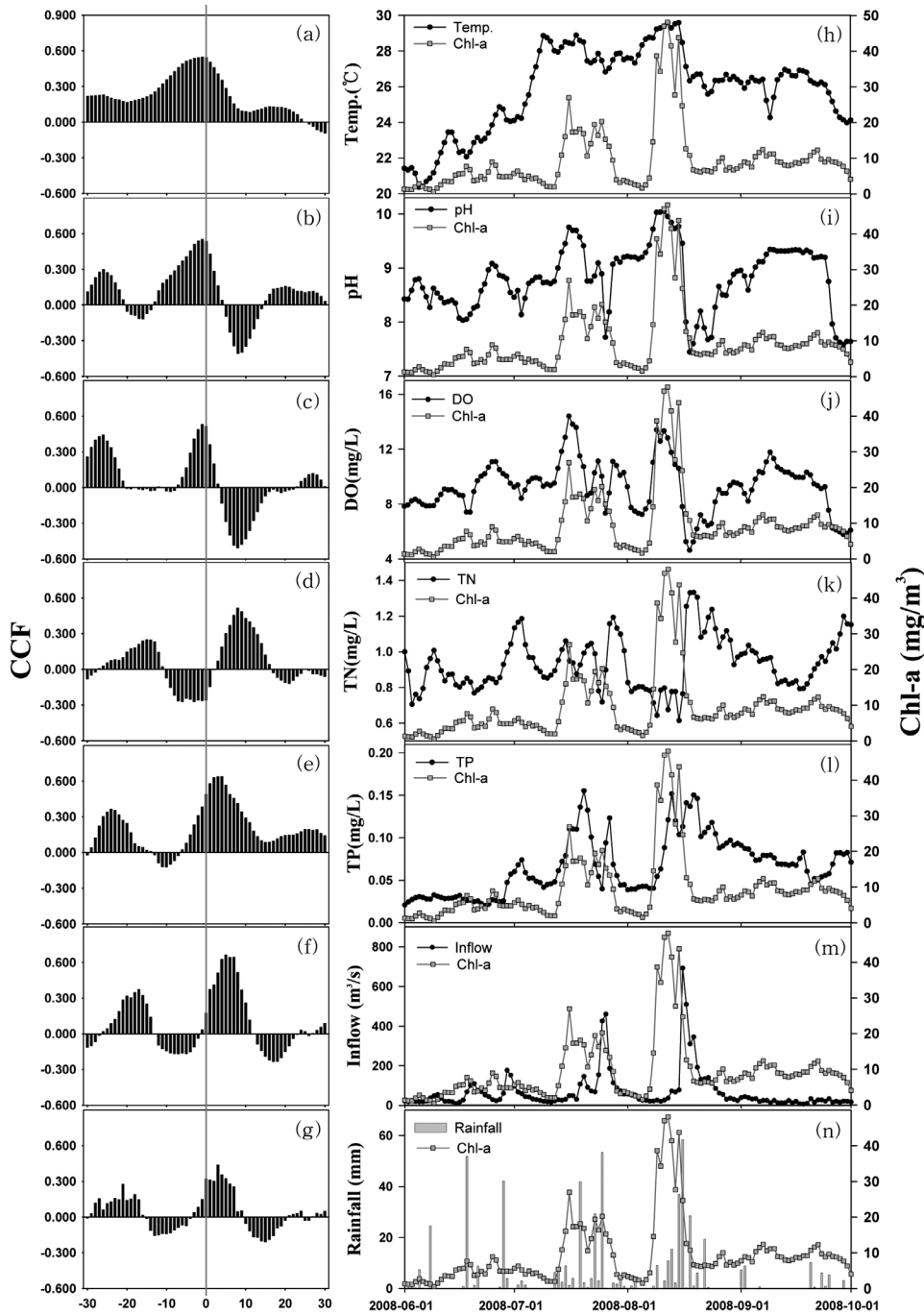


Fig. 3. The comparison between cross correlation analysis (a-g) and fluctuating trend of water quality data (h-n) in the summer. (a) Water temperature with Chl-a, (b) pH with Chl-a, (c) Dissolved oxygen with Chl-a, (d) Total nitrogen with Chl-a, (e) Total phosphorus with Chl-a, (f) Inflow with Chl-a, (g) Rainfall with Chl-a, (h) Water temperature with Chl-a, (i) pH with Chl-a, (j) Dissolved oxygen with Chl-a, (k) Total nitrogen with Chl-a, (l) Total phosphorus with Chl-a, (m) Inflow with Chl-a, (n) Rainfall with Chl-a

관성이 대체로 높은 것으로 나타났으며, 동일시점과 1일 전, 후 자료 간의 교차상관계수가 가장 큰 것으로 나타났다. 특히 DO와 pH의 최고 피크가 Chl-a의 최고농도 시기와 일치하는 것으로 나타나 조류증식에 의한 광합성의 영향이 수질에 큰 영향을 미치고 있음을 설명하고 있다.

상관분석 결과를 변동성향 그래프와 비교해보면, 여름철 강우량과 유입량은 조류발생 이전보다 이후에 뚜렷한 연관성을 보였는데, 특히 조류증식이 이미 진행된 이후에 큰 규모의 강우량과 유입량은 수온 감소 및 플러싱(flushing) 효과를 유발함으로써 조류감소 효과를 가져 오는 것으로 분석되었다.

Chl-a의 농도 예측을 위해서는 조류발생 사전에 원인을 제공하는 인자에 대한 분석이 필요하다. 조류발생의 큰 영향인자로 알려진 강우량, 유입량, 영양염류인 N, P는 이번 교차상관분석에서 조류발생 사전에 미치는 영향이 뚜렷하게 나타나지는 않았으나, 교차상관분석 결과의 시차 비교를 통해서 TP는 초기강우와 TN은 댐 유입량(Inflow)과 유사한 시기에 Chl-a에 대한 영향을 미치는 것으로 분석되었다 (Fig. 3). 국내에서 조류의 영양제한인자로 알려진 TP의 경우, 강우 초기에 농도가 증가되는 것으로 나타났으며, TP 유입 후 3주일 후에 Chl-a의 농도가 증가하는 경향을 보였다. TN의 경우는 교차상관분석에서 TP와 같은 경향을 보이나 상관성이 작게 나타났으며, 변동성향 그래프에서 Chl-a의 농도가 가장 높았던 8월에 TN 농도가 상대적으로 낮은 것을 보아 TN이 제한인자가 아닌 것으로 추정될 수 있으나, 향후 자동측정 자료의 N/P 비율에 대한 영향 분석

은 검토되어야 할 것으로 판단된다.

3.2. 신경망 모형의 적용 및 파라미터 평가

조류농도의 예측을 위해서는 연관성이 큰 자료가 효과적이며, 사전 분석을 통해 1차적인 검토를 할 수 있다. 입력 자료로서 TN, 강우량 등 상관성이 크지 않았던 항목은 제외하고, 입력 자료의 구성에 따라 모형을 8개(No. 1~No. 8)까지 구분하여 적용하였으며, 검증 자료를 통해서 예측 간격별, 입력 자료별 적정 모형에 대한 결과를 정리하여 Table 2에 나타내었다. 각 신경망 모형은 은닉층 노드수와 학습회수를 변경하며 모의되었으며, 다만, 수온만을 입력 자료로 사용하였을 때는 수렴되지 않는 경우가 있어 몇 가지 추가적인 테스트를 수행하였다.

각 모형별 파라미터인 은닉층 노드와 학습회수는 예측 간격별, 입력 자료별로 다양한 결과를 보였는데, 원인은 초기 연결강도가 랜덤(random)하게 세팅되어 있기 때문인 것으로 판단된다. 초기 연결강도가 운영자에 의해 고정된 값으로 초기화되어 있을 경우에 학습회수의 크기가 중요한 파라미터로서 의미를 가질 수 있겠으나, 난수 발생에 의해 랜덤하게 처리되는 경우에는 국부 최소점에 빠지지 않을 가능성이 많은 대신, 수렴하는 속도가 모두 달라서 학습회수의 크기는 큰 의미를 부여받지 못한다. 다만, 입력층, 은닉층, 출력층의 노드수와 입력 자료 등 다른 파라미터들이 동일한 조건에서 어느 정도의 충분한 학습이 이루어진다면 결과는 유사할 것으로 기대될 수는 있으나, 실제 자

Table 2. The optimal structure by estimation of least mean errors

No.	Input factor	Input nodes - Hidden nodes - Output node (Training epoch)				
		1day ①	3day ②	7day ③	Combination1 (①+②+③)/3	Combination2 Optimal
No. 1	W.T.	1-3-1 (10000)	1-2-1 (433)	1-9-1 (2000)	1-3-1 (10000) 1-2-1 (433) 1-9-1 (2000)	1-6-1 (2000)
No. 2	W.T., DO	2-5-1 (2000)	2-5-1 (2000)	2-2-1 (2000)	2-5-1 (2000) 2-5-1 (2000) 2-2-1 (2000)	2-5-1 (2000)
No. 3	W.T., DO, pH	3-6-1 (2000)	3-3-1 (5000)	3-3-1 (5000)	3-6-1 (2000) 3-3-1 (5000) 3-3-1 (5000)	3-3-1 (5000)
No. 4	W.T., DO, pH, Lagged Chl-a	4-4-1 (10000)	4-4-1 (5000)	4-4-1 (5000)	4-4-1 (10000) 4-4-1 (5000) 4-4-1 (5000)	4-4-1 (5000)
No. 5	W.T., DO, pH, Lagged Chl-a, Inflow	5-5-1 (5000)	5-5-1 (2000)	5-5-1 (5000)	5-5-1 (5000) 5-5-1 (2000) 5-5-1 (5000)	5-5-1 (5000)
No. 6	W.T., Impulse, TP	3-3-1 (10000)	3-3-1 (5000)	3-12-1 (5000)	3-3-1 (10000) 3-3-1 (5000) 3-12-1 (5000)	3-3-1 (10000)
No. 7	W.T., W.L., TP	3-7-1 (5000)	3-3-1 (10000)	3-7-1 (2000)	3-7-1 (5000) 3-3-1 (10000) 3-7-1 (2000)	3-3-1 (10000)
No. 8	W.T., TP	2-2-1 (10000)	2-2-1 (2000)	2-2-1 (5000)	2-2-1 (10000) 2-2-1 (2000) 2-2-1 (5000)	2-2-1 (10000)

료들은 다양한 원인으로 인해 비선형성이 강하여 수렴이 잘 되지 않은 경우가 많다. 따라서 랜덤한 초기 연결강도를 사용하여 현장에 적용할 경우에는 학습과정, 검증과정, 테스트 과정을 거쳐 모형을 최종 판단하는 것이 적절할 것이다.

이번에 대청호 Chl-a의 1일, 3일, 7일 예측에 적용된 모형들 중, 적절한 결과를 보인 모형들은 뚜렷한 패턴을 보이지는 않았으나, 대체로 은닉층 노드수와 입력층 노드수가 같은 경우에 많았다. 1, 3, 7일 예측에 모두 최상의 결과는 아니더라도 전체적으로 적절한 결과를 보임으로써 선정된 모형들(조합모형2, Combination2)을 중심으로 분석해보면, 은닉층 노드수가 많을수록, 그리고 상관성이 높은 입력 인자들을 사용하였을 경우에 수렴을 위한 학습회수가 비교적 적었다(Table 2).

Chl-a와의 상관성이 높게 나타난 인자들을 중심으로 입력 인자를 변경해 가며 모형을 구성하여 적용한 결과, 대체로 예측 간격이 커질수록 예측오차(Root mean square error: RMSE)가 증가되는 것으로 나타났다. 예측모형 중 2008년 자료를 통한 검증과정에서 양호한 결과를 보인 경우는 수온과 함께 조류발생의 영향으로 변동되는 DO와 pH를 입력인자로 구성한 모형(No. 3)이다. 이 구성의 Chl-a 농도에 대한 예측오차는 1일 예측모형이 5.89 mg/m³, 3일 예측모형이 6.17 mg/m³, 7일 예측모형이 6.87 mg/m³로 나타났다. 또한 각 모형들로부터 사전에 예측되어진 Chl-a 농도를 산술평균한 예측오차가 4.20 mg/m³로 나타나, 예측 모형들의 조합이 효과적인 것으로 판단된다(Table 3).

입력 자료로 수온만을 사용했을 때(No. 1)에는 1일 예측 결과가 비교적 큰 오차를 보였다. 수리·수문 인자(수온, 댐 수위)와 영양염류(TP)가 포함된 모형(No. 7)의 1일 예측오차는 8.94 mg/m³, 3일 예측오차는 8.56 mg/m³, 7일 예측오차는 8.50 mg/m³로 나타났고, 모형을 조합한 경우의 예측오차가 8.30 mg/m³로 계산된 오차는 모두 비교적 큰 것으로

분석되었다. 물벼룩의 Impulse 를 사용한 모형(No. 6) 또한 예측오차가 큰 것으로 나타났는데 연속 측정된 Impulse는 물벼룩 사멸 등의 이유로 Chl-a와 상관성이 낮은 것으로 조사된 바(연인성 등, 2010) 있다. 그러나 수질악화로 인한 물벼룩의 사멸 직전에 Impulse 가 급증하므로, 수체의 급변을 통해서 Chl-a의 시기를 분석하는 데는 유용하게 이용될 수 있다. 이번 연구에서와 같이 연속측정 자료를 통한 농도변화 예측을 위해서는 입력 자료간의 독립성 문제나 입·출력간 인과관계에 있는 다양한 항목을 고려하기보다 예측력 향상에 도움을 주는 항목만을 사용하는 것이 효과적인 것으로 판단된다.

검증과정에서 선정된 모형들의 파라미터를 이용하여 학습과 검증에 사용되지 않은 2009년 자료를 대상으로 예측 성능을 테스트하였다. 2009년 테스트 자료를 통해서도 조합모형이 상대적으로 좋은 결과를 보였는데, 예측 간격별로 최소오차를 보인 경우의 입력 자료들은 상이하였다. 1일 예측에서는 입력자료로 수온, DO, pH, 과거 Chl-a를 사용한 경우가 6.27 mg/m³, 3일 예측에서는 수온, DO, pH를 사용한 경우가 5.89 mg/m³, 7일 예측에서는 수온, DO, pH, 과거 Chl-a, 유입량을 사용한 경우가 6.06 mg/m³로 비교적 작은 오차를 보였으며, 조합된 결과에서는 수온, DO, pH, 과거 Chl-a의 경우가 4.36 mg/m³로 가장 작은 오차를 보였다. 이상 3개의 입력 자료를 사용한 경우(No. 3, No. 4, No. 5)에 2009년 검증결과에서도 모형을 조합한 결과들 중 상대적으로 작은 오차를 나타내었다. 또한 조합된 모형이 2008년 검증결과와 유사하게 비교적 작은 오차를 보였는데, 수온, DO, pH를 입력 자료로 사용하였을 때(No. 3), 예측오차는 각각 1일 예측모형으로부터 7.75 mg/m³, 3일 예측모형으로부터 5.89 mg/m³, 7일 예측모형으로부터 8.79 mg/m³, 각 예측모형 조합으로부터 5.79 mg/m³ 만큼 산정되었다(Table 3).

2008년과 2009년에 공통적으로 3일 예측결과가 1일 예측

Table 3. The forecasting errors classified by input data and time steps

No.	Input factor	RMSE (mg/m ³)				
		1day	3day	7day	Combination1	Combination2
Verifi- -cation (2008)	No. 1 W.T.	9.67	7.54	7.54	7.97	8.41
	No. 2 W.T., DO	7.17	7.22	7.89	5.98	6.33
	No. 3 W.T., DO, pH	5.89	6.17	6.87	4.49	4.20
	No. 4 W.T., DO, pH, Lagged Chl-a	6.13	7.15	10.39	5.07	5.85
	No. 5 W.T., DO, pH, Lagged Chl-a, inflow	6.21	7.55	8.60	5.35	5.48
	No. 6 W.T., Impulse, TP	9.55	8.55	9.74	7.24	8.44
	No. 7 W.T., W.L., TP	8.94	8.56	8.50	6.59	8.30
	No. 8 W.T., TP	7.83	7.75	6.97	6.51	6.10
Test (2009)	No. 1 W.T.	8.64	10.31	8.90	8.09	8.36
	No. 2 W.T., DO	8.72	8.83	10.51	7.38	8.15
	No. 3 W.T., DO, pH	7.75	5.89	8.79	5.94	5.79
	No. 4 W.T., DO, pH, Lagged Chl-a	6.27	8.15	7.71	4.13	4.36
	No. 5 W.T., DO, pH, Lagged Chl-a, inflow	6.44	11.30	6.06	5.20	4.49
	No. 6 W.T., Impulse, TP	9.68	8.68	13.90	7.44	8.11
	No. 7 W.T., W.L., TP	9.75	10.04	10.99	9.78	9.29
	No. 8 W.T., TP	8.90	9.46	7.86	7.32	7.23

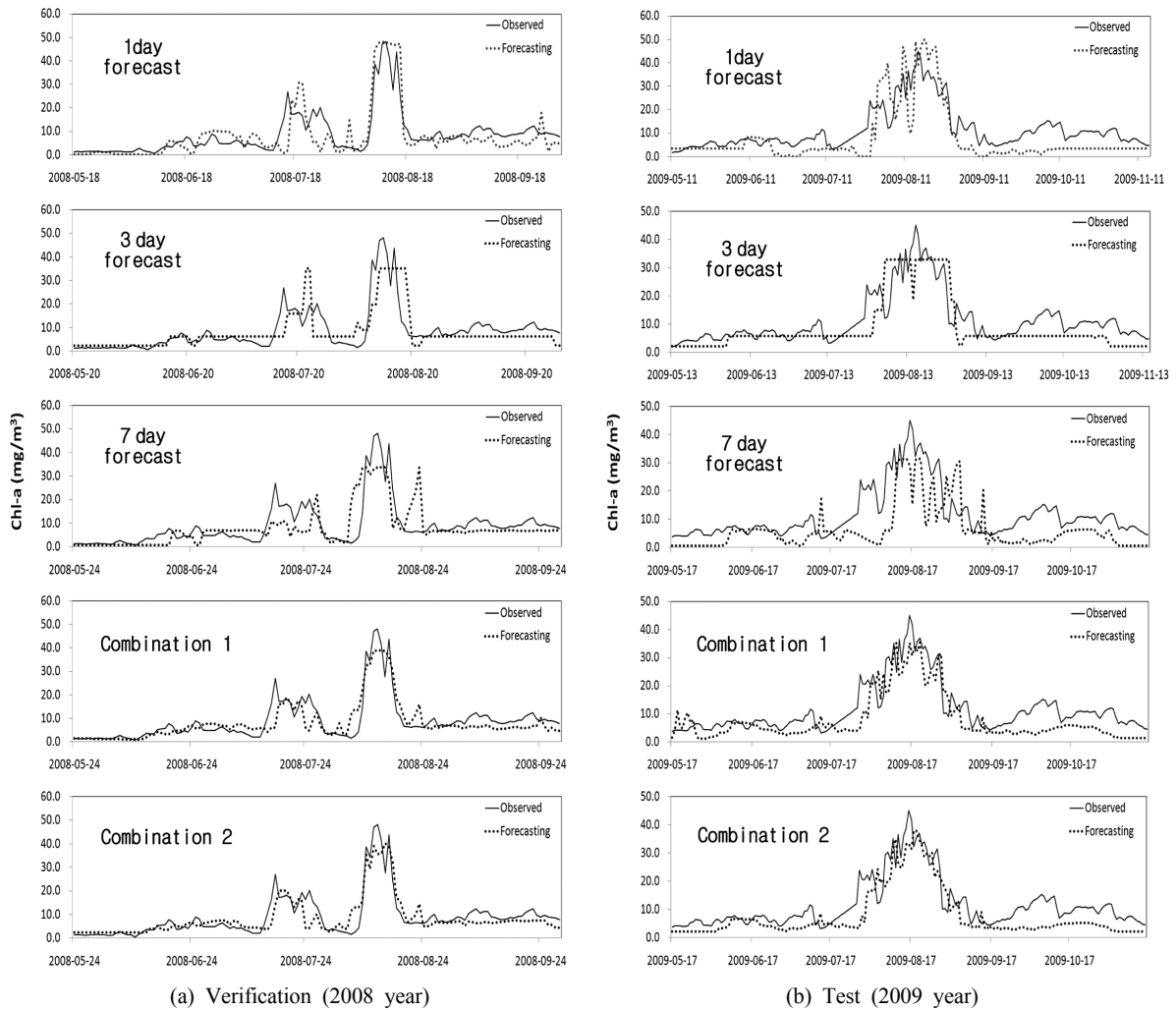


Fig. 4. The forecasting results of No. 3 model.

결과보다 오차가 작았던 경우는 수온, Impulse, TP를 사용했던 경우(No. 6)이며, 7일 예측 오차가 가장 작았던 경우는 수온과, TP를 사용했던 경우(No. 8)이었다. 수온과 Impulse, TP는 조류발생 시기를 사전에 분석하는데 효과적으로 이용될 수 있을 것으로 판단된다.

입력 자료로 수온, pH, DO를 사용한 No. 3 모형의 예측 결과를 Fig. 4에 도시하였다.

4. 결론

대청호의 조류발생을 조기에 예측하기 위하여 수질자동측정망의 Chl-a와 수질 및 수문인자와의 상관분석을 수행하고, 상관성이 높은 수질항목을 중심으로 신경망 기법을 적용하여 입력 자료별, 예측 간격별, 은닉층 노드수별, 학습 회수별 분석을 통해 다음과 같은 결론을 얻었다.

1) 대청호 측정소의 Chl-a와 상관성이 큰 인자들 중 수온은 일정한 시차를 두고 밀접한 변동을 보였고, pH와 DO는 Chl-a와 최고 농도 시기가 일치하여 조류증식에 의한 광합성이 수질에 영향을 미치는 것으로 분석되었다. 조류발생 이후에 나타난 여름철 강우량과 유입량은

수온 감소 및 플러싱(flushing)으로 Chl-a 농도의 감소를 가져오는 뚜렷한 양상을 보였으며, TP는 초기강우와 TN은 유입량과 유사한 시기에 Chl-a에 대해 영향을 미치는 것으로 분석되었다.

- 2) 검증과 테스트 과정에서 오차 분석을 통해 적용 모형들의 성능을 분석한 결과, Chl-a와 상관성이 높았던 수온, DO, pH를 입력인자로 사용한 모형(No. 3)이 예측오차가 작았으며, 과거의 Chl-a를 No. 3에 추가한 모형(No. 4), 또 댐 유입량을 No. 4에 추가한 모형(No. 5)이 상대적으로 좋은 결과를 보여, 신경망 모형에 사용되는 입력 자료로는 독립성에 상관없이 종속변수와 상관성 큰 인자들이 효과적일 것으로 판단되었다.
- 3) 대체로 예측 간격이 커질수록 예측 오차가 증가되는 것으로 나타났으나, 수온, Impulse, TP를 입력 자료로 사용했던 경우(No. 6)에는 3일 예측오차가 1일 예측오차보다 작았고, 수온과 TP를 사용했던 경우(No. 8)에는 7일 예측오차가 상대적으로 가장 낮았다. 예측오차를 낮추기 위해서는 입력 자료를 선정할 때에 교차상관분석을 통한 Chl-a와 입력 자료간의 시차가 고려되어야 함을 시사한다.

4) 신경망을 이용한 대청호 조류예측을 위해 1일, 3일, 7일 예측모형이 적용된 이번 연구에서 각 모형의 예측 오차보다 3개의 모형의 예측값을 산술평균한 오차가 더 작은 것으로 분석되었으며, 안정적인 결과를 보였다. 또한 대체로 작은 오차를 보인 모형들은 은닉층 노드수가 입력층 노드수와 같은 경우가 많았으며, 상관성 높은 입력 자료를 사용하였을 때 혹은 은닉층 노드수가 많을수록 학습회수가 적어 수렴이 빠른 것으로 나타났다.

추가적으로 장기 예측을 통한 예측 가능범위나 타 유역에서의 검증이 필요하나, 이상에서와 같이 대청호내의 자동 측정자료를 이용하여 Chl-a의 일변화를 예측할 때에는 자료 간에 상관성이 큰 인자만을 이용하여 복잡하지 않은 신경망 모형을 구성하고, 여러 모형의 결과를 조합하여 안정화된 값으로 최종 산출되는 모듈라형(modular type) 기법을 고려한다면 안정적인 조류 사전예측 시스템을 구축할 수 있을 것으로 판단된다.

사 사

본 연구는 2010년도 국립환경과학원 박사후연수과정 지원사업에 의해 이루어진 것임.

참고문헌

국립환경과학원(2008). 대청호 조류상승 발생수역이 분류 수질에 미치는 영향 분석(II).

국립환경과학원(2009). 대청호 수질 및 조류발생 실태조사.

김동균(2008). Plankton community dynamics in Nakdong river basin: the prediction based on ecological modeling. 박사학위논문, 부산대학교.

박혜경, 김은경, 정동일(2005). 팔당호 조류 발생 예측 모델. 공동출판회 논문집, 한국물환경학회·대한상하수도학회, pp. 115-118.

서동일(1997). 댐 유역 오염물질 유입특성 및 영향에 관한 연구(2차년도). 한국수자원공사.

심순보, 고덕구, 이요상, 오이성(1995). WASP5 수질모형을 이용한 대청호의 부영양화 예측. 건설기술논문집, 14(2), pp. 41-53.

연인성, 홍지영, 홍은영, 임병진(2010). 연속 측정된 대청호 Chlorophyll-a의 자료 특성 및 상관 분석. 수질보전 한국물환경학회지, 26(6), pp. 994-999.

유병로(2001). WASP5 수질모형을 이용한 대청호 부영양화 인자의 특성연구. 한국환경기술학회지, 2(3), pp. 297-306.

유순주, 황종연, 채민희, 김상용(2006). 대청호 유속에 따른 조류이동 영향. 수질보전 한국물환경학회지, 22(5), pp.

887-894.

정건희(2001). 신경망 모형을 이용한 대청댐의 수질예측에 관한 연구. 석사학위논문, 고려대학교.

정광석(2004). Application of machine learning to pattern and predict the phytoplankton time-series data in a flow-regulated river system (the lower Nakdong river). 박사학위논문, 부산대학교.

정세용, 박재호, 김유경, 윤성환(2007). 대청호 부영양화 모의를 위한 CE-QUAL-W2 모형의 적용. 수질보전 한국물환경학회지, 23(1), pp. 52-63.

정종혁, 정하규, 권원태(2008). 조류를 이용한 수계모니터링 시스템에서 뉴럴 네트워크에 의한 실시간 독성물질 판단. 수질보전 한국물환경학회지, 24(1), pp. 1-6.

Chen, S. H., Jakeman, A. J., and Norton, J. P. (2008). Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, 78(2), pp. 379-400.

Hornsby shire council. (2008). *Estuary management program (2007-2008 annual report)*. Australia.

Karul, C., Soyupak, S., Cilesiz, A. F., Akbay, N., and Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling*, 134(2), pp. 145-152.

Lee, J. H. W., Huang, Y., Dickman, M., and Jayawardena, A. W. (2003). Neural network modelling of coastal algal blooms. *Ecological Modelling*, 159(2), pp. 179-201.

Maier, H. R., Dandy, G. C., and Burch, M. D. (1998). Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray. *Ecological Modelling*, 105(2), pp. 257-272.

Oh, H. M., Ahn, C. Y., Lee, J. W., Chon, T. S., and Choi, K. H. (2007). Community patterning and identification of predominant factors in algal bloom in Daechung Reservoir (Korea) using artificial neural networks. *Ecological Modelling*, 203(1), pp. 109-118.

Palani, S., Liong, S. Y., and Tkalich, P. (2008). An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), pp. 1586-1597.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K. I. (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, 96(2), pp. 11-28.

Velo-Suarez, L. and Gutierrez-Estrada, J. (2007). Artificial neural network approaches to one-step weekly prediction of *Dinophysis acuminata* blooms in Huelva. *Harmful Algae*, 6(3), pp. 361-371.

WAMIS (2007-2009). <http://www.wamis.go.kr/>.