

# 신경망모형을 이용한 하천의 수질예측 연구

○김만식\* 이요상\*\* 심규철\*\*\* 심순보\*\*\*\*

## 1. 서론

하천은 인류문명이 시작되면서 생활터전으로서 우리의 생활환경에 중요한 부분을 차지하여 왔고 자연적인 것으로 조건없이 사용되어 왔다. 그러나, 우리 나라는 1970년대 이후 경제발전에 힘입어 국민 생활수준의 급격한 향상으로 물 사용량이 급격히 증가되었고, 이에 따라서 수질오염도 심각한 수준으로 되어 80년대 이후에는 수질오염의 방지 및 저감, 오·폐수 처리시설의 확대와 첨단화를 도모하여 환경 및 생태계 변화에 적극 대응하고 있는 실정이다. 수질예측을 위한 모형으로는 개념적이고 물리적인 모형과 시스템 또는 Box모형으로 구분할 수 있다. 물리적모형은 하천 수질과 관련되는 자연현상을 제약조건으로 수식화하고 실험을 통한 경험적인 관계를 기초로 하여 모형을 구성하며, Block-Box모형은 시스템내의 물리적 특성을 구체화하지 않고 입력-출력간의 관계만을 나타낼 수 있는 모형이다. 기존에 사용되어온 개념적모형은 WQRRS, WASP5, QUAL2E 등이 있지만, 활용시 많은 입력자료가 요구되고 최적의 매개변수를 산정하기가 어렵다.

신경망이론은 인간의 두뇌와 같이 생각하고 판단하며, 인식할 수 있는 능력을 구현하기 위해 인간뇌의 정보처리 메커니즘을 수학적으로 표현한 이론으로서 다수의 원소가 동시에 작동하는 병렬 처리 기계처럼 하드웨어 및 소프트웨어적인 병렬 분산 처리를 하는 시스템이며, 다른 환경에 맞도록 변화시킬 수 있는 학습 능력을 가지고 있다. 또한, 모형 개발에 있어 특별한 구조나 매개변수 산정, 자료의 변환 등이 필요치 않고 자료의 축척에 따라 학습을 통하여 모형의 예측 능력을 향상시킬 수 있다. 따라서 역전파알고리즘과 최적 매개변수 추정에 사용되는 모멘트-적용학습을방법을 이용하여 학습회수 및 은닉층 그리고 노드수에 변화를 주어 여러 다층신경망 모형을 구축하여 점진강유역의 쌓치와 관촌지점에 적용하였다. 그리고 학습된 결과중에 효율이 좋은 모형을 선정하여 학습되지 않은 자료의 수질인자를 예측을 실시하였다. 또한, 예측결과를 분석·고찰하였으며, 하천의 단기간 수질예측에 신경망모형의 실용가능성을 검증하였다.

## 2. 신경망이론

### 2.1 다층신경망모형

다층신경망모형은 입력자료를 받아들이는 입력층(input layer), 결과를 나타내는 출력층(output layer), 그리고 두 층 사이에 은닉층(hidden layer)으로 구성되어 있으며, 이 층들을 결합하는 점을 시냅스(synapse)라 하고 이 시냅스의 연결강도(결합강도) 변화에 따라 입력에 따른 출력이 결정되

\* 정회원, 신성대학 토목과, 교수

\*\* 정회원, 한국수자원공사 수자원연구소, 선임연구원

\*\*\* 정회원, 충북대학교 건설기술연구소, 선임연구원

\*\*\*\* 정회원, 충북대학교 토목공학과, 교수

어진다. 다층신경망모형에서 시냅스의 연결강도는 모형에 따라 미리 계산에 의해 주어지는 경우와 학습에 의해 결정되는 경우가 있으며, 본 연구에서의 연결강도는 임의의 초기값에서 지속적인 개선이 이루어지며 적절한 연결강도를 구하는 학습(learning)과정을 사용하였다. 학습방식에 따라 지도학습(supervised learning)모형과 자율학습(unsupervised learning)모형으로 나눌 수 있으며, 입력자료와 출력자료가 2진수(binary)의 형태를 갖거나 실수형태의 연속적인 값을 가지는 두 가지 형태로 나누어진다. 본 연구에 적용된 다층신경망(multilayered neural network)모형은 1개 이상의 은닉층이 존재하는 모형으로, 입력층에서 은닉층으로 또는 은닉층에서 출력층으로 연결될 때 연결강도의 백터들과 곱해져서 특정한 출력함수를 거쳐 출력을 산정한다. 이때 사용하는 출력함수(또는 전이함수)는 대표적인 비선형 함수이며 로그-시그모이드(log-sigmoid)함수를 사용하였다.

## 2.2 역전파알고리즘

다층구조의 신경망에서 각 뉴런과 각 층(layer)의 모든 뉴런에 임계값(threshold values) 사이에 연결 가중치를 주기 위해서 학습 알고리즘(learning algorithm)이 필요하다. 본 연구에서는 학습 알고리즘으로 러멜하트(rumelhart, 1986) 등에 의해 제안된 역전파 알고리즘(backpropagation algorithm)을 사용하였다. 역전파 알고리즘은 입력층의 각 유니트에 입력패턴이 주어지면 이 신호는 각 유니트에서 가중값에 의해 변환되어 은닉층에 연결되고 마지막에 출력층으로 전파해 나가는데, 이때 네트워크에서 출력된 신호 패턴이 목표 패턴과 일치하지 않을 경우 역방향인 출력 패턴에서 목표 패턴으로 오차의 값들을 감소시킬 수 있도록 네트워크의 연결강도를 조절하여 학습하는 방법이다.

## 2.3 모멘트(moment)법과 적응학습율(adaptive learning rate)

다층신경망모형에서 역전파 알고리즘에 의한 학습과정에서 발생할 수 있는 지역최소값 문제와 오차를 적게 하고, 수렴속도를 빨리하여 학습속도를 향상시킬 수 있는 모멘트법과 적응학습율을 사용하였다. 모멘트법은 간단하면서도 효율적이기 때문에 자주 이용되는 방법으로 모멘트항을 학습에 포함시킨다. 적응학습율은 모멘트 방법과 병행하여 학습을  $\mu$ 를 조절하며 학습속도를 향상시킨다. 학습 후에 오차함수가 감소하였나 확인을 하여 크게 감소하였을 때에는 학습율을 일정 비율만큼 증가시키고, 반대로 오차함수가 정해진 비율에 이상으로 증가하였을 때에는 학습율을 정해진 비율만큼 감소시키게 된다.

## 3. 하천의 수질 예측모형

신경망모형은 수집된 또는 측정된 입력자료와 출력자료의 학습패턴을 정리해야 하는데 자료의 특성을 변화시키지 않는 범위에서 입·출력 학습패턴을 정규화하여 입력값과 출력값의 차원을 일치시킨다. 다음 단계로 입·출력패턴으로 학습을 수행하여 최적 매개변수를 산정하기 위해 신경망모형의 구조를 입·출력자료의 수와 각 층의 수, 각 층에 대한 노드의 수, 각 층의 출력함수 등에 따라 결정하여 모형을 구축하고, 학습을 완료한 후 학습되지 않은 실측자료를 이용하여 모형의 적합성을 검증하게 된다.

### 3.1 하천의 수질 모형화

다목적댐 유입량의 수질은 발생시간에 따라 일치하지 않는 비정규성 및 비선형성을 나타낸다. 신경망은 수질인자와 유량과의 상관관계에 의하여 입력되는 변수와의 연결강도를 조정하여 활성

화값을 가감시켜 매개변수를 추정하므로 학습을 통해 시스템 동정과정을 자동화할 수 있다. 수질 예측의 모형화를 위해 시간흐름에 따른 비선형 수질인자 및 유하량을 이용한 수질모형은 식(1)과 같이 정의할 수 있다.

$$QUAL(t) = f[QUAL(t-k_i), Q(t-k_j)] \text{ for } k_i=0, \dots, n, k_j=0, \dots, m \quad (1)$$

여기서,  $QUAL(t)$  : 수질예측인자,  $QUAL(t-k_i)$  :  $k_i$ 전의 수질예측 인자,  $Q(t-k_j)$  :  $k_j$ 전의 유량,  $k_i, k_j$  :  $n, m$ 간격의 시계열(시간, 일, 달, 년)을 의미한다.

### 3.2 신경망모형 구성

예측되는 수질인자는 학습과정을 통하여 추정된 연결강도와 임계값을 이용하여 계산된다. 연결강도는 각 층의 상태를 매개변수로 저장시킨 값으로 각 처리소자간의 입력과 출력의 상관관계를 나타내며 연결되어 있는 처리소자간의 영향력에 정도를 나타낸다. 따라서 수질인자의 예측은 추정된 연결강도의 행렬을 이용하여 계산된다. 시간의 흐름에 따른 비선형 수질인자-유량의 관련성을 나타낸 식(1)를 이용하여 모형화 과정을 통해 신경망에 의한 예측모형을 정의하면 식(2)와 같이 나타낼 수 있다.

$$QUAL_i(t) = g \left( \sum_{k=1}^K W_{jk} g \left( \sum_{j=1}^I W_{ij} (QQ(t-k) + \theta_j) + \theta_i \right) \right) \quad (2)$$

여기서,  $QUAL_i(t)$  : 예측된 수질인자,  $QQ(t-k)$  : 학습된 시계열로서 과거의 수질인자와 유량,  $k$ 는 지체시간이다. 표 1에 수질예측을 위한 신경망모형들의 구성을 나타내었다.

표 1. 수질예측을 위한 신경망모형의 구성

water quality index	model	lag time	inflow lag time	input layer	hidden layer node	output layer	learning number
BOD	BOD-1, BOD-2 BOD-3, BOD-4	3	3	6	6, 12 18, 24	1	10,000 ~ 25,000
COD	COD-1, COD-2 COD-3, COD-4	3	3	6	6, 12 18, 24	1	
SS	SS-1, SS-2 SS-3, SS-4	3	3	6	6, 12 18, 24	1	

본 연구에서 예측하는 수질인자는 BOD, COD 및 SS농도로 하였으며 수질예측을 위한 최적 다층신경망모형을 구성하기 위해 표 1과 같이 입력층, 은닉층, 출력층의 노드수와 학습회수 그리고 지체시간을 정하였다.

## 4. 적용 및 결과

### 4.1 대상유역과 검증자료

본 연구를 위한 대상유역은 섬진강댐 저수지 상류유역으로 섬진강은 전라북도 장수, 임실, 진안 등 3개군의 경계를 이루는 팔공산에서 발원하여 약 84.2km를 남서류하여 섬진강 다목적댐 저수지에 이르게 된다. 섬진강 다목적댐까지의 하천현황은 섬진강 본류와 39개의 제 1, 2 지류들로 구성되어 있으며, 유역면적이 약 763.82km<sup>2</sup>로 관촌유역이 301.776km<sup>2</sup>, 댐저수지유역이 344.145km<sup>2</sup>, 쌍치유역이 117.895km<sup>2</sup>이며, 하천 총연장은 356.5km이다. 하천 수질예측모형을 개발하기 위해 섬진강 다목적댐 저수지의 상류유역인 관촌지점과 쌍치지점의 1998년(4회)과 1999년(6회)의 BOD, COD, SS 수질자료와 유출량자료를 사용하였다. 그리고 모형의 검증을 위해 학습되지 않은 자료를 이용하여 수질예측을 실시하였다.

#### 4.2 모형검증을 위한 통계지표

예측결과가 실측자료와 비교하여 얼마나 적합한 것인가 하는 문제는 단기 수질모형에 대한 신경망이론의 적용타당성과 관련된다. 따라서 본 연구에서는 예측문제에 적용되고 있는 통계적 검증 방법으로 각 모형에서 계산된 오차와 비교 검토하여 예측에 사용될 최적 모형을 선정하기 위해 SSE(sum square error), CC(correlation coefficient), RMSE(root mean square error),  $R^2$ 를 이용하여 통계분석을 실시하였다.

#### 4.3 적용 및 분석

단기 수질예측모형을 구축하기 위해 1998년부터 1999까지 측정된 관촌과 쌍치지점의 수질자료 중 9회의 자료에 대해 자기상관과 교차상관분석을 실시하여 유량은 2시간전, 수질자료는 3시간전까지 영향을 받는 것으로 분석되었다. 그리고 각 수질인자별로 12가지의 신경망모형을 구축하여 학습을 실시하였으며 통계분석 결과는 표 2와 같다. 단기 수질예측을 위한 다층신경망 모형의 학습과정에서 오차는 모든 모형에서 학습회수가 100회 이내에서 급격히 감소하였으며, 은닉층 노드 수가 많을수록 좋은 결과를 나타내었다. 표 2에 나타난 결과를 볼 때, 거의 모든 모형들이 상관계수가 0.9이상이고, RMSE는 0.1이하, SSE는 0.001 이하로 나타나 매우 양호한 결과를 보여주고 있다. 이 모형들의 예측능력을 검토하기 위해서 학습에 사용하지 않은 1999년의 1회 수질자료에 대해 예측을 실시하였다.

표 2. 신경망모형의 통계분석 결과

지점	model	SSE	CC	RMSE	model	SSE	CC	RMSE	model	SSE	CC	RMSE
관촌	BOD-1	0.00249	0.979	0.0499	COD-1	0.00201	0.976	0.0448	SS-1	0.00089	0.988	0.0299
	BOD-2	0.00228	0.981	0.0478	COD-2	0.00214	0.979	0.0462	SS-2	0.00072	0.990	0.0268
	BOD-3	0.00226	0.979	0.0475	COD-3	0.00240	0.981	0.0490	SS-3	0.00090	0.989	0.0300
	BOD-4	0.00218	0.980	0.0467	COD-4	0.00307	0.983	0.0554	SS-4	0.00066	0.992	0.0257
쌍치	BOD-1	0.00311	0.965	0.0557	COD-1	0.00200	0.984	0.0447	SS-1	0.00172	0.974	0.0415
	BOD-2	0.00199	0.969	0.0446	COD-2	0.00060	0.985	0.0246	SS-2	0.00065	0.976	0.0278
	BOD-3	0.00134	0.970	0.0367	COD-3	0.00061	0.985	0.0247	SS-3	0.00077	0.978	0.0278
	BOD-4	0.00043	0.972	0.0207	COD-4	0.00069	0.987	0.0263	SS-4	0.00022	0.982	0.0148

#### 4.4 예측 및 분석

선정된 모형을 학습에 이용되지 않은 1999년 1회 수질자료를 예측한 결과에 대해 통계분석한 결과를 상기 표 3에 나타냈으며, 예측결과는 그림 2~3과 같다. 통계분석은 RMSE와  $R^2$  을 실시하였다.

표 3. 수질예측결과에 대한 통계분석

지점	model	RMSE	$R^2$	model	RMSE	$R^2$	model	RMSE	$R^2$
관촌	BOD-1	0.131	0.642	COD-1	0.053	0.942	SS-1	0.105	0.846
	BOD-2	0.170	0.594	COD-2	0.027	0.947	SS-2	0.089	0.938
	BOD-3	0.163	0.224	COD-3	0.042	0.964	SS-3	0.107	0.964
	BOD-4	0.240	0.284	COD-4	0.059	0.955	SS-4	0.093	0.866
쌍촌	BOD-1	0.134	0.739	COD-1	0.222	0.390	SS-1	0.209	0.699
	BOD-2	0.122	0.666	COD-2	0.084	0.478	SS-2	0.263	0.525
	BOD-3	0.173	0.564	COD-3	0.134	0.479	SS-3	0.184	0.931
	BOD-4	0.147	0.641	COD-4	0.173	0.465	SS-4	0.267	0.168

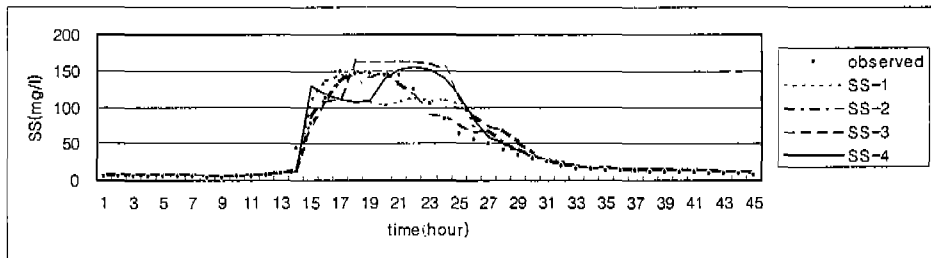
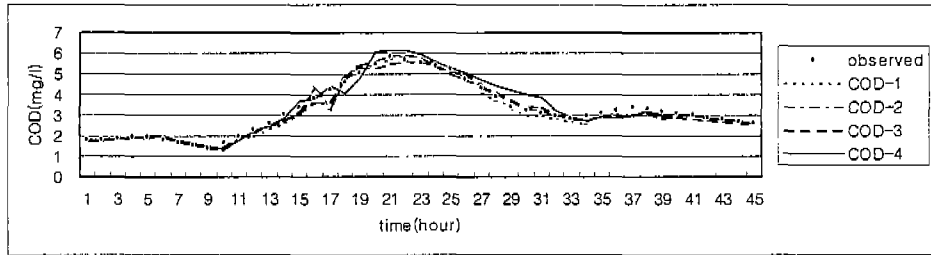
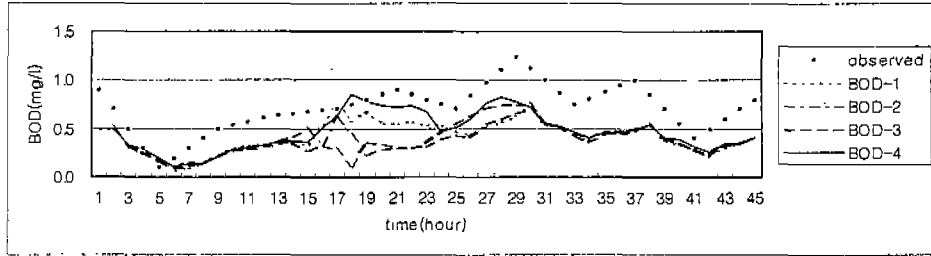
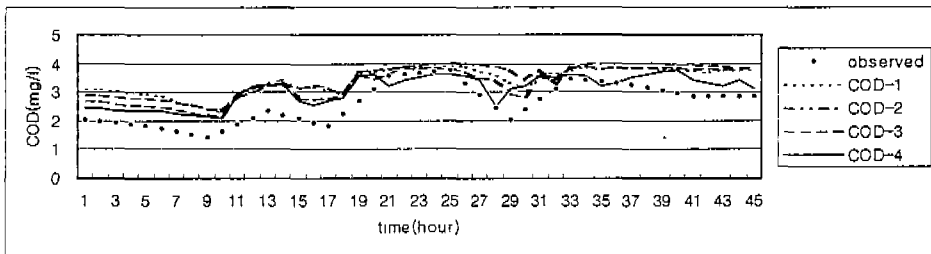
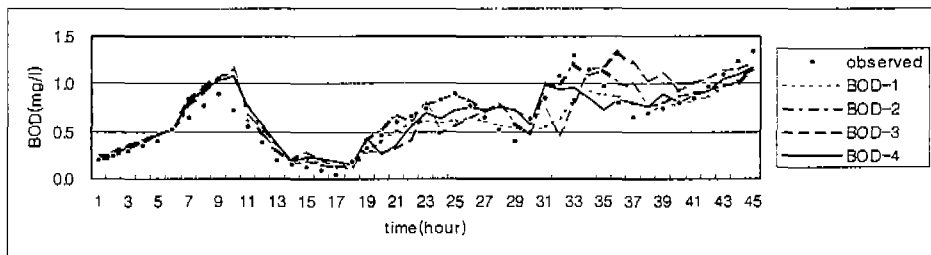


그림 2. 관촌지점 수질 예측결과



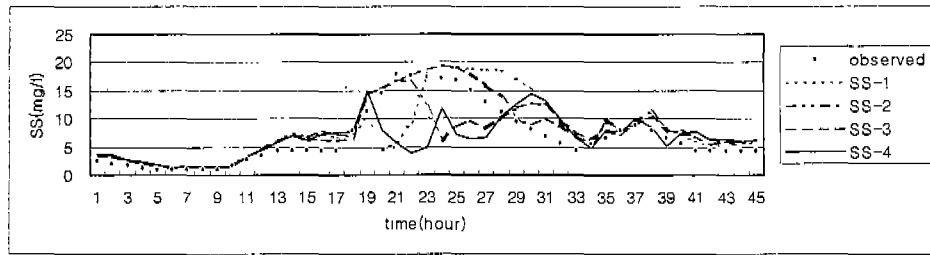


그림 3. 쌍촌지점 수질 예측결과

관촌지점의 예측된 결과를 보면, BOD모형에서는 다소 차이가 있으나 BOD-4모형의 예측결과가 관측치와 비슷한 경향을 나타내었고, COD모형에서는 모든 모형들이 실측값과 거의 비슷한 경향을 나타내었으며, SS모형은 SS-3모형이 오차가 가장 적은 결과를 나타내었다.

쌍치지점은 BOD모형에서 다소 차이가 있으나 BOD-2모형이 예측결과가 가장 관측치와 비슷한 경향을 나타내었고, COD모형에서는 모든 모형들이 경향은 비슷하나 통계치의 정도가 많이 떨어져 나타났으며, SS모형은 SS-3모형이 오차가 가장 적은 결과를 나타내었다.

이와 같이 예측결과가 BOD, SS, COD가 모두가 다소 차이가 있으나 거의 유사한 경향을 나타냈고, 은닉층의 노드개수는 입력층의 개수가  $n$ 개라면  $2n \sim 4n$ 개, 그리고 학습회수는 20,000~25,000회가 적당한 것으로 사료된다. 학습시간과 효율 등을 고려하여 수질모형에 대한 은닉층 노드수와 학습회수의 설정에 대한 세심한 검토가 요구되었으며, 좀더 많은 자료와 적절한 수질인자간의 상호 연관된 물리적 수질모형을 활용할 수 있다면 신경망이론을 응용한 수질예측 및 추정에 대해서도 충분히 활용할 수 있을 것으로 판단되었다.

## 5. 결론

본 연구에서는 신경망이론을 이용하여 섬진강다목적댐 상류유역의 관촌지점과 쌍치지점에 대한 단기 수질예측을 위한 활용가능성을 판단하기 위해 다층 신경망모형을 구성하였다. 1998년의 5회 수질자료, 1999년의 7회 수질자료, 그리고 동일한 기간의 유출량자료를 이용하여 BOD, COD, SS에 대해 총 12가지 모형을 학습하였고, 모형의 검증에 대해 학습에 이용되지 않은 1회 수질자료에 대해 예측하였다. 예측 결과는 비교적 양호한 결과를 나타냈으며 좀더 향상된 결과를 얻기 위해서는 다양한 구조의 신경망모형의 구성, 양질의 많은 자료 그리고 수질자료의 특성을 명확히 구분할 수 있는 신경망모형에 대한 연구가 진행되어야 할 것으로 판단되었다.

## 참고문헌

1. Demuth, H., and M., Beale, 1994. *Neural Network Toolbox : For Use with MATLAB User's Guide*, The Math Works Inc., pp. 5.1~5.38.
2. Haykin, S., 1994. *Neural Networks : A Comprehensive Foundation*. Prentice Hall.
3. Peter A. K., and V., Novotny, 1980. *Water Quality Management*. Academic Press.
4. 김만식, 1998. 다목적 저수지의 홍수유입량 예측을 위한 신경망모형, 충북대학교, 박사학위논문.
5. 한국수자원공사, 1998. 섬진강 다목적댐 수질관리 기법 연구(1차년도), 연구보고서.
6. 한국수자원공사, 1999. 섬진강 다목적댐 수질관리 기법 연구(2차년도), 연구보고서.
7. 이요상, 강병수, 2001. 댐저수지 유역의 오염부하 유출특성 - 섬진강댐 저수지를 중심으로 -, Vol. 33, No. 6, pp.757~764.