

# 신경망을 이용한 영산강 수질예측에 관한 연구

○오창렬<sup>1)</sup>, 김용구<sup>2)</sup>, 박성천<sup>3)</sup>

## 1. 서론

수질관리란 수자원의 양적인 면과 질적인 면 중에서 질적인 면을 대상으로 하여 수자원을 용도에 맞도록 관리하는 것이라 할 수 있다. 이러한 수질관리의 문제에 있어 합리적이며 경제적인 수질관리를 하기 위해서는 자연적 정화능력인 하천의 자정작용을 최대한 이용하고, 타당성 있고 현실적인 수질기준을 설정하여 경제적 손실을 최소화하는 작업이 필요하다. 그리고 이러한 작업을 위해 수질예측 모형과 선형계획을 기초로 하는 최적화 기법이 많이 사용되고 있다.

따라서 본 연구는 영산강 유역의 대표지점이며 동시에 수질관리 대상지점인 나주지점을 선정하여 비선형적인 수질의 변화를 예측하고 그에 따른 효과적인 수질보전정책을 수립하기 위하여 수계 내의 물리적·화학적·생물학적 반응 과정을 나타내지 않으면서 시행착오방법(trial and error method)으로 적절한 입력변수를 선택하는 인공신경망의 역전파학습알고리즘을 이용하여 비선형 모형을 개발하였다.

## 2. 신경망 이론

과거로부터 인간들은 뇌가 어떻게 외부로부터의 입력에 반응하며 그 원리가 무엇인지에 대하여 의문을 가지고 연구하여 왔으며 인공신경망(Artificial Neural Networks, ANN)에 관한 연구는 뇌가 작용하는 원리를 수학적이고 전산적인 처리에 어떻게 접목할 수 있을까? 라는 의문속에서 시작되었다. 인간의 뇌(Brain)는 매우 복잡하고 비선형적이며 자료들을 병렬적으로 처리하는 일종의 컴퓨터와 유사하다. 또한, 그러한 행위 및 계산을 가능하게 하는 것을 뉴런(neuron)이라 하며, 자체의 구성체를 조직하는 능력을 가지고 있다. 이러한 인공신경망의 인공 뉴런이 갖는 입력과 가중치 벡터를 곱해서 더한 결과를 하나의 전달함수에 적용하는 방식으로 계산되는 처리요소이다.

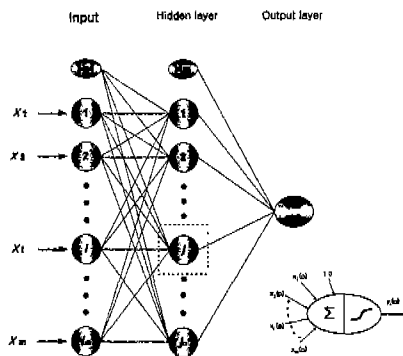


그림 1. 다층신경망 모형의 구조

하나의 특정 뉴런에서 임계값을 원하지 않은 경우에는 생략할 수 있다. 보편적으로 전달함수는 설계자에 의해 선택되고, 인자들은 뉴런 입출력과 관계가 어떤 특정 목표에 도달될 수 있도록 하는 학습규칙에 의해 조정된다. 가끔 전달함수는 활동함수로, 임계값은 편입력 또는 offset로 일컬어진다. 가중치는 생물학적인 뉴런에서 시냅스의 강도에 해당되고, 세포 몸체는 덧셈

1) 동신대학교 환경공학과 공학석사  
 2) 동신대학교 환경공학과 석사과정  
 3) 동신대학교 토목환경공학부 부교수

과 전달함수로 표현되며, 뉴런출력은 축색에서 신호를 나타낸다.

이처럼 인공신경망 역시 인공적인 뉴런들로 구성되며, 인간의 두뇌가 특별한 기능이나 임무를 수행하도록 하는 방법을 모형화하도록 구성되어진 구조이다.

인공신경망은 인간 두뇌의 뉴런과 같은 기능을 수행하며, 수학적 연산능력을 가진 노드들이 상호 연결된 것으로 적절한 학습규칙에 의하여 작동한다. 즉, 각 노드들은 결합함수와 전달함수에 의해 수학적 연산을 수행한다. 노드로 입력되는 신호는 식 1의  $S_j$  같이 가중치가 부여된 값들의 합이 되며, 이 값은 다시 전달함수라 불리는 비선형 함수를 통과하여 식 2의  $y_j$ 가 된다.

$$s_j = \sum_{i=0}^m w_{ji} x_i \quad \dots\dots \text{식 1} \quad y_j = f(s_j) = \frac{1}{1 + \exp(-s_j)} \quad \dots\dots \text{식 2}$$

여기서,  $s_j$ 의 범위는  $\pm\infty$ 이며,  $y_j$ 의 범위는 0에서 1까지 이다.

본 연구에 사용된 인공 신경망의 종류는 다층(multi-layer) 신경망이며, 입력층과 은닉층, 그리고 출력층으로 구성된다. 또한 매개변수 최적화를 위하여 사용된 학습 알고리즘은 모멘텀 상수와 적응식 학습율이 적용된 최급 강하법을 이용한 역전파 학습 알고리즘이며, 이에 포함된 활성화 함수는 단극성 시그모이드(sigmoid) 함수와 선형(linear) 함수를 사용하였다.

### 3. 대상지점 및 수집자료

본 연구의 대상지점은 우리나라 5대강의 하나인 영산강 수계의 나주지점에 대하여 수질농도 예측을 시행하였으며, 입력 자료는 1990년 ~ 1999년까지의 수질자료를 이용하였다. 수질자료는 영산강 환경관리청에서 조사 보관중인 자연하천의 수질 자료 24개 항목 중 BOD, DO, T-N, T-P, SS, 수온 항목의 월자료와 영산강 홍수통제소의 일수위 자료를 해당 년의 수위-유량관계곡선식에 대입하여 유출량으로 환산한  $Q_{max}$ ,  $Q_{min}$ ,  $Q_{avg}$  자료를 사용하였다.

### 4. 모형의 구성

본 연구에서는 영산강 유역의 나주지점에 대한 하천의 수질 조사항목인 DO농도, BOD농도, T-N농도, T-P농도를 예측하기 위하여 신경망 모형을 구성하였으며 모형의 입력층 노드의 수에 따라 은닉층 노드의 수가 조절되도록 하였다. 즉, 각 수질농도를 예측하기 위한 모형은 은닉층의 노드의 수를  $n \sim 3n$ 개까지 변화시키면서 각각의 모형을 적용하였으며, 이러한 일련의 과정을 거친 모형들 중  $CC$ 와  $MSE$ 를 평가 기준으로 하여 모형을 선별하였다.

수질농도 예측모형의 입력자료의 구축은 시행착오법에 의하여 적절한 입력변수를 선정하였다. 예측 수질항목 중 DO, BOD농도를 예측하기 위하여 DO농도 및 BOD농도는 시차를 4로 고정하였고, T-N, T-P농도 수질항목을 예측하기 위하여 T-N농도 및 T-P농도의 시차를 3으로 고정하였다. 수온, SS농도 및 월최소 평균 유출량, 월평균 유출량, 월최대 평균 유출량 자료는 식 3~6과 같이 각각 시차를 달리하여 모형을 구성하였다.

$$\text{Model I, II, III, IV} \begin{bmatrix} do_{k-1}, & do_{k-2}, & do_{k-3}, & do_{k-4} \\ bod_{k-1}, & \dots & \dots & bod_{k-4} \\ temp_{k-1}, & \dots & \dots & temp_{k-4} \\ qmin_{k-1}, & \dots & \dots & qmin_{k-4} \\ qavg_{k-1}, & \dots & \dots & qavg_{k-4} \\ qmax_{k-1}, & \dots & \dots & qmax_{k-4} \end{bmatrix} \dots \text{식 3}$$

$$\text{Model I, II, III, IV} \begin{bmatrix} bod_{k-1}, & bod_{k-2}, & bod_{k-3}, & bod_{k-4} \\ temp_{k-1}, & \dots & \dots & temp_{k-4} \\ t-n_{k-1}, & \dots & \dots & t-n_{k-4} \\ qmin_{k-1}, & \dots & \dots & qmin_{k-4} \end{bmatrix} \dots \text{식 4}$$

$$\text{Model I, II, III } \begin{bmatrix} t^{-n}_{k-1}, & t^{-n}_{k-2}, & t^{-n}_{k-3} \\ t^{-p}_{k-1}, & \dots & t^{-p}_{k-3} \\ qmin_{k-1}, & \dots & qmin_{k-3} \\ qavg_{k-1}, & \dots & qavg_{k-3} \end{bmatrix} \dots \text{식 5} \quad \text{Model I, II, III } \begin{bmatrix} t^{-p}_{k-1}, & t^{-p}_{k-2}, & t^{-p}_{k-3} \\ t^{-n}_{k-1}, & \dots & t^{-n}_{k-3} \\ SS_{k-1}, & \dots & SS_{k-3} \\ qmax_{k-1}, & \dots & qmax_{k-3} \end{bmatrix} \dots \text{식 6}$$

여기서,  $do_k$ 는 나주 지점의 DO예측농도로서  $do_{k-1}, do_{k-2}, do_{k-3}, do_{k-4}$ 는 시차 1에서 4까지 DO농도,  $bod_k$ 는 나주 지점의 BOD예측농도로서  $bod_{k-1}, bod_{k-2}, bod_{k-3}, bod_{k-4}$ 는 시차 1에서 4까지 BOD농도,  $T-N_k$ 는 나주 지점의 T-N예측농도로서  $t-n_{k-1}, t-n_{k-2}, t-n_{k-3}, t-n_{k-4}$ 는 시차 1에서 4까지 T-N농도,  $T-P_k$ 는 나주 지점의 T-P예측농도로서  $t-p_{k-1}, t-p_{k-2}, t-p_{k-3}$ 는 시차 1에서 3까지 T-P농도,  $temp_{k-1}, temp_{k-2}, temp_{k-3}, temp_{k-4}$ 는 시차 1에서 4까지의 수온,  $q-min_{k-1}, q-min_{k-2}, q-min_{k-3}, q-min_{k-4}$ 는 시차 1에 4까지의 월최소 평균유출량,  $q-avg_{k-1}, q-avg_{k-2}, q-avg_{k-3}, q-avg_{k-4}$ 는 시차 1에서 4까지의 월평균 유출량,  $q-max_{k-1}, q-max_{k-2}, q-max_{k-3}, q-max_{k-4}$ 는 시차 1에 4까지의 월최대 평균유출량이다.

### 5. 모의결과 및 고찰

일반화된 모형을 선별하기 위해 수치적인 기준과 도식적인 기준을 적용하였으며 수치적인 기준으로 아래의 식의 CC 및 MSE를 사용하였다.

$$CC = \frac{Cov(X, Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}} \dots \dots \dots \text{식 7} \quad MSE = \frac{1}{n} \sum_{i=1}^n (y_i - Y_i)^2 \dots \dots \dots \text{식 8}$$

모형의 선별과정은 각 농도의 예측모형 중 훈련 및 보정과정에서 예측의 적용성이 뛰어난 모형을 선택하였고, 검증과정은 훈련 및 보정과정에서 선별된 모형을 대상으로 검증을 실시하여 수치적인 기준과 도식적인 기준인 산포도를 적용하여 우수한 모형을 최종 모형으로 선택하였다.

#### 5-1. DO농도 모형

본 연구에서 DO농도의 예측 모형을 개발하기 위하여 식 3과 같이 구성한 모형 중 Model II의 예측력은 타 모형에 비하여 예측력의 우수성이 인정되었다. 상기 Model II에 대하여 은닉층의 노드의 수를 달리하여 적용하였으며 그 통계적 특성치의 결과는 표 1과 같다.

표 1. Statistical analysis of Model II (DO) and selected models

Model	Hidden layer	Training		Calibration		Verification		Model	Hidden layer	Training		Calibration		Verification	
		MSE	CC	MSE	CC	MSE	CC			MSE	CC	MSE	CC	MSE	CC
DO_ANN(14, 14, 1)	n	0.921	0.847	1.591	0.811	1.036	0.803	DO_ANN(14, 26, 1)	n+12	0.626	0.886	0.964	0.897	1.179	0.801
DO_ANN(14, 15, 1)	n+1	0.721	0.881	1.041	0.881	1.314	0.844	DO_ANN(14, 30, 1)	2n+2	0.636	0.894	1.325	0.845	0.698	0.870
DO_ANN(14, 21, 1)	n+7	0.807	0.868	1.310	0.847	1.119	0.805	DO_ANN(14, 35, 1)	2n+8	0.787	0.870	0.629	0.929	1.024	0.808
EO_ANN(14, 24, 1)	n+10	0.660	0.885	1.342	0.856	1.160	0.875	DO_ANN(14, 38, 1)	2n+10	0.806	0.872	1.169	0.865	1.077	0.809

최종 모형의 선택은 1차적으로 선별된 모형 중 검증에서 MSE와 CC가 가장 우수한 DO\_ANN(14, 30, 1)의 모형을 본 연구에서 DO농도 모형으로 선택하였다. 그림 2는 본 모형의 훈련 및 검증과정에서의 도식적인 표현을 나타내고 있다.

검증과정에서 본 모형이 DO\_ANN(14, 24, 1)모형에 비하여 예측값이 실측값에 근사한 예측력을 보이고 있다. 검증과정의 이중누가곡선을 볼 때 DO\_ANN(14, 24, 1) 모형은 실측값에 과소평가가 되고 있으나 본 모형은 실측값과 거의 유사한 예측력을 보임으로써 본 모형이 나주지점 수계의 DO농도에 대한 특성을 가장 잘 반영하고 있는 것으로 판단된다.

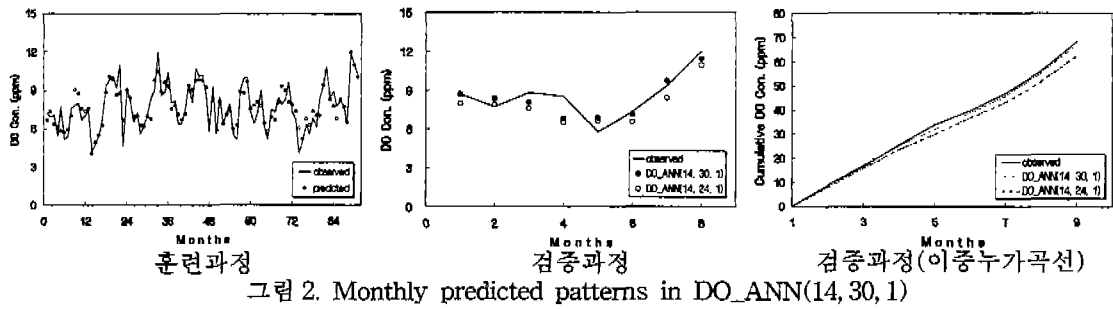


그림 2. Monthly predicted patterns in DO\_ANN(14, 30, 1)

### 5-2. BOD농도 모형

본 연구에서 BOD농도의 모형은 식 4와 같이 4종류의 모형을 구성하여 적용한 결과 Model II가 타 모형에 비하여 예측력의 우수성이 인정되었다. Model II에 대하여 은닉층의 노드의 수를  $n$ 부터  $3n$ 까지 순차적으로 증분시키면서 모형을 적용하였으며 그 통계적 특성치는 표 2와 같다. 최종 모형의 선택은 1차적으로 선택된 모형 중 검증에서 MSE와 CC가 가장 우수한 BOD\_ANN(10, 19, 1)의 모형을 본 연구에서 BOD농도 모형으로 선택하였다.

표 2 Statistical analysis of Model II (BOD) and selected models.

Model	Hidden layer	Training		Calibration		Verification		Model	Hidden layer	Training		Calibration		Verification	
		MSE	CC	MSE	CC	MSE	CC			MSE	CC	MSE	CC	MSE	CC
BOD_ANN(10, 10, 1)	n	0.047	0.990	0.240	0.887	3.817	0.799	BOD_ANN(10, 19, 1)	n+9	0.221	0.960	0.568	0.732	1.141	0.887
BOD_ANN(10, 12, 1)	n+2	0.085	0.982	0.349	0.812	1.958	0.813	BOD_ANN(10, 20, 1)	2n	0.059	0.987	0.476	0.754	3.968	0.883
BOD_ANN(10, 14, 1)	n+4	0.826	0.974	0.312	0.826	3.196	0.860	BOD_ANN(10, 23, 1)	2n+3	0.063	0.986	0.446	0.793	4.257	0.781
BOD_ANN(10, 15, 1)	n+5	0.080	0.982	0.378	0.790	2.732	0.793	BOD_ANN(10, 25, 1)	2n+5	0.106	0.978	0.389	0.790	3.494	0.874
BOD_ANN(10, 17, 1)	n+7	0.020	0.996	0.363	0.849	3.008	0.886	BOD_ANN(10, 28, 1)	2n+8	0.013	0.997	0.370	0.888	3.176	0.819

그림 3은 본 모형에 대한 도식적인 표현을 나타내고 있으며 본 모형과 유사한 예측력을 보이는 BOD\_ANN(10, 17, 1)과 비교 검토하여 보았다.

검증과정에서의 이중누가곡선 그림을 볼 때 BOD\_ANN(10, 17, 1) 모형은 실측값에 과대평가가 되고 있으나 본 모형은 BOD\_ANN(10, 17, 1) 모형에 비하여 실측값과 거의 유사한 예측력을 보임으로서 본 모형이 나중지점 수계의 BOD농도에 대한 특성을 가장 잘 반영하고 있는 것으로 평가되었다.

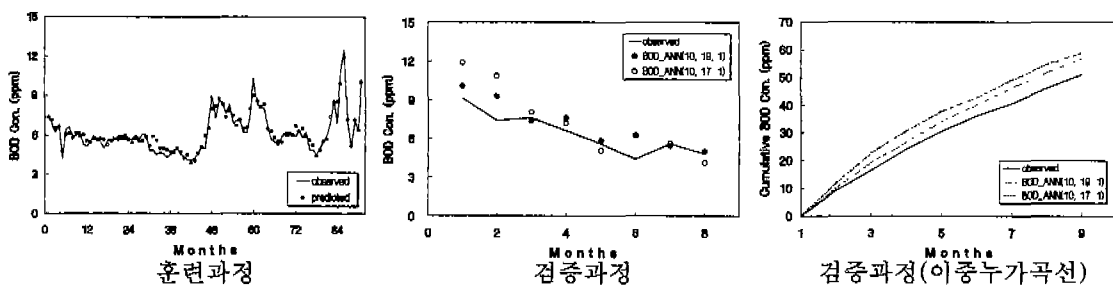


그림 3. Monthly predicted patterns of BOD\_ANN(10, 19, 1)

### 5-3. T-N농도 모형

본 연구에서 T-N농도의 모형을 개발하기 위하여 식 4와 같이 4개의 모형을 구성하여 적용한 결과 Model II의 예측력의 우수성이 인정되었으며, Model II에 대하여 DO농도 모형과 같이 은닉

층의 노드의 수를  $n$ 부터  $3n$ 까지 순차적으로 증분시켜 모형을 적용하였다. 그에 대한 통계적 특성치는 표 3에 나타내었다. 최종 모형의 선택은 1차적으로 선택된 모형 중 검증에서  $MSE$ 와  $CC$ 가 가장 우수한 T-N\_ANN(9, 12, 1)의 모형을 본 연구에서 T-N농도 모형으로 선택하였다.

표 3. Statistical analysis of Model II(T-N) and selected models.

	Hidden layer	Training		Calibration		Verification			Hidden layer	Training		Calibration		Verification	
		MSE	CC	MSE	CC	MSE	CC			MSE	CC	MSE	CC		
T-N_ANN(9, 9, 1)	n	1.925	0.929	5.583	0.776	0.472	0.879	T-N_ANN(9, 18, 1)	2n	1.405	0.949	5.737	0.710	0.792	0.871
T-N_ANN(9, 11, 1)	n+2	0.212	0.993	5.457	0.711	0.392	0.919	T-N_ANN(9, 20, 1)	2n+2	1.205	0.958	3.857	0.801	0.649	0.904
T-N_ANN(9, 12, 1)	n+3	<b>0.861</b>	<b>0.969</b>	<b>4.148</b>	<b>0.781</b>	<b>0.362</b>	<b>0.920</b>	T-N_ANN(9, 22, 1)	2n+4	0.579	0.980	5.418	0.759	0.870	0.909
T-N_ANN(9, 14, 1)	n+5	0.387	0.987	5.465	0.765	1.006	0.959	T-N_ANN(9, 23, 1)	2n+5	0.560	0.980	4.979	0.737	0.556	0.879
T-N_ANN(9, 15, 1)	n+6	0.935	0.968	6.424	0.702	0.500	0.892	T-N_ANN(9, 24, 1)	2n+6	1.022	0.963	4.036	0.792	1.087	0.841
T-N_ANN(9, 16, 1)	n+7	0.685	0.975	5.672	0.710	0.981	0.834	T-N_ANN(9, 25, 1)	2n+7	0.860	0.970	5.138	0.745	1.057	0.859

그림 4는 본 모형에 대한 도식적인 표현을 나타내고 있으며 본 모형과 유사한 예측력을 보이는 T-N\_ANN(9, 14, 1)과 비교 검토하여 보았다. T-N\_ANN(9, 14, 1)의 모형은 본 모형보다는 검증 과정에서 높은  $MSE$ 를 나타내고 있으나  $CC$ 는 높게 나타나고 있다.

검증과정에서는 그림 4와 같이 전체적으로 본 모형이 T-N\_ANN(9, 14, 1)모형에 보다는 우수한 것으로 판단되었으며 이증누가곡선 볼 때 본 모형은 우수한 예측력을 보이고 있으며 T-N\_ANN(9, 14, 1)모형은 전반적으로 과대평가를 보이고 있다. 따라서, 본 모형인 T-N\_ANN(9, 12, 1)이 나주지점 수계의 T-N농도에 대한 특성을 가장 잘 반영하고 있는 것으로 평가되었다.

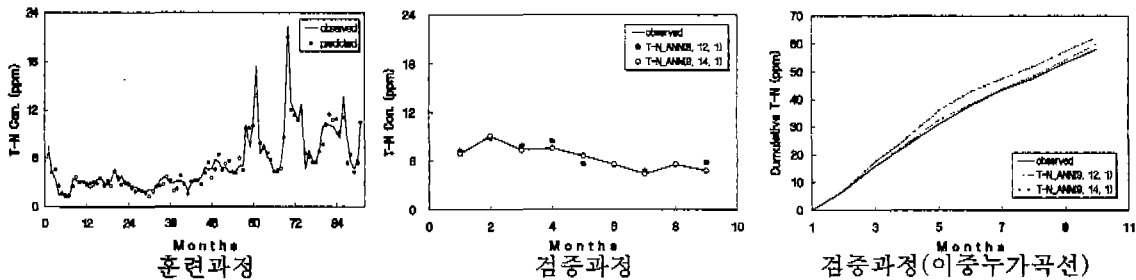


그림 4. Monthly predicted patterns of T-N\_ANN(9, 12, 1)

#### 5-4. T-P농도 모형

본 연구에서 T-P농도의 모형을 개발하기 위하여 식 6과 같이 시차에 따라 3종류의 모형을 구성하여 적용한 결과 Model II의 예측력이 타 모형보다는 우수한 것으로 판단되었으며 상기 Model II에 대하여 은닉층의 노드의 수를  $n$ 부터  $3n$ 까지 순차적으로 증분시켜 모형을 적용하였다. 은닉층 노드 수의 변화에 따른 통계적 특성치를 표 4에 나타내었다.

표 4. Statistical analysis of Model II(T-P) and selected models.

	Hidden layer	Training		Calibration		Verification			Hidden layer	Training		Calibration		Verification	
		MSE	CC	MSE	CC	MSE	CC			MSE	CC	MSE	CC		
T-P_ANN(9, 9, 1)	n	0.025	0.784	0.014	0.748	0.009	0.477	T-P_ANN(9, 21, 1)	2n+3	0.016	0.872	0.012	0.830	0.016	0.162
T-P_ANN(9, 17, 1)	n+8	0.013	0.892	0.007	0.894	0.014	0.412	T-P_ANN(9, 23, 1)	2n+4	0.017	0.860	0.016	0.715	0.013	0.338
T-P_ANN(9, 19, 1)	2n+1	0.022	0.817	0.015	0.744	0.007	0.607	T-P_ANN(9, 24, 1)	2n+5	0.021	0.826	0.016	0.717	0.008	0.603
T-P_ANN(9, 20, 1)	2n+2	0.025	0.792	0.014	0.746	0.010	0.410	T-P_ANN(9, 25, 1)	2n+6	<b>0.024</b>	<b>0.811</b>	<b>0.014</b>	<b>0.745</b>	<b>0.006</b>	<b>0.688</b>

최종 모형의 선택은 1차적으로 선택된 모형 중 검증과정에서  $MSE$ 와  $CC$ 가 가장 우수한 T-P\_ANN(9, 25, 1)의 모형을 본 연구에서의 T-P농도 모형으로 선택하였다.

T-P농도 예측을 위한 모형을 다른 수질농도 예측모형과 비교하여 볼 때 검증과정에서 다소 낮은 상관계수를 나타내고 있는 것을 알 수 있으나 T-P농도 예측을 위한 여러 모형 중에서 본 모형이 가장 우수한  $CC$ 값을 나타내고 있으며, 도식적인 표현을 기준으로 비교하여 볼 때 우수한 예측력을 보이는 것으로 평가되었다.

이중누가곡선은 그림 5와 같이 본 모형은 실제값에 근사한 예측력을 보이고 있으나 T-P\_ANN(9, 19, 1)모형은 전반적으로 과대평가를 보이고 있다. 따라서, 본 모형이 나주지점 수계의 T-P농도에 대한 특성을 가장 잘 반영하고 있으며 우수한 예측력을 보이는 것으로 평가되었다.

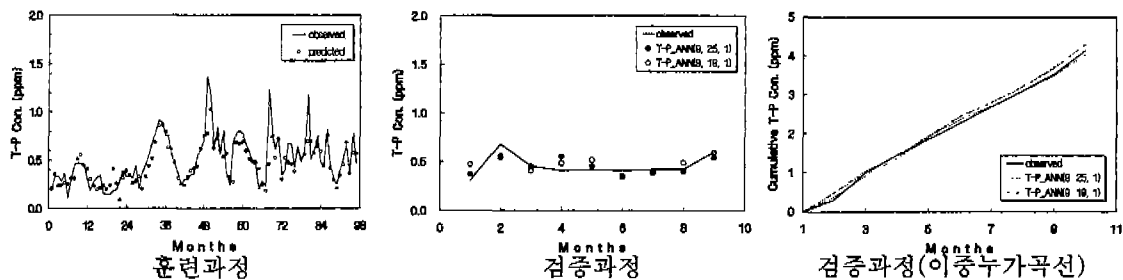


그림 5. Monthly predicted patterns of calibration in T-P\_ANN(9, 25, 1)

## 6. 결론

본 연구는 영산강 유역의 대표지점인 나주지점을 선정하여 비선형적인 수질의 변화를 예측하고 그에 따른 효과적이며 경제적인 수질보전정책을 수립하는 등 보다 적극적인 수질관리를 도모하기 위하여 인공 신경망이론을 적용한 DO농도, BOD농도, T-N농도, T-P농도의 수질예측에 대한 모형을 개발하였다.

인공신경망 모형은 은닉층의 뉴런의 수가 많다고 하여 반드시 모형의 적합성 및 일반화가 구현되는 것은 아니라는 것을 알 수 있었으며, 본 모형은 수질자료와 유량자료를 입력자료로 구축하여 개발한 모형으로 유량과 수질과의 상관성을 다소 입증할 수 있었다. 또한, 본 연구에서 사용한 자료는 월자료로서 자료의 빈약성에도 불구하고 우수한 모형을 개발할 수 있었으나 일자료 또는 시자료의 구축이 선행된다면 보다 더 우수한 모형을 개발할 수 있을 것으로 판단된다.

## [참고문헌]

1. 오창렬, "인공신경망을 이용한 하천수 수질예측", 동신대학교 석사학위 논문, 2001.
2. 영산강 홍수통제소, "한국수문조사연보", 건교부, 1990~1999
3. 영산강 환경관리청, "환경연보", 환경부, 1990~1999
4. DeSilets, L., Golden, B., Wang, Q., kumar R., "Prediting salinity in the Chesapeake Bay using backpropagation", Computer and Operation Research Vol.19, No. 3/4, pp. 227~285, 1992
5. Ching-gung Wen, Chin-Sheng Lee "A neural network approach to multiobjective optimization for water quality management in a river basin" Water Resources Research, Vol. 34, No. 3, pp. 427~436, 1998.