

MCSM과 신경망기법을 연계한 하천수질의 불확실성 해석

○김성원*·지흥기**·이순탁***

1. 서론

수질모델링은 허용 오염물의 용량을 결정하고 여러 오염물질의 배분적인 측면을 예측하는 효율적인 방법을 제공함으로써 환경적인 관리에 기여한다(Lung, 1998). 그러나 전통적인 수질모형화과정은 사용자자료와 모형의 매개변수에 있어서 많은 불확실성을 이론적 및 효과적으로 설명하기가 난이하기 때문에 이용에 있어서 제한적인 것이 사실이다. 공학시스템에 있어서 모형화의 불확실성을 취급하는 방법 중 지금까지 가장 많이 사용되는 방법은 확률적인 이론법이며, 그 중에서도 Monte Carlo Simulation Method(MCSM)은 환경공학 분야에서 실무에 가장 빈번하게 그리고 성공적으로 적용되어 온 방법론이다(Ayyub and Gupta, 1997; Duke et al., 1998). MCSM은 모형 매개변수들을 확률밀도함수(Probability Density Functions, PDFs)에 의하여 나타난 무작위변수로 가정하고, 무작위변수들의 확률분포는 무작위 매개변수들의 상호작용에 의해 결정되어진다. 본 연구에서는 하천 및 호소에서의 부영양화의 원인이 되는 수질오염인자인 낙동강유역 삼랑진지점의 T-N과 T-P의 월별 수질오염예측을 위하여 수질오염예측 신경망모형을 구성하였으며, 입력자료는 왜관대교와 고령교지점의 월별 T-N과 T-P자료로 구성되어 있다. 또한 신경망모형의 훈련자료이자 입력자료인 왜관대교와 고령교지점의 월별 T-N과 T-P 자료의 부족으로 인하여 MCSM에 의하여 입력자료를 무작위 발생시켰으며, 훈련자료의 불확실성에 따른 하천수질예측의 불확실성을 해석하였다. 즉 MCSM에 의하여 수질오염예측 신경망모형의 입력자료를 난수로서 무작위 발생시킨 후, 무작위 발생시킨 각 자료를 이용하여 수질오염예측 신경망모형을 훈련시켜 최적 매개변수를 산정한다. 그리고 산정된 매개변수를 이용하여 수질오염예측 신경망모형을 검증하여 수질오염인자의 불확실성에 따른 하천수질오염을 예측 및 분석하고자 한다.

2. 자료의 수집

인공신경망을 이용하여 하천의 수질오염인자의 모형화기법중 BOD, DO 및 T-N의 변화에 대한 모형화과정에 대한 연구가 선행되었다(안상진 외, 2001; 조현경, 2000). 그러나 본 연구에서 대상 수질오염인자는 T-N과 T-P자료이다. 낙동강유역의 주요 유량관측지점인 왜관대교와 금호강 합류후인 고령교지점 및 낙동강 하류의 삼랑진지점의 월별 수질자료인 T-N과 T-P자료이다. 왜관대교, 고령교지점 및 삼랑진지점의 수질오염인자 관측의 경우 1997년부터 시작되었으며, 대구지방환경청 및 낙동강유역환경청 홈페이지의 월평균 자료는 1997년 1월부터 2003년 2월까지의 자료가 구축되어 있으며, 1997년 1월부터 2002년 12월까지의 자료를 구축하여 이용하였다.

3. Monte Carlo Simulation Method(MCSM)의 개요

본 연구에서는 수질오염예측 신경망모형의 입력자료를 무작위 발생시키기 위하여 MCSM기법을 사용하였다. MCSM기법의 원래의 의미는 Random Walks를 포함하는 모의기법을 의미하고 있으나, 현재에 와서는 난수(Random Number)를 발생시키는 모든 종류의 모의기법을 나타내고 있다. 따라서 본 연구에서는 수질오염예측 신경망모형의 입력자료를 구성하기 위한 난수를 무작위 발생시키기 위하여 Multiplicative Congruential Generator를 이용하였으며 다음 식[1] 및 [2]와 같다.

* 정희원, 동양대학교 지구환경시스템공학과, 전임강사(E-mail : swkim68@phenix.dyu.ac.kr)

** 정희원, 영남대학교 토목도시환경공학부, 교수(E-mail : jeehk@yucc.yeungnam.ac.kr)

*** 정희원, 영남대학교 토목도시환경공학부, 교수(E-mail : leest@yucc.yeungnam.ac.kr)

$$X_{i+1} = aX_i \pmod{m} \quad [1]$$

$$U_i = \frac{X_i}{m} \quad [2]$$

여기서 $a=16,807$, $m=2^{31}-1$, \pmod{m} 은 modulus m 에 의한 나머지 값, X_i 는 균등난수를 발생시키기 위한 초기 값, U_i 는 균등분포에 의한 난수를 나타낸다. 따라서 MCSM과 신경망을 연계한 모형화과정은 다음과 같이 나타낼 수 있다.

[Step1] 수질오염예측 신경망모형의 입력자료인 왜관대교, 고령교지점 및 삼랑진지점의 T-N 및 T-P자료의 확률분포형을 정규분포(Normal Distribution)로 가정한다.

[Step2] 가정한 확률분포형인 정규분포에 따라서 왜관대교, 고령교지점 및 삼랑진지점의 T-N 및 T-P자료의 난수를 200, 500 및 1,000개씩 무작위 발생시킨다.

[Step3] 입력층노드의 자료가 왜관대교와 고령교지점의 T-N 및 T-P자료이고, 출력층노드의 자료가 삼랑진지점의 T-N 및 T-P자료로 구성된 4-4-2 구조의 수질오염예측 신경망모형을 구성한다.

[Step4] [Step2]에서 발생한 난수를 이용하여 수질오염예측 신경망모형을 훈련시켜 매개변수인 최적 연결강도와 편차를 산정한다.

[Step5] 대구지방환경청 및 낙동강유역환경청의 실제 수질자료를 입력자료로 하여 [Step4]에서 산정된 매개변수인 최적 연결강도와 편차를 이용하여 수질오염예측 신경망모형을 검증한다.

[Step6] 발생한 난수의 자료수에 따른 수질오염예측 신경망모형의 불확실성을 검토한다.

4. 확률분포형의 선정

본 연구에서는 난수발생을 위한 확률분포형으로 정규분포를 선택하였으며, 확률분포형에 대한 난수는 각각의 평균과 표준편차를 가지는 자료를 200, 500 및 1,000개씩 무작위 발생시켰다. 왜관대교, 고령교지점 및 삼랑진지점의 수질오염인자에 대한 평균과 표준편차는 다음 표 1과 같이 나타낼 수 있으며, 그림 1과 2는 왜관대교, 고령교지점 및 삼랑진지점의 T-N 및 T-P 수질오염인자의 월별변화량을 나타낸 것이다.

표 1. T-N 및 T-P 수질오염인자의 평균과 표준편차(1997년 1월 - 2002년 12월)

지점	자료	평균값(PPM)	표준편차(PPM)
왜관대교	T-N	3.357	0.132
	T-P	0.696	0.072
고령교지점	T-N	5.078	0.256
	T-P	1.368	0.088
삼랑진지점	T-N	3.851	0.124
	T-P	1.213	0.043

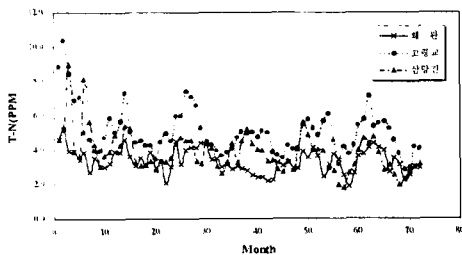


그림 1. T-N 수질오염인자의 월 변화 추이도

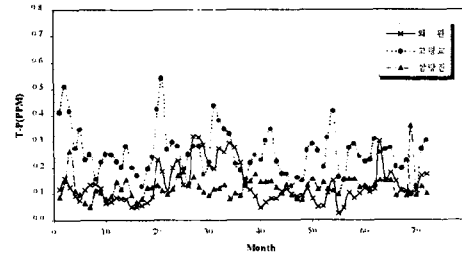


그림 2. T-P 수질오염인자의 월 변화 추이도

5. 수질오염예측 신경망모형의 구축

본 연구에서 제시하는 신경망모형의 구조는 다음 그림 3과 같다. 또한 본 연구의 신경망모형은 역전파 알고리즘을 이용한 다층퍼셉트론(MultiLayer Perceptron, MLP)형태의 기법이며, 역전파 알고리즘은 주어진 문제에 대하여 비최적화 해법을 나타내는 구배강하법이다. 본 연구에 사용된 역전파 알고리즘은 일단계 활성역전파(One Step Secant BackPropagation

OSSBP)알고리즘과 탄성역전파(Resilient BackPropagation, RBP)알고리즘이며, 상세한 내용은 다음과 같다.

5.1 OSSBP 알고리즘

OSSBP 알고리즘은 출력치에 높은 정도의 값이 요구될 때 많은 차수에 의하여 훈련단면을 가속화할 수 있으며, 할선법(Secant method)과 공액구배법(Conjugate gradient method)사이에서 강한 연결성을 제공한다(Battiti, 1992; Broomhead and Lowe, 1988; 김성원 외, 2001). OSSBP 알고리즘의 역전파는 연결강도와 편차벡터에 관해서 오차의 평방합이 허용치이하가 되는 미분함수를 계산하는데 사용된다.

5.2 RBP 알고리즘

RBP 알고리즘은 국부적인 구배정보에 기초를 둔 연결강도나 편차벡터 단계의 직접적인 적합성을 수행하는 알고리즘이며, 이 알고리즘을 수행하기 위하여 각 연결강도나 편차벡터의 갱신크기(Update size)를 결정하는 각 연결강도나 편차벡터의 갱신값(Update value)인 Δ_k 를 도입한다(Riedmiller and Braun, 1993).

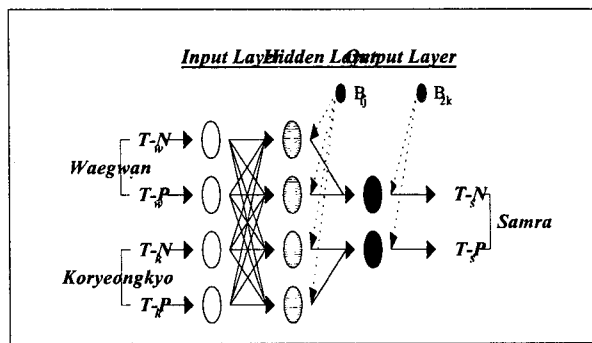


그림 3. 수질예측 신경망모형의 구조

6. 수질오염예측 신경망모형의 훈련과 검증

6.1 수질오염예측 신경망모형의 훈련

수질오염예측 신경망모형의 훈련을 위하여 훈련임계치(Training Tolerance)의 허용한계를 0.005로 설정하였다. 또한 무작위 발생된 난수의 수는 200, 500 및 1,000개로서 각각에 대하여 수질오염예측 신경망모형의 훈련을 실시하였으며, 훈련결과 무작위 발생된 난수의 수가 많을수록 즉 훈련자료의 수가 많을수록 훈련결과가 양호하고 훈련에 걸리는 시간이 단축되는 것으로 나타났다. 다음 그림 4는 훈련에 사용된 난수의 수가 200개일 경우에 OSSBP 및 RBP 수질오염예측 신경망모형의 훈련에 있어서 반복횟수에 따른 평균제곱오차의 수렴과정을 나타낸 것이다. 그리고 OSSBP 및 RBP 수질오염예측 신경망모형의 훈련에 의해 산정된 최적 연결강도와 편차를 수질오염예측 신경망모형의 검증에 이용하였다.

6.2 수질오염예측 신경망모형의 검증

수질오염예측 신경망모형의 검증은 무작위 발생된 난수의 수에 따른 훈련을 통하여 산정된 매개변수인 최적연결강도와 편차를 이용하였다. 즉 실제 관측자료를 입력자료로 하여 삼랑진지점의 T-N 및 T-P 수질오염예측을 실시하였다. 그 결과 삼랑진지점의 T-N 및 T-P의 수질오염예측은 OSSBP-WQNNM과 RBP-WQNNM에서 양호한 결과를 유도하였다.

7. 입력자료에 따른 수질오염예측의 불확실성 분석

본 연구의 낙동강 주요 수질관측지점인 삼랑진지점에서 수질오염예측을 위하여 수질오염예측 신경망모형의 훈련자료를 MCSM에 의해 200, 500 및 1,000개씩의 난수를 무작위 발생시켰다. 무작위 발생된 난수 그 자체가 자료의 불확실성을 포함하고 있으며, 무작위 발생된 난수의 수도 자료의 불확실성을 포함하고 있다. 따라서 불확실한 입력자료는 일반적으로 수질모형화 과정의 많은 변수들 중에서 주요 부분을 차지하고 있으며, 본 연구에서도 입력자료에 많은 불확실성을 내포하고 있는 실정이다. 그러나 본 연구에서는 수질오염인자의 확률분포형을 정규분포로 가정하고 MCSM에 의해 난수를 무

작위 발생시켰으며, 발생된 난수의 수가 많을수록 훈련결과가 양호한 것으로 분석되었으며, 검증결과도 양호한 것으로 나타났다. 따라서 가정한 확률분포형인 정규분포와 그에 따른 무작위 발생된 난수의 수는 타당한 것으로 사료된다.

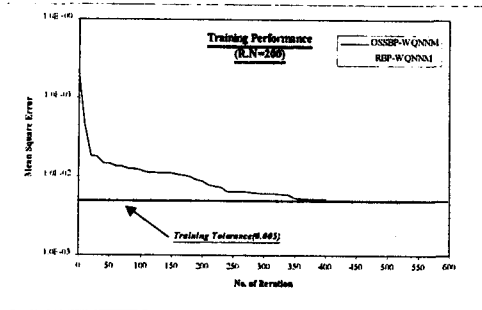


그림 4. 수질오염예측 신경망모형의 오차수렴도(R.N=200)

8. 결론 및 추후 연구과제

본 연구에서는 하천 및 호소에서 부영영화의 원인이 되는 수질오염인자인 낙동강유역 삼랑진지점의 T-N과 T-P 수질 오염인자의 월별 수질오염예측을 위하여 수질오염예측 신경망모형을 구성하였다. 수질오염예측 신경망모형의 훈련자료는 정규분포를 가지는 난수를 200, 500 및 1,000개씩 발생시켜 훈련자료의 발생조건에 따른 하천수질오염의 불확실성을 해석하였다. 훈련결과 생성된 난수의 수가 많을수록 즉 훈련자료의 수가 많을수록 훈련결과가 양호하고 훈련에 걸리는 시간이 단축되는 것으로 나타났으며, 검증결과도 유사한 결과를 나타내었다. 그러나 수질모형화과정이 많은 불확실성 인자를 포함하고 있으므로 입력자료의 불확실성을 제거하기 위한 결정론적 모형(Deterministic Model)과 MCSM을 포함한 신경망모형을 연계한 지속적인 연구가 수행되어야 할 것이다.

참고문헌

1. 김성원, 이순탁, 조정석(2001). "중소하천유역에서 Hybrid Neural Networks에 의한 수문학적 예측.", **한국수자원학회 논문집**, 한국수자원학회, 제 34권, 제 4호, pp. 303-316.
2. 낙동강유역환경청 홈페이지, <http://ndg.me.go.kr/>
3. 대구지방환경청 홈페이지, <http://daegu.me.go.kr/index.htm>
4. 안상진, 연인성, 한양수, 이재경(2001). "신경망모형을 적용한 금강 공주지점의 수질예측.", **한국수자원학회 논문집**, 한국수자원학회, 제 34권, 제 6호, pp. 701-711.
5. 조현경(2000). "인공신경망기법을 이용한 하천수질인자의 예측모델링 -BOD와 DO를 중심으로-.", **한국환경과학회지**, 한국환경과학회, 제 9권, 제 4호, pp. 455-462.
6. Ayyub, B.M., and Gupta, M.M.(1997). *Uncertainty analysis in engineering and sciences : Fuzzy logic, statistics and neural network approach*, Kluwer Academic, Boston.
7. Battiti, R.(1992). "First and second order methods for learning : between steepest and descent Newton's method." *Neural Computation*, Vol. 4, No. 2, pp. 141-166.
8. Broomhead, D.S. and Lowe, D.(1988). "Multivariable functional interpolation and adaptive networks." *Complex Systems*, Vol. 2, pp. 321-355.
9. Duke, Z.D., Rong, Y., and Harmon, T.C.(1998). "Parameter-induced uncertainty in modeling vadose zone transport of VOCs." *J. of Environ. Eng.*, Vol. 124, No. 5, pp. 441-448.
10. Lung, W.S.(1998). "Trends in BOD/DO modeling for wasteload allocations." *J. of Environ. Eng.*, Vol. 124, No. 10, pp. 1004-1007.
11. Riedmiller, M., and Braun, M.(1993). "A direct adaptive method for faster backpropagation learning : The RPROP algorithm." *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco.