

# 신경망 모형을 이용한 상수관로계통에서의 수질예측

## Prediction of Water Quality in Water Supply Systems Using Artificial Neural Network

○김주환\*, 윤재홍\*\*

### 1. 서 론

먹는 물의 수질악화 요인으로는 정수처리 과정에서 생성되기도 하지만 공급계통의 노후화로 인한 질적인 저하와 저수조의 관리부실 등이 큰 요인으로 작용하기도 한다. 지금까지 정수장에서 처리과정에서 발생하는 미량의 유기 및 무기 오염물질에 대한 연구는 활발하게 진행되어 왔으나 수도물의 공급계통, 즉 관로를 통한 공급과정에서 발생하는 문제점에 대해서는 아직 기초적인 수준에 머물고 있다. 정수장에서 소독과정에서 투여되고 있는 염소는 상수관로를 통해 흐르면서 물속의 용존유기물과 반응함으로써 소독부산물을 발생시키는데 대표적인 것이 염소계의 소독부산물이다. 일반적으로 물속의 용존유기물은 원수나 처리공정의 특성에 따라 정도의 차이는 있지만 모든 수중에 존재하고 있으며, 대부분의 국내 정수장에서 염소투입이 이루어지고 있는 만큼 트리할로메탄(THMs)의 생성은 필수적으로 일어나게 된다. 그러나 트리할로메탄의 생성여부 및 감시는 분석방법 등의 복잡성 때문에 효율적인 감시방법의 적용이 곤란한 경우가 대부분이며, 측정이 가능한 지점에서의 자료를 근거로 모형식을 개발하여 농도를 추정하는 경우가 주를 이루고 있다. 현재 국내에서의 트리할로메탄의 생성은 정수장내에서 처리공정별로 이루어지고 있으며 관망에서의 증가분은 고려하지 않는 경우가 대부분이다. 특히 관망에서는 시료채취의 대표성과 유달시간 등 수리적 요소들의 불확실성 때문에 관망에서의 트리할로메탄 농도를 정확히 알아낸다는 것은 현실적으로 불가능한 일이기 때문에 관망에서의 트리할로메탄 추정을 위한 모형의 역할은 더욱 중요하다 할 수 있다.

본 연구에서는 실제 수질변화 현상의 추정이 어려운 관로내 트리할로메탄의 생성을 추정하기 위하여 Black Box 모형의 일종인 신경망 모형을 적용하였다. 또한 이를 평가하기 위하여 미국수도협회(AWWA)로 부터 제시된 모형을 근거로 대상관망으로부터 측정된 수질자료의 통계적 분석에 의하여 선형 및 지수형 중회귀모형을 제시하였고 그 결과를 신경망 모형과 비교함으로써 신경망의 관내 수질예측 수단으로의 적용성을 평가하였다.

### 2. 모형의 구성

트리할로메탄 생성에 미치는 인자들로는 pH, TOC, UV254흡광물질, 수온, 초기염소농도, Br이온농도 및 접촉시간 등이 있으며, 이들 인자들은 트리할로메탄 생성에 비례의 관계를 갖는 것이 일반적이다. 그러나 트리할로메탄의 생성은 모든 수질인자들의 복합적인 조건에 의해 결정되며 우점 및 비우점 인자들의 작용에 따라 상관관계를 달리하는 경우도 보고되고 있다.

---

\* 한국수자원공사 수자원연구소 선임연구원  
\*\* 한국수자원공사 수자원연구소 수도연구부장

트리할로메탄 생성에 직접 영향을 미치는 인자는 트리할로메탄 전구물질의 농도와 초기염소농도이다. 특히 전구물질의 농도는 그것을 정량화하는데 어려움이 있으며, 따라서 유기물질의 농도로 대변하게 된다. 이때 사용되는 트리할로메탄 생성 전구물질의 지표인자로써는 TOC, DOC 및 UV254 흡광물질의 농도가 주로 사용되어 왔으며, 최근에는 TOX를 전구물질의 지표인자로 사용할 수 있는지의 상관성 여부가 연구 중에 있다. 현재까지 발표된 트리할로메탄 생성농도 추정식들은 앞의 인자들을 근거로 가감이 이루어지며, 현장조건에 맞게 각 인자별로 적당한 지수값 및 계수값 등이 산정 된다. 현재 사용되고 있는 트리할로메탄 농도 추정 모형중에서 대표적인 것으로 식(1)의 미국수도협회가 제시한 모형과 Urano에 의해 제시된 식(2), 그리고 식(3)의 Engerholm-Amy에 의해 제시된 모형 등이 있다.

$$DBP = k \times TOC^a \times pH^b \times T^c \times (Cl_2)_0^d \times Br^e \times UV254^f \times t^g \quad (1)$$

$$TTHM = k \times (pH - 2.8) \times TOC^l \times (Cl_2)_0^m \times t^n \quad (2)$$

$$TTHM = k \times (UV254 \times TOC)^q \times (Cl_2)_0^m \times t^n \times T^p \times (pH - 2.6)^r \times (Br + 1)^x \quad (3)$$

위 식에서 TOC는 총유기탄소농도(mg/L), pH는 수소이온농도, T는 수온(°C),  $(Cl_2)_0$ 는 초기염소농도(mg/L), Br은 브롬이온농도(mg/L), UV254는 UV254 흡광물질 농도이며 t는 반응시간(hr)을 나타낸다. 또한 k, a, b, c, d, e, f, g, l, m, n, p, q, r, x는 모형의 계수를 의미한다.

본 연구에서는 식(1), (2) 및 (3)의 모형 중에서 미국수도협회에서 제시한 모형을 도입하여 트리할로메탄 생성 농도를 추정하고자 하였으며, 이를 위한 방법으로서 통계학적 분석을 통한 중회귀 모형과 신경망 이론을 적용한 모형을 선정하였다.

## 2.1 중회귀 모형 구성

일반적으로 대부분의 자연적, 사회적인 현상을 설명하는데 있어서 종속변수 y의 변화는 2개 이상의 독립변수에 의해 좌우된다. 일반적으로 변수들간의 관계를 보다 효과적으로 설명하기 위해서 수 개의 독립변수를 사용하여 종속변수와와의 관계를 선형으로 가정하는 회귀모형을 선형중회귀모형(Multiple Linear Regression Model)이라 하고 수식으로 표현하면 다음과 같다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (4)$$

또한 지수형 중회귀모형은 다음 식 (5)과 같으며 이를 선형중회귀모형식의 형태로 변환하여 모형의 매개변수를 추정한다.

$$y = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \dots \cdot x_n^{\beta_n} \quad (5)$$

여기서,  $\beta_0, \beta_1, \dots, \beta_n$ 는 매개변수,  $x_1, x_2, \dots, x_n$ 는 독립변수, y는 종속변수를 나타낸다. 자료로부터 추정된 모형식이 그 특성자료에 어느 정도 적합한가를 측정하는 척도로는 결정변수(Determination Coefficient)를 사용하는데, 이는 총 변동중에서 추정된 모형식이 처리하는 비율을 말하며 다음 식 (6)으로 표현된다.

$$R^2 = \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \bar{x}_i)^2} \quad (6)$$

위 식에서,  $x_i$ 는 자료값,  $\hat{x}_i$ 는 모형에 의한 추정치,  $\bar{x}_i$ 는 자료집단의 평균치이며 N은 자료수를 나타낸다. 결정계수는 보통 0에서 1사이 범위의 값을 가지며 1에 가까울수록 추정된 모형에 대한 그 적합도 및 신뢰도가 높은 것으로 평가될 수 있다.

본 연구에서 트리할로메탄 생성에 따른 농도를 추정을 위해 도입한 미국수도협회 모형을 선형

및 지수형 중회귀식의 형태로 변환하여 다음 식(7) 및 식(8)과 같이 선정하였다.

$$\text{Model I} \quad DBP = k \cdot T^a \cdot pH^b \cdot TOC^c \cdot (Cl_2)_0^d \cdot Br^e \cdot UV_{254}^f \cdot t^g \quad (7)$$

$$\text{Model II} \quad DBP = k' + l \times T + m \times pH + n \times TOC + o \times (Cl_2)_0 + p \times (UV_{254}) + q \times t \quad (8)$$

## 2.2 신경망 모형 구성

신경망은 인간의 두뇌를 모사한 수학적 모형으로서 가장 폭 넓게 사용되어온 것으로는 Rumelhart 등(1986)에 의해 소개된 다층퍼셉트론 신경망이며, 이를 이용한 모형의 개발과정은 특별한 구조나 매개변수의 산정 및 자료의 변화 등이 필요치 않고 자료의 축적에 따라 모형의 능력을 향상시킬 수 있는 장점을 가지고 있다. 상수관망에서 트리할로메탄 농도를 추정하기 위한 신경회로망 모형은 관로의 각 지점에서 관측된 수질인자로 구성된 입력뉴런과 트리할로메탄 농도를 추정하기 위한 출력뉴런들로 구성된다. 이는 신경회로망의 다층퍼셉트론으로서 여러 개의 처리소자를 각 계층, 즉 입력자료를 받아들이는 입력층(Input Layer), 결과를 나타내는 출력층(Output Layer) 그리고 두 개의 층 사이에 중간층(Hidden Layer)을 두어 각 패턴마다 필요한 정보를 학습시킬 수 있도록 한 것이다.

신경망을 이용한 모형의 구성을 위해서는 먼저 적용하고자 하는 문제를 설정하여야 하는데 여기에서는 얻고자 하는 입력과 출력자료의 패턴을 결정하여야 하며 입출력 패턴에 따라 신경망의 구조를 선택한 후 전처리 과정을 거친다. 다음 단계에서는 입출력 패턴을 사용하여 신경망을 학습시키는데, 학습은 신경망 이론에서 사용되는 고유용어로서 일반적인 의미에서 매개변수의 산정절차와 유사하다. 학습이 완료되어 신경망의 매개변수인 가중치들이 모두 적절하게 산정된 후에는 학습과정에서 사용되지 않았던 측정자료를 이용하여 선택된 모형의 수행능력 및 적합성을 검증한다. 학습의 목적은 학습시 사용되지 않았던 자료에도 선택된 신경망 모형이 일반화될 수 있도록 하는 것이므로 모형의 학습과 검증과정은 동시에 수행되어야 하며 이 과정을 통하여 적정한 신경망 모형의 구조 및 가중치가 결정되어야 한다

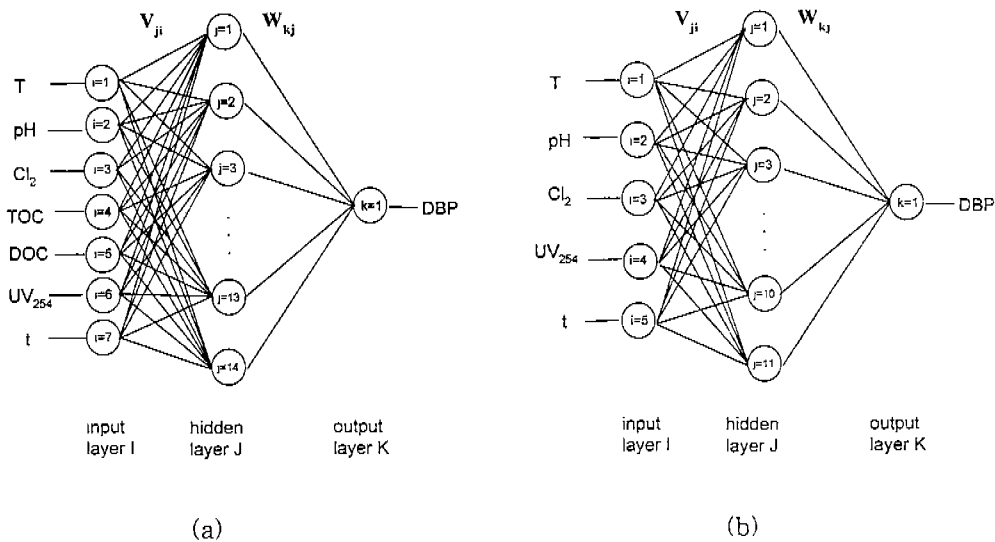


Fig. 1 Neural network architecture of Model III(a) and IV(b)

본 연구에서는 상수관로내에서의 트리할로메탄 생성에 영향을 미치는 인자들로서 pH, TOC, UV254 흡광물질, 수온, 초기염소농도, Br이온농도 및 접촉시간 등을 고려하여 신경망 모형을 다음 식(9)와 (10)과 같이 선정하여 구성하였으며 Fig. 1과 같다.

$$\text{Model III } DBP = ANN [ T, pH, Cl_2, TOC, DOC, UV254, t ] \quad (9)$$

$$\text{Model IV } DBP = ANN [ T, pH, Cl_2, UV254, t ] \quad (10)$$

### 3. 적용 및 결과분석

상수관로계통에서 트리할로메탄 농도분포 및 추정은 D-광역상수도 상수관로를 대상으로 총 13회에 걸쳐 측정된 수질자료를 사용하였다. 관로내에서 소독제와 전구물질이 접촉, 반응하는 유달시간은 트리할로메탄 생성에 있어 중요한 인자중 하나로서 관측지점별 유달시간은 수리학적으로 계산하였으며 관로시스템의 수리학적 특성은 KYPIPE모형을 이용하여 파악하였다.

본 연구의 대상관로로부터 측정된 관련 수질인자들을 대상으로 분석한 결과 트리할로메탄 생성농도 추정을 위한 지수형 및 선형 중회귀모형식은 각각 다음 식(11) 및 식(12)와 같이 나타났다.

$$TTHM = 6.188 \times T^{0.4794} \times pH^{-2.869} \times Cl_2^{0.5107} \times UV254^{0.3598} \times t^{0.06916} \quad (11)$$

$$TTHM = 0.05867 + 0.000687T - 0.009643pH + 0.01095Cl_2 + 0.4876UV254 + 0.0002916t \quad (12)$$

본 연구에서 제시한 트리할로메탄 생성농도 추정모형에 대한 분석결과 지수형 및 선형 중회귀모형식의 결정계수값은 각각 0.841 및 0.878로서 선형 중회귀식에 의한 추정식이 더 양호한 것으로 나타났으며 신경망에 의한 트리할로메탄 농도추정 결과에서는 유기물 인자를 모두 고려한 경우의 결정계수값이 0.972로 나타난 반면, UV254 흡광물질의 농도만을 고려한 경우의 결정계수값은 0.955로 다소 낮게 나타났으며 이 결과를 요약한 것이 Table 1이다. Table 2.에서는 본 연구에서 트리할로메탄 생성농도를 추정하기 위해 선정한 4가지의 모형에 대하여 계산된 추정치 및 관측치간의 상관관계를 요약한 것으로 신경망 이론을 적용하여 입력변수를 T(수온), pH,  $Cl_2$ , TOC, DOC, UV254, t(접촉시간)로 설정한 Model III에 의한 결과가 가장 양호한 것으로 나타났다.

Table 1. Correlation coefficient and determination coefficient for each models

index	Model I	Model II	Model III	Model IV
Correlation Coefficient	0.917	0.937	0.986	0.977
Determination Coefficient	0.841	0.878	0.972	0.955
Sample Size	81	81	90	90

Model I 및 Model II에 의해 추정된 결과와 측정값간의 상관성을 알아보기 위하여 Fig. 2 및 Fig. 3과 같이 도시하였다. 여기서 점선의 구간은 평균에 대한 95%신뢰구간을 의미하며 실선구간은 추정치에 대한 95%신뢰구간을 나타낸다. 또한 Model III와 Model IV에 의해 얻은 결과와 측정값과의 관계는 Fig. 4 및 Fig. 5에 도시하였다. 신경망을 이용한 Model III와 Model IV의 결과에서 트리할로메탄의 생성은 전구물질의 농도로서 UV254 흡광물질의 농도가 비교적 잘 반영한다고 볼 수 있으나 그 외의 유기물질에서도 트리할로메탄의 생성이 발생할 수 있으며, 현장의 실제조건에서 트리할로메탄 생성의 추정은 이들을 가능한 한 모두 고려할 경우만이 가장 실제와 근접한 추정식을 도출할 수 있음을 알 수 있다.

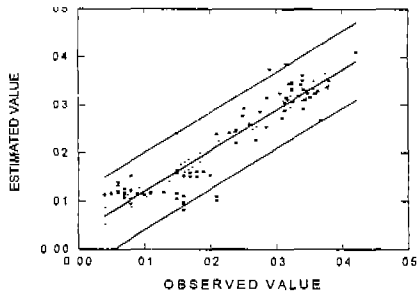


Fig. 2 Result (Model I)

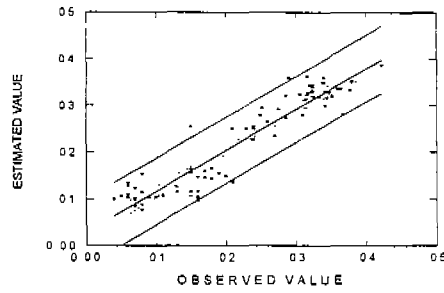


Fig. 3 Result (Model II)

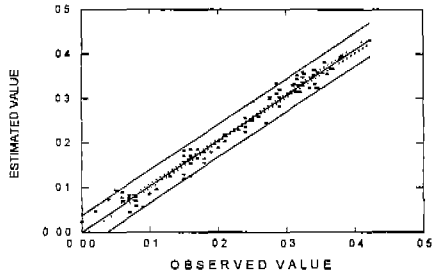


Fig. 4 Result (Model III)

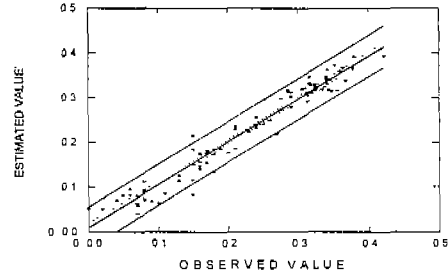


Fig. 5 Result (Model IV)

#### 4. 결 론

본 연구에서는 상수관로내에서 수질변화 및 거동에 대한 물리·화학적 접근이 어려운 경우, 모형화 방법으로 사용되는 신경망 모형을 적용해 상수관로내 트리할로메탄 생성에 따른 농도추정 과정을 모형화 하였으며 그 적용성을 평가하기 위하여 미국수도협회에서 제시한 모형을 기초로 구성된 중회귀모형에 의한 결과와 비교하였다. 여기에서는 신경망 모형에 의해 추정된 결과가 중회귀모형으로부터 얻은 결과보다 양호한 것으로 나타남을 알 수 있었는데 이는 신경망 모형은 자료간의 비선형 관계를 표현할 수 있는 비선형함수를 내포하고 있으며 자료의 입출력 관계를 패턴으로 인식함으로써 자료간의 집단적 특성을 분류해 낼 수 있는 반면, 중회귀 모형은 자료의 선형적 관계를 단순히 수학적으로 나타내는 것이기 때문이다. 또한 대상관로에 대하여 미국수도협회에서 제시한 지수모형에 의한 트리할로메탄 추정결과 보다는 선형 중회귀모형에 의한 결과가 더 높은 상관성을 보여주었다는 사실에 주목할만 하다.

신경망 모형의 적용결과에서는 모형구성시 트리할로메탄 전구물질을 가능한 한 모두 포함시키는 것이 효과적인 것으로 나타났으며 이를 고려한 Model III가 트리할로메탄 생성특성에 따른 수질패턴을 가장 잘 반영하는 것으로 나타났다.

본 연구에서 도입, 적용된 신경망 모형은 어떤 과정이나 현상에 대한 재현성이 뛰어나며 상수관로계통에서 모형화하기 어려운 수질예측을 위한 수단으로 적용될 수 있음을 알 수 있었으며 트리할로메탄 외 다른 수질인자에 대해서도 적용이 가능할 것으로 판단되어 이에 대한 연구가 기대된다.

## 참 고 문 헌

1. 강용태 등, "급배수시스템에서의 수질변화", 한국수처리기술연구회, 제5권 제2호, pp.3-15, (1997)
2. 임정아 등, "THM 형성인자에 의한 THM 농도예측 및 고도정수처리 공정에서의 적용", 대한상하수도학회지 제11권 제4호, pp. 126-132, (1997)
3. 오정우 등, "상수도 관로내에서의 THMs 생성 특성", 한국물환경학회지 제14권 제2호, pp.223-228, (1998)
4. 한국수자원공사, "G-7 수돗물의 2차오염 방지기술", 1997.
5. 김주환 (1993) "신경회로망을 이용한 하천유출량의 수문학적 예측에 관한 연구." 박사학위논문, 인하대학교.
6. K. Urano et al, "Empirical rate Equation for Trihalomethans Formation with Chlorination of Humic Substances in Water", Wat. Res., Vol 12, 1983
7. "Mathematical Modeling of the Formation of THMs and HAAs in Chlorinated Natural Water", AWWA, 1993
8. Lee, B.H., "Locating Monitoring Station in Water Distribution Networks", A Doctorial Dissertation, The Univ. of Michigan, 1990.
9. Bryant, E.A. and Fulton, G.P. and Budd, G.C., "Disinfection Alternatives for Safe Drinking Water", Van Nostrand Reinhold, New York, 1992.
10. AWWA, "Disinfection/Disinfection By-Products Database and Model Projects," 1993.
11. Singer, P. C., Barry, J. J. III., Palen, G. M. and Scriver, A. E., "Trihalomethane Formation in North Carolina Drinking Waters", AWWA, Aug., 1981
12. Wasserman P.D. *Neural Computing Theory and Practice*, Van Nostrand Reinhold, 1989.
13. Lippman R.P. "An Introduction to Computing with Neural Net." IEEE, ASSP Magazine, April, 1987.
14. Reckhow, D.A., Singer, P.C., and Malcom R.L., "Chlorination of Humic Materials: Byproduct Formation and Chemical Interpretation. *Envir. Sci. & Technol.*, 24:11, 1990.