

지하수의 질산성질소($\text{NO}_3\text{-N}$) 추세 예측을 위한 MLP 모델의 적용성 검토

김규범^{1,†} · 황찬익²

¹대전대학교 건설안전공학과

²대전대학교 산학협력단

Application of the MLP model for a trend prediction of $\text{NO}_3\text{-N}$ in groundwater

Gyoo-Bum Kim^{1,†} · Chan-Ik Hwang²

¹Department of Construction Safety and Disaster Prevention, Daejeon University, Daejeon 34520, Republic of Korea

²Industry-Academic Cooperation Foundation, Daejeon University, Daejeon 34520, Republic of Korea

요 약

최근들어 국내 농촌지역의 대표 오염 항목으로 대두되는 지하수내 $\text{NO}_3\text{-N}$ 농도에 대한 체계적인 관리가 요구되는 실정이다. 국가지하수측정망에서 5년 이상 분석된 $\text{NO}_3\text{-N}$ 농도 시계열 자료를 활용하여 추세분석을 실시한 결과 전체의 13.2%가 상승 추세를 보이는 것으로 분석되었다. 본 연구에서는 GIS 기법을 활용하여 추출한 각 지하수 측정망 지점의 지형, 수문지질, 토양, 관정밀도 및 지표피복 등의 특성 DB를 입력 변수로, 측정망의 $\text{NO}_3\text{-N}$ 상승 추세 여부를 출력 변수로 설정하여 91% 이상의 예측 정확도를 보이는 다층퍼셉트론(Multilayer perceptron) 모델을 도출하였다. k-분할 교차검증(k=5) 결과, 정확도, 정밀도, 재현율 및 F1 Score 등이 대부분이 0.9 이상을 보여 양호한 것으로 평가되었다. 목포-무안 지역에 본 모델을 적용한 결과, 각 그리드에서 $\text{NO}_3\text{-N}$ 의 실제 상승 추세를 예측 모델이 상승으로 예측하는 경우가 비상승으로 예측하는 경우보다 예측 확률값이 높은 것으로 나타나 유의미한 결과가 도출되었다. 한편, 본 $\text{NO}_3\text{-N}$ 의 상승 추세 예측을 위한 MLP 모델에서는 오염원의 종류 및 농도, 오염원과의 거리 등이 입력 인자로 고려되지 못하였으나, 향후 이들 데이터의 수집을 통하여 보다 정확한 예측 모델이 개발될 수 있으며, 현재 농도 수준 뿐 아니라 미래의 변화 가능성까지 고려한 새로운 지하수 수질 오염 관리 체계가 마련될 수 있을 것으로 판단된다.

주요어: 지하수, 질산성질소, 추세분석, 농도 예측, 다층퍼셉트론

ABSTRACT: Systematic management of groundwater is required due to high concentration of $\text{NO}_3\text{-N}$ in groundwater, Korea. The trend analysis for the time series of $\text{NO}_3\text{-N}$ concentration, which have been analyzed at least over 5 years at the national groundwater monitoring wells, indicates that 13.2% of total wells has shown an upward trend. Characteristic information, which is related to topography, soil, hydrogeology, well density, and land cover, is extracted by using GIS technology. An MLP (multilayer perceptron) model, which prediction accuracy is about over 91%, is developed using the information as input variable and $\text{NO}_3\text{-N}$ trend as output variable. As a result of k-fold validation (k=5), accuracy, precision, recall, and F1 score show mostly higher than 0.9, and the model is evaluated as good. The MLP model application to Mokpo-Muan region shows that it makes a little sense, because the average value of predicted probability in the cases in which the actual upward trend is predicted as the upward trend by the MLP model is higher than the value in the cases in which it is predicted as the non-upward trend. On the other hand, the type and concentration of the pollutant and the distance to the pollution source are not considered as input variables in this model, but a more accurate prediction model can be developed when collecting these data in the future. Consequently, a new groundwater management system considering not only the present concentration but also future changes will be prepared.

Key words: groundwater, $\text{NO}_3\text{-N}$, trend analysis, concentration prediction, MLP

[†]Corresponding author: +82-42-280-4574, E-mail: geowater@dju.kr

1. 서 언

지하수 수질의 변동성은 오염원의 증감, 오염배출량의 증감, 오염 경로의 변경 등 다양한 요인에 의하여 발생되게 된다. 특히, 염소(CI)와 질산성질소($\text{NO}_3\text{-N}$)는 흔히 발생하는 대표적인 지하수 수질 오염 항목으로서 농도 값 뿐 아니라 추세성도 중요한 관리 인자이므로 국가 지하수 측정망에서 지속적으로 관측되어 왔다.

국내에서 지하수 수질 분포와 추세성에 대한 연구는 다양하게 수행된 바 있다. 다수의 수질 분석 자료를 보유하고 있는 제주도를 대상으로 지하수내 염소(CI)와 질산성질소($\text{NO}_3\text{-N}$)의 광역적 추세를 Sen의 추세 분석 기법과 GIS를 융합하여 분석하였으며, 중산간 지역의 수질 악화 추세를 진단한 바 있다(Kim *et al.*, 2007). 한편, 내륙에서의 수질 추세성에 대한 연구는, 국내의 지하수 오염우려가 높은 781개 지점에 설치된 수질측정망의 연 2회 수질 분석 자료를 이용하여 이들 지점의 수질 농도가 비오염지역보다 높음을 분석한 바 있다(Kim *et al.*, 2010). 또한, 지하수 수질 관리를 보다 효율적으로 수행하기 위한 수단으로서 전국의 지하수 수질 측정망의 분석 자료의 수질 문턱값(Threshold value)과 Sen의 추세분석 결과를 활용하는 방안을 제시하기도 하였다(An *et al.*, 2015). 최근에는, 제주도에서 1993년부터 2015년까지 측정된 21,568건의 지하수 수질 자료 중 질산성질소를 대상으로 통계 특성과 Mann-Kendall 및 Sen의 분석 방법을 적용하여 공간상의 추세 분포를 평가한 바 있으며(Kim *et al.*, 2018), 경기도에서 운영하는 지역지하수측정망의 수질 분석 자료를 토대로 평균농도, 추세성, 초과횟수 등의 통계량을 이용하여 도내 수질의 분포 특성을 분석하고 수질 등급을 분류한 바 있다(Seo *et al.*, 2020). 이와 같이, 지하수 수질에 대한 연구는 농도의 통계적 특성을 토대로 공간적 분포를 파악하거나, 항목별 추세성을 토대로 오염원 관리의 필요성 등을 제시하는 등에 치중되어 왔다. 그러나, 기존의 통계적인 접근법이나 추세성 분석 방법은 정기적인 수질 분석이 이루어지는 계측 지점에 대하여 가능한 해석법으로서, 미계측 지역에 대한 수질 농도의 예측이나 추세성의 진단이 이루어진다면 보다 체계적인 수질관리가 가능할 것이다.

최근 해외에서는 Big data 기반의 인공지능망 기

법과 수질 예측을 접목하는 다양한 연구들이 수행되고 있다. 하천수의 수질을 예측하기 위하여 지역의 강우량, 하천 유량, 탁도 등을 입력 인자로 사용하는 단층(Single hidden-layer) 인공지능망을 개발한 바 있으며(Bowers and Shedrow, 2000), 싱가포르 해안 지역 해수의 온도, 용존산소, Chlorophyll-a 등을 입력 인자로 사용하는 인공지능망 모델을 개발하여 해수의 수질을 예측하는 연구를 수행한 바 있다(Palani *et al.*, 2008). 체계적인 수질 관리를 위하여 다양한 머신러닝 기법(ANFIS, RBF-ANN, MLP-ANN, WDT-ANFIS 등)을 이용하여 예측의 정확도를 비교, 평가함으로써 하천수 수질 예측의 실효성을 높일 수 있는 모델들이 개발되기도 하였다(Ahmed *et al.*, 2019). 모로코에서는 농업용 지하수의 수질(TDS, Salinity, Sodium adsorption ratio 등)을 예측하기 위하여 전기전도도, 온도, pH 등을 입력 인자로 활용한 인공지능망 및 SVR (Support vector regression) 모델을 개발하여 적용성을 검토한 바 있다(Bilali *et al.*, 2021). 이와 같이, 최근에는 수질의 예측 분야 뿐 아니라 수위, 유량, 퇴적, 홍수 등의 여러 분야에서 인공지능망이 활용되고 있어 Big data 시대의 예측 기술이 다양화하고 있다(Minns and Hall, 2009; Mustafa *et al.*, 2012; Shiri *et al.*, 2013; Gholami *et al.*, 2016; Sakizadeh, 2016).

질산성질소($\text{NO}_3\text{-N}$)는 축사, 비료, 오수 등 다양한 오염원에 의하여 국내에서 광범위하게 나타나는 항목이므로, 지하수 측정망의 관측자료에만 의존하여 관리하는 것은 제한적이다. 임의의 미계측 지점에서의 지하수내 $\text{NO}_3\text{-N}$ 의 악화 추세성을 예측하는 모델이 개발되어 적용된다면, 이들 지역에 대하여 선제적인 오염원 관리를 수행함으로써 지하수 수질 악화의 가능성을 사전에 예방하는데 도움이 될 것이다. 본 연구에서는 국가 지하수 측정망 운영 과정에서 분석된 $\text{NO}_3\text{-N}$ 의 시계열 자료를 활용하여 추세성을 평가한 후, 각 측정망의 지점 특성과의 상관성에 대한 인공지능망 기법을 적용하여 추세 예측 모델을 개발하고 미계측 지역에 적용성을 평가하고자 하였다.

2. 연구 방법

2.1 선형 추세 분석

지하수내 $\text{NO}_3\text{-N}$ 농도의 추세성에 대한 평가는

선형회귀분석과 T-검정에 의하여 수행하였으며, 유의수준 0.05하에서 추세성을 판단하였다. 선형회귀 추세선의 일반식은 다음과 같이 표현된다.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

여기에서, β_0 는 원점에서의 값이며, β_1 은 경향의 크기(Trend magnitude), X 는 시간, ϵ_t 는 오차항으로서 평균이 0이고 분산 σ_ϵ^2 을 갖는 독립 확률변수(Independent random variables)이다. 위 식은 다음과 같이 전통적인 선형회귀모델(Linear regression model)에 의하여 표현될 수 있다.

$$h_t = b_0 + b_1 t$$

여기에서, b_0 와 b_1 은 β_0 및 β_1 의 추정치이다.

선형경향성을 인지하는 것은 경향성의 경사도에 대한 통계적 검정으로서, 가설 검정에서 다음 조건을 고려하게 된다.

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

유의수준이 $(1-\alpha)$ 라면, 이 검정의 통계량 t 는 다음과 같다.

$$t = \frac{|b_1|}{s_{b_1} / \sqrt{\sum_{t=1}^n t^2 - \frac{(\sum_{t=1}^n t)^2}{n}}} - t_{\alpha/2}$$

여기에서 n 은 표본의 크기이고, $t_{\alpha/2}$ 는 유의수준이 $\alpha/2$ 이고 $n-2$ 의 자유도를 갖는 Student T 분포의 분위수이다. s_{b_1} 는 위에서 정의한 자료의 표본 표준편차이다.

귀무가설 H_0 는 검정통계량 $t \leq 0$ 인 경우에 받아들여지게 되어 추세가 없다고 판단하며, $t > 0$ 인 경우에는 귀무가설을 기각하여 추세가 존재한다고 판단한다. 추세의 유형은 상승 추세, 하강 추세 및 무변동 등으로 구분이 가능하나, 실제 지하수 관리의 주 대상은 상승 추세이므로, NO₃-N 농도를 상승 추세와 비상승 추세로 이분화하여 사용하였다.

2.2 다층퍼셉트론(Multilayer perceptron, MLP) 모델

인공신경망 모델인 다층퍼셉트론은 은닉층이 1개인 모델로서, 비선형적으로 분리되는 자료에 대해서는 제대로 된 학습이 이루어지기 어렵다는 한계점이 존재하였다. 이를 극복하기 위하여 입력층과 출력층 사이에 하나 이상의 은닉층을 두어 비선형적으로 분리되는 자료에 대해서 학습이 가능하도록 한 것이 다층퍼셉트론(MLP)이다(Kim and Oh, 2018). MLP는 각 은닉층간의 뉴런의 연결강도를 최적의 상태로 적응하는 학습과정을 통하여 최적 모델이 구성된다.

출력 뉴런에 의하여 전달되는 은닉층의 출력은 아래 식과 같이 각 시냅스의 가중치(w_{ij})와 노드의 입력자료(x_j)에 의하여 결정된다.

$$t_j = g(\sum w_{ij} x_i)$$

또한, 출력층에서 출력값(y_k)은 시냅스의 가중치와 은닉층 및 출력층에서의 활성화 함수(g 및 \bar{g})에 의하여 결정된다.

$$y_k = \bar{g}\left(\sum_{j=0}^n w_{kj} \cdot g\left(\sum_{i=0}^m w_{ij} x_i\right)\right)$$

활성화 함수란 입력함수의 결과값을 특정 임계값과 비교하여 특정값으로 출력하는 것으로서, 본 연구의 은닉층에서는 입력함수 결과값이 이 임계값보다 크면 1, 그렇지 않으면 -1로 출력하는 방식의 하이퍼볼릭 탄젠트(tanh) 함수를 사용하였고, 최종 예측확률을 계산하기 위한 출력층에서는 소프트맥스(softmax) 활성화 함수를 사용하였다. 하이퍼볼릭 탄젠트(tanh) 및 소프트맥스(softmax) 활성화 함수는 각각 아래와 같이 표현된다.

$$\tanh(x_i) = \frac{\exp(x_i) - \exp(-x_i)}{\exp(x_i) + \exp(-x_i)}$$

$$y_i = \frac{\exp(x_i)}{\sum_{j=1}^k \exp(x_j)}$$

Table 1. Input and output variables of MLP model for a trend prediction of NO₃-N in groundwater (from Kim *et al.*, 2022).

	Input variables	Output variable
Continuous variables	· Elevation of land surface	Trends (upward and non-upward) of NO ₃ -N in groundwater
	· Highest elevation in 250 m radius	
	· Upward slope in 250 m radius	
	· Lowest elevation in 250 m radius	
	· Downward slope in 250 m radius	
	· Average slope	
	· Distance to double-lined stream	
	· Elevation of nearby double-lined stream	
Categorical variables	· Density of pumping wells	
	· Hydrogeologic unit (3 classes)	
	· Subsoil feature (7 classes)	
	· Top soil feature (7 classes)	
	· Soil drainage class (6 classes)	
	· Land use (7 classes)	

여기에서, y_i 는 예측 분류 결과의 확률을 의미하며, k 는 출력층의 뉴런의 수, x_i 는 이전 층의 뉴런에서 넘어오는 시그날의 가중합을 의미한다.

3. 데이터베이스 구축

본 연구는 NO₃-N 농도의 추세성을 판단하기 위한 모델을 개발하는 것이므로, 국가지하수측정망의 5년 이상 기간 동안 연 2회(일부는 4회) 측정된 NO₃-N 농도의 추세성(상승 또는 비상승(여기에서 비상승이란 하강 및 유지를 포함함))을 MLP 모델의 주요 예측 변수로 설정하였다. 한편, 각 성분 농도의 추세성을 예측하는 MLP 모델의 입력 변수는 지하수 측정망의 지점 특성 자료를 활용하였는데, GIS 기법을 활용하여 디지털 지도로부터 추출하였다. 입력 변수의 선정은 자료 취득의 용이성, 균질성, 정확성 및 정부 또는 공공기관에서 제공된 공인된 데이터 등의 조건을 고려하였다.

모델의 입력 변수는 연속형 변수와 범주형 변수를 모두 포함하고 있는데, 연속형 변수로는 지형도(1:50,000 축척) 및 DEM (90 m) 자료로부터 추출 가능한 지하수 흐름에 영향을 미칠 수 있는 지형 인자를 검토하였으며, 범주형 변수는 국내 발간된 다양한 특성 도면(토양도, 수문지질도, 지표피복도 등)

에서 추출하였다. 최종적으로 모델에 사용된 연속형 변수는 지형 표고 등 8개의 변수와 전국 약 160만 공의 지면 위치를 사용하여 작성한 관정밀도도로부터 추출된 관정밀도 등 총 9개로 결정되었다. 범주형 변수로는 각 주제도로부터 수문지질단위, 심토토성, 표토토성, 배수등급, 지표피복 등 5개의 속성을 사용하였다(표 1). 연속형 변수에서 각 속성의 추출은 관측 지점 주변 반경 250 m의 범위를 기준으로 이루어졌는데, 이 규모는 지형 변화 특성, 지하수의 유동성, 지하수 영향 반경 등을 고려하여 결정하였다.

총 521개 지하수 측정망 지점의 입력 변수에 대한 통계 특성은 아래와 같다(표 2). 연속형 변수를 보면, 지형 표고를 제외한 대부분의 지형 특성 인자는 국내 일반적인 지점에서 나타날 수 있는 지형 특성의 범위를 대부분 포함하고 있다. 다만, 지하수 측정망의 설치 표고가 최대 981.7 EL.m로서 고지대의 특성을 반영하기에는 부족해 보이므로 고지대 예측 시 유의할 필요가 있다.

각 범주형 변수의 속성을 지하수 유동 측면에서 1부터 7까지(수문지질단위는 1~3, 배수특성은 1~6) 재분류한 후 각 지하수측정망 지점의 해당 특성을 도출하였으며, 일부 속성을 제외하면 지하수측정망의 분포가 해당 속성의 실제 분포 비율과 비교할 때 큰 차이가 없이 비교적 골고루 분포하고 있는 것으

Table 2. Statistics of continuous variables at the groundwater monitoring sites.

Variables	Elevation (EL.m)	Highest elevation (EL.m)	Upward slope	Lowest elevation (EL.m)	Downward slope	Average slope	Distance to stream (m)	Elevation of nearby stream (EL.m)	Density of wells (n/km ²)
Min	1.2	2.1	0.0	0.3	0.0	0.0	1.7	1.0	0
Max	981.7	1,038.5	0.5	958.1	0.3	0.3	13,731.4	895.2	124
Mean	123.6	153.4	0.1	113.0	0.1	0.1	407.3	116.9	10

Table 3. Classification of groundwater monitoring site characteristics for each categorical variable.

Code	Top soil		Subsoil		Hydrogeologic unit		Soil drainage		Land use	
	Group	Count	Group	Count	Group	Count	Group	Count	Group	Count
1	Clay	0	Clay	50	Metamorphic rock	159	Very poor	13	Domestic; Commercial; Industrial; Transportation; Culture/athletic	210
2	Silty clay loam; Silty clay	29	Fine silty	22	Intrusive igneous rock; Sedimentary rock; Non-porous volcanic rock	285	Poor	7	Rice farming	44
3	Silt	0	Coarse silty	3	Unconsolidated soil; Limestone; Porous volcanic rock	77	Slightly poor	82	Grass and golf course	53
4	Silt loam; Sandy clay	66	Fine loamy	169	-	-	Slightly good	120	Dry farming; Orchards; Greenhouses	91
5	Loam; Sandy-clay loam	200	Coarse loamy	228	-	-	Good	231	Broad-leaved trees; Coniferous trees; Mixed forests	23
6	Sandy loam	218	Sandy	14	-	-	Very good	68	Unimproved land; Mining land	79
7	Sand; Loamy sand	8	Sandy skeletal	35	-	-	-	-	Wetlands; Surface water	21

로 나타났다(표 3).

4. 추세 분석 결과

T-검정에 의한 분석 결과 도출되는 추세성 유무와 추세성의 기울기(+ 및 -)를 토대로 상승과 비상승

으로 구분하였다. 최종적으로 지하수 측정망 452 개소(86.8%)가 비상승 추세, 69 개소(13.2%)가 상승 추세로 나타났다(그림 1).

NO₃-N의 상승과 비상승 추세별로 특성 인자들의 통계적인 차이는 유의하지 않으나, 일부 인자에서는 개략적인 차이가 보이기도 한다(표 4). 평균 통계량

을 보면, NO₃-N의 상승 추세로 나타난 지하수측정망은 비상승 추세인 측정망에 비하여 상류경사 및 평균경사가 급한 곳, 하천에서 먼 곳 등에서 주로 나타나는 것으로 보인다.

범주형 인자에 대하여 NO₃-N의 상승 추세와 비상승 추세의 통계적 차이를 비교해 보면, 4개의 특성(표토토성, 심토토성, 배수등급, 지표피복)에 대하여 상승 및 비상승 추세를 보이는 관측지점의 속성

의 분포 차이가 크지 않은 것으로 나타났다(그림 2). 이와 같은 결과를 보면, 지하수내 NO₃-N의 상승 추세를 보이는 것은 지점의 특성별로 미세한 차이는 존재하나 통계적으로 유의한 수준의 뚜렷한 차별성을 보이지 않는 것으로 평가되므로 각 지점의 특성과 성분의 추세성을 선형 관계에 의하여 설명하는 것은 불가능하다.

5. 추세예측을 위한 MLP 모델 개발

상기 지하수측정망 지점의 특성인 입력 변수와 지하수내 NO₃-N의 추세성 여부(상승 및 비상승 추세)에 대한 출력 변수간의 비선형 관계를 도출하기 위하여 IBM SPSS modeler (v.18.1)을 활용하여 MLP 모델을 구성하였다. MLP 모델의 특징으로는 사용되는 은닉층의 활성화 함수가 출력값의 경우 -1과 1 사이 값만을 반환하기 때문에 모델에 사용되는 입출력인자들의 다양한 변환 가정이 필요하게 된다. 본 연구에서는 입출력 인자들이 동일 표준 편차와 평균을 갖도록 표준화 변환을 실시한 후 모델의 변수로 활용하였다.

$$x_{ij}^k = \frac{X_{ij} - \bar{X}_{ij}}{\sigma}$$

여기에서, x_{ij}^k 는 표준화된 값, X_{ij} 는 변수 값, \bar{X}_{ij} 는 변수의 평균, σ 는 변수의 표준편차를 의미한다.

NO₃-N에 대한 MLP 모델의 은닉층은 반복 학습을 통하여 2개, 뉴론은 5 및 4개로 결정하였다. 최종 개발된 MLP 모델을 보면, 훈련자료와 검증자료에



Fig. 1. Distribution of NO₃-N trends at groundwater monitoring sites.

Table 4. Statistics of continuous variables for two types of NO₃-N trend at groundwater monitoring sites.

Trend	Stat.	Elevation (EL.m)	Highest elevation (EL.m)	Upward slope	Lowest elevation (EL.m)	Downward slope	Average slope	Dist. to stream (m)	Elevation of nearby stream (EL.m)	Density of wells (n/km ²)
Non-upward	n						452			
	Mean	123.2	146.1	0.091	112.0	0.045	0.068	301.9	113.8	11.8
	St.dev	153.7	164.5	0.086	149.0	0.062	0.068	441.4	142.6	11.2
Upward	n						69			
	Mean	123.7	154.5	0.123	113.2	0.042	0.083	423.4	117.3	9.6
	St.dev	134.5	146.9	0.098	130.6	0.044	0.062	1351.9	132.4	13.4

대한 추세성(비상승 및 상승) 예측의 정확도는 91.3% 및 91.8% 정도로 높게 나타났다(표 5).

도출된 모델의 적용성을 진단하기 위하여 k-분할 교차검증(k=5)을 통하여 모델의 정확도(Accuracy), 정밀도(Precision), 재현율(Recall) 및 F1 Score를 평가하였다. 여기에서,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Table 5. Accuracy of MLP models for the trend predictions of NO₃-N.

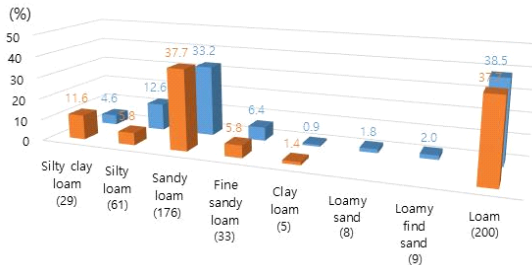
Decision	Training	Test
Correct	230 (91.3%)	247 (91.8%)
Wrong	22 (8.7%)	22 (8.2%)

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

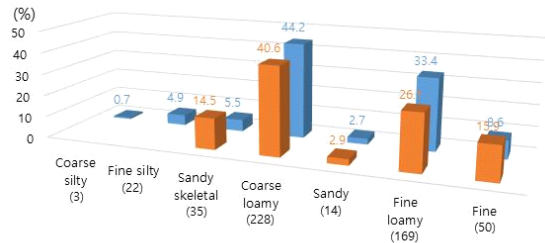
로 표현된다. 표에서 보는 바와 같이, 대부분의 모델에서 정확도, 정밀도 및 재현율이 최적 모델과 유사하며, 적용 가능한 수준으로 도출되었음을 알 수 있다(표 6).

6. 모델의 평가

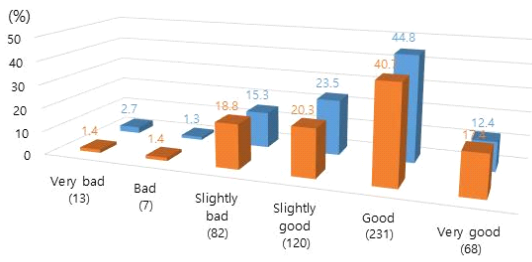
상기 NO₃-N의 추세성 예측 인공지능망 모델의 실제 적용성을 파악하기 위하여 목포-무안 지역을 대상으로 실제 추세와 모델 예측 추세를 비교해 보았다. 2009년부터 2018년까지 지하수 관정의 NO₃-N 수질 데이터를 수집하여 500 m 그리드내 관정의 수질 자료를 시계열로 배열한 후 선형회귀 및 T-검정에 의한 추세평가를 실시하였다. 관정 수질 자료 갯수는 제한적으로서 추세 분석이 가능한 그리드는 76개에 불과하였으며, 상승 기울기의 그리드는 38개, 하강 추세의 그리드는 38개로 나타났다(그림 3a, 3b).



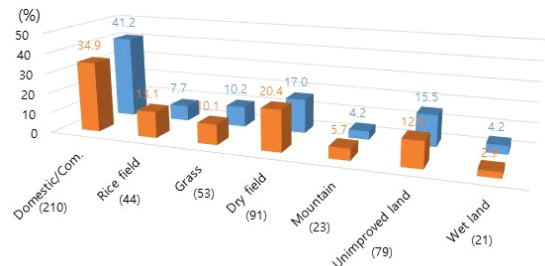
(a) Top soil



(b) Subsoil



(c) Soil drainage



(d) Land use type

Fig. 2. Percentage of groundwater monitoring wells according to the properties of the input variables for NO₃-N (red: upward trend, blue: non-upward trend, numbers in parentheses: number of monitoring wells).

Table 6. Results of 5-fold cross-validation and evaluation for the MLP model.

Data set	Training partition				Validation partition			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
1	89.4%	88.9%	98.8%	93.6%	94.2%	95.6%	97.7%	96.6%
2	86.1%	88.3%	95.2%	91.6%	84.6%	88.9%	93.0%	90.9%
3	90.9%	91.7%	97.6%	94.6%	86.5%	93.3%	91.3%	92.3%
4	92.8%	94.4%	97.1%	95.8%	84.6%	86.7%	95.1%	90.7%
5	89.9%	90.6%	97.6%	93.9%	90.4%	95.6%	93.5%	94.5%
Ave.	89.8%	90.8%	97.3%	94.1%	88.1%	92.0%	94.1%	94.1%

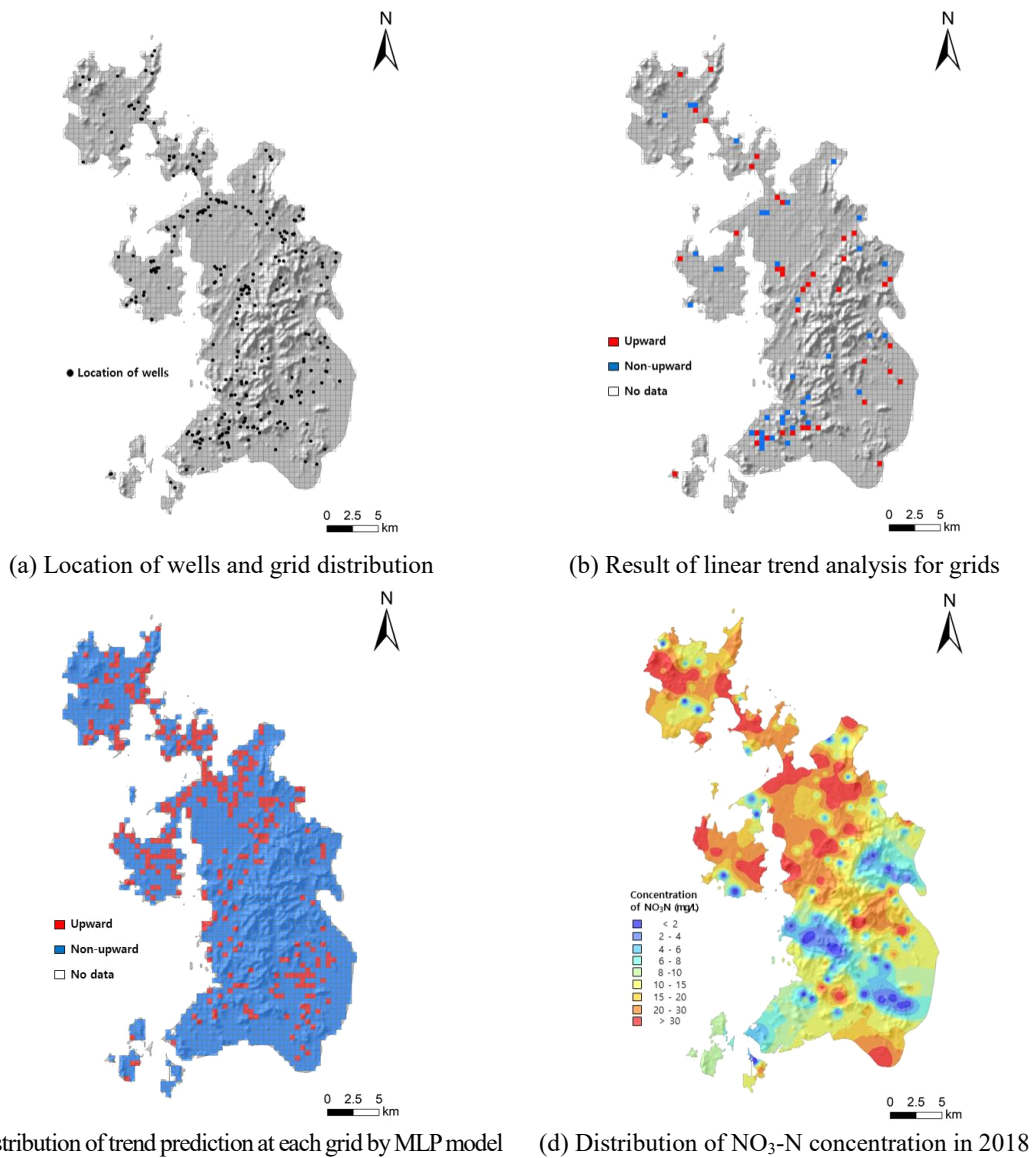


Fig. 3. Actual and predicted trends of NO₃-N in groundwater in Mokpo-Muan region.

Table 7. Comparison of average probability of trend prediction for MLP model between actual NO₃-N trend and non-trend at groundwater wells.

Actual trend	Trend by MLP model	
	Upward trend	Non-upward trend
Upward trend	0.933	0.768
Non-upward trend	0.883	0.899

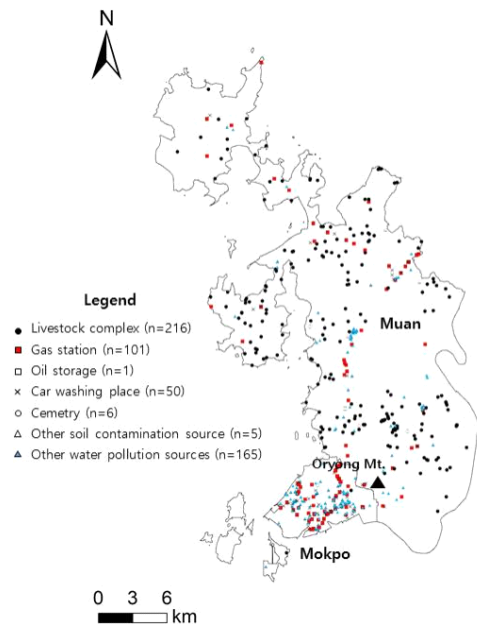
목포-무안 지역내 500 m 격자를 구성한 후 각 중심점에서 모델 입력 인자를 추출한 후 해당 지점의 NO₃-N의 추세를 예측하였다. GIS를 이용하여 전체 그리드 2016개에 대하여 14개의 입력변수 값을 추출하였으며 이들로부터 추정된 NO₃-N의 상승 예측 그리드는 345개로 나타났다. 북측의 신안군 및 함평군과 인접한 무안군의 북쪽 지역이 상승 추세 가능성이 높은 지역이며, 목포시 인근의 영산강 주변도 상승 추세 가능성이 높은 것으로 나타났다(그림 3c).

그림에서 보는 바와 같이, 관정내 NO₃-N의 실제 추세 분포와 예측 모델에 의한 추세의 분포가 유사한 경향을 보이고는 있다(그림 3b, 3c). 그러나, 이와 같이 실제 관정의 수질 자료가 추세를 보이는 그리드와 인공신경망 예측 모델에서 추세를 보이는 그리드의 수에 많은 차이가 존재하기 때문에 두 결과를 이용하여 모델의 예측 정확도를 평가하기에는 한계가 있어 보인다. 실제 관정의 수질 분석 자료의 개수(그리드 수)가 76개로 너무 적은 점, 관정의 수질 분석에서 불합격되는 경우는 재분석 등 추가 분석을 통하여 수질 기준 이내로 분석한다는 점(이 점은 MLP 모델에 사용된 지하수 측정망의 수질 분석 방법과 차이가 있음) 및 분석 주기가 짧은 수질 자료가 존재하는 등이 예측 모델의 추세와 실제 추세의 비교를 어렵게 하는 것으로 보인다.

따라서, MLP 모델 예측의 정확도를 평가하기 위하여, 올바르게 예측되는 경우와 틀리게 예측되는 경우의 MLP 모델의 예측 확률을 비교해 보았다(표 7). 지하수 NO₃-N의 실제 추세가 상승인 경우에 MLP 모델에 의하여 동일하게 상승 추세로 예측되는 38개 각각의 예측확률의 평균값은 0.933으로서 틀리게 예측되는 경우의 0.768과 비교하여 훨씬 높은 것으로 나타났다. 또한, 비상승 추세인 경우에도 올바르게 예측한 경우의 예측확률 평균값이 0.899로서 미세하게나마 정확한 것으로 평가되었다. 이 결과는 본 MLP 예측 모델을 이용한다면 실제 미세측 지역

에서의 상승과 비상승을 구분하는 것이 의미가 있음을 보여준다. 또한, 향후 모델의 입력 변수가 추가적으로 반영되고, 장기간의 다수 지점에서의 수질 시계열 데이터에 의한 추세 결과와 비교하게 된다면 MLP 모델에 의한 지하수 NO₃-N의 상승 예측을 보다 정확하게 평가할 수 있을 것이다.

한편, 목포-무안지역 지하수 기초조사시 수행되었던 2018년 기준의 지하수 수질 분석자료로부터 NO₃-N 농도 분포도를 작성한 결과, 수질 기준(10 mg/L)을 초과하는 지역은 MLP 모델에 의한 상승 추세 예측 지역과 유사하게 북부 지역에 집중되는 것을 알 수 있다(그림 3d). 목포-무안 지역내 주요 오염원은 축산단지, 주유소, 세차장 등 다양하며, 도시지역의 인간활동에 의한 오염원은 목포 지역에 집중되어 있는 반면에 축산 오염원은 농촌지역의 구릉지 및 평지에 주로 위치하고 있다(그림 4). 북측 지역의 경우는,

**Fig. 4.** Location of main pollution sources in Mokpo-Muan region (data from MOE and K-water, 2018).

전반적으로 축산 오염원 분포와 $\text{NO}_3\text{-N}$ 의 상승 추세 예측 및 고농도 지점이 유사한 특성을 보인다. 한편, 동남측의 오룡산 북동쪽 지역에 축산 오염원이 다수 분포하나 $\text{NO}_3\text{-N}$ 의 농도는 상대적으로 낮으며 (그림 3d 참조), MLP 모델에 의한 상승 추세는 존재하는 것으로 예측되었다(그림 3c). 이와 같은 결과는, $\text{NO}_3\text{-N}$ 의 상승 추세 모델에서 오염원이 입력인자로 고려되지 않았음에도 불구하고, 장기간의 토지 이용 과정에서 $\text{NO}_3\text{-N}$ 오염이 지속되어 왔으며 궁극적으로 수질 기준을 초과하는 결과를 가져오는 것으로 보인다. 반면에 동남부 지역에서 보듯이 $\text{NO}_3\text{-N}$ 의 상승 추세성이 존재하나 수질 기준을 초과하지 않은 경우는 수질 시료 채취 및 분석 과정의 한계점을 고려하더라도 축산 등 오염원이 비교적 최근에 위치하였거나 오염원에 대한 관리의 결과로 추정할 수도 있으며 추가적인 조사가 필요하다.

7. 토의 및 결론

본 연구의 MLP 모델에 의한 $\text{NO}_3\text{-N}$ 의 추세성 예측은 훈련 데이터의 정확성, 데이터의 양, 입력 인자의 적정 선정 등 다양한 요인에 의하여 달라질 수 있다. 본 연구가 수행되는 시점에서 취득 가능한 입력 데이터의 양과 수준을 고려하였을 때, 개발된 모델의 적용성은 어느 정도 유효한 것으로 평가된다. 추후 $\text{NO}_3\text{-N}$ 의 농도와 관련성이 있는 제반 자료의 취득이 추가로 이루어진다면 보다 개선된 모델로 발전할 수 있을 것이다.

본 연구에서 지하수 오염 농도에 직접 영향을 미치는 오염원의 종류 및 농도, 오염원과의 거리 등이 입력 변수로 사용되지 못하였는데, 이는 이들 데이터의 취득에 어려움이 존재하기 때문이다. 이와 같이 모델 개발시 데이터 취득의 용이성도 중요한 요소로 작용하게 되는데, 모델 개발 시점의 데이터의 취득 뿐 아니라 추후 모델 적용시에도 취득 가능한 자료가 입력 변수로 사용되어야 할 것이다.

현재 지하수 기초조사 등에서는 지하수 수질 관리를 위하여 오염취약성도를 작성하여 매질이 오염에 노출될 경우 취약한 정도를 제공하고 있다. 기존의 오염취약성도는 미국의 지형 및 지질 조건을 토대로 개발된 DRASTIC을 기반으로 하고 있으며, DRASTIC에서는 자연적인 매질 특성을 입력인자로 사용하기

때문에 매질 자체의 오염취약성을 평가하는 도구에 해당한다. 국내의 미고결층이 얇고 산악이 많은 지역에서는 DRASTIC의 적용성이 낮은 점을 고려하여 지질구조선 등을 추가한 변형된 DRASTIC이 사용되고 있다(Chung *et al.*, 2018). 그러나, 하나의 오염 취약성도에 의하여 다수의 오염 항목의 확산을 설명하기에는 한계가 있으며, 정책 집행자가 도면을 활용하기에 어려움이 존재한다. 따라서, 특정 수질 오염원 또는 수질 성분에 노출될 경우 오염의 정도 또는 오염의 악화 정도에 대한 정보를 제공할 수 있는 도면이 분리되어 제공된다면 그 활용도는 훨씬 높아질 것이며, 본 연구와 같이 인공신경망 기반의 예측 모델이 그 역할을 할 수 있을 것이다.

수질 예측 인공신경망 모델 개발은 수 십 년간의 지하수 수질 자료를 활용하고, 국내의 지형, 지질, 토양, 오염원 등의 다양한 인자를 활용할 수 있어, 국내 환경에 적합하고 보다 정확한 예측 모델이 개발될 수 있으며, 지속적인 데이터 누적을 통하여 모델의 개선도 가능해질 것이다. 이를 통하여 새로운 지하수 수질 관리 기법의 도입이 가능할 것이다.

감사의 글

이 논문은 대전대학교 2021년도 교내학술연구비 지원에 의하여 수행되었습니다.

REFERENCES

- Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., Ehteram, M. and Elshafie, A., 2019, Machine learning methods for better water quality prediction. *Journal of Hydrology*, 578, 124084, <https://doi.org/10.1016/j.jhydrol.2019.124084>.
- An, H.S., Jeon, S.W., Lee, S.J., Hyun, Y.J., Yoon, H.S. and Kim, R.H., 2015, Suggestion of a groundwater quality management framework using threshold values and trend analysis. *Journal of Soil and Groundwater Environment*, 20, 112-120 (in Korean with English abstract).
- Bilali, A. El., Taleb, A. and Brouzyne, Y., 2021, Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, 245, 106625, <https://doi.org/10.1016/j.agwat.2020.106625>.
- Bowers, J.A. and Shedrow, C.B., 2000, Predicting stream water quality using artificial neural networks (ANN). In: *Development & application of computer techniques*

- to environmental studies VII (ed. by Brebbia, C.A., Zannetti, P. and Ibarra-Berastegiet, G.), WIT Transactions on Ecology and the Environment, 41, 89-97.
- Chung, S.Y., Elzain, H.E., Senapathi, V., Park, K.H., Kwon, H.W., Yoo, I.K. and Oh, H.R., 2018, Assessment of Groundwater Contamination Vulnerability in Miryang City, Korea using Advanced DRASTIC and fuzzy Techniques on the GIS Platform. *Journal of Soil and Groundwater Environment*, 23, 26-41 (in Korean with English abstract).
- Gholami, V., Sebghati, M. and Yousefi, Z., 2016, Integration of artificial neural network and geographic information system applications in simulating groundwater quality. *Environmental Health Engineering and Management Journal*, 3, 173-182, <https://doi.org/10.15171/EHEM.2016.17>.
- Kim, G.B., Choi, D.H., Yoon, P.S. and Kim, K.Y., 2010, Trends of groundwater quality in the areas with a high possibility of pollution. *Journal of the Korean GEO-environmental Society*, 11, 5-16 (in Korean with English abstract).
- Kim, G.B., Kim, J.W., Won, J.H. and Koh, G.W., 2007, Regional trend analysis for groundwater quality in Jeju Island - focusing on chloride and nitrate concentrations. *Journal of Korea Water Resources Association*, 40, 469-483 (in Korean with English abstract).
- Kim, G.B. and Oh, D.H., 2018, Determination of the Groundwater Yield of horizontal wells using an artificial neural network model incorporating riverside groundwater level data. *The Journal of Engineering Geology*, 28, 583-592, <https://doi.org/10.9720/kseg.2018.4.583> (in Korean with English abstract).
- Kim, G.B., Son, Y.C. and Hwang, C.I., 2022, Determination of new national groundwater monitoring sites using artificial neural network model in South Korea. *Geosciences Journal*, <https://doi.org/10.1007/s12303-021-0044-0>.
- Kim, H.R., Oh, J., Do, H.K., Lee, K.J., Hyun, I.H., Oh, S.S., Kam, S.K. and Yun, S.T., 2018, Spatial-temporal variations of nitrate levels in groundwater of Jeju Island, Korea: Evaluation of long-term (1993-2015) monitoring data. *Economic and Environmental Geology*, 51, 15-26 (in Korean with English abstract).
- Minns, A.W. and Hall, M.J., 2009, Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal*, 41, 399-417, <https://doi.org/10.1080/02626669609491511>.
- MOE and K-water, 2018, Basic survey for Mokpo-Muan region (revised ed.) Appendix II, Daejeon, 445 p.
- Mustafa, M.R., Rezaur, R.B., Saiedi, S. and Isa, M.H., 2012, River suspended sediment prediction using various multilayer perceptron neural network training algorithms: a case study in Malaysia. *Water Resources Management*, 26, 1879-1897, <https://doi.org/10.1007/s11269-012-9992-5>.
- Palani, S., Liong, S.-Y. and Tkalich, P., 2008, An ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56, 1586-1597, <https://doi.org/10.1016/j.marpolbul.2008.05.021>.
- Sakizadeh, M., 2016, Artificial intelligence for the prediction of water quality index in groundwater systems. *Modeling Earth Systems and Environment*, 2, 8, <https://doi.org/10.1007/s40808-015-0063-9>.
- Seo, I., Choi, Y., Kim, T., Guen, S., Song, J., Lee, H., Jung, S., Kim, C., Yoon, H. and Kim, T., 2020, Water quality characteristics of the Gyeonggi-do groundwater quality monitoring network. *Journal of Environmental Analysis, Health and Toxicology*, 23, 141-154 (in Korean with English abstract).
- Shiri, J., Kisi, O., Yoon, H., Lee, K.-K. and Nazemi, A.H., 2013, Predicting groundwater level fluctuations with meteorological effect implications-a comparative study among soft computing techniques. *Computers & Geosciences*, 56, 32-44, <https://doi.org/10.1016/j.cageo.2013.01.007>.

Received : March 23, 2022

Revised : April 8, 2022

Accepted : April 8, 2022