



Prediction of Total Phosphorus (T-P) in the Nakdong River basin utilizing In-Situ Sensor-Derived water quality parameters

Kang, YuMin^a · Nam, SuHan^b · Kim, YoungDo^{c*}

^aMaster's Course, Department of Civil & Engineering, MyongJi University, Yongin, Korea

^bPh.D Course, Department of Civil & Engineering, MyongJi University, Yongin, Korea

^cProfessor, Department of Civil & Engineering, MyongJi University, Yongin, Korea

Paper number: 24-011

Received: 21 February 2024; Revised: 21 June 2024; Accepted: 21 June 2024

Abstract

This study aimed to predict total phosphorus (T-P) to address early eutrophication caused by nutrient influx from various human activities. Traditional T-P monitoring systems are labor-intensive and time-consuming, leading to a global trend of using direct reading sensors. Therefore, this study utilized water quality parameters obtained from direct reading sensors in a two-stage T-P prediction process. The importance of turbidity (Tur) in T-P prediction was examined, and an analysis was conducted to determine if T-P prediction is possible using only direct reading sensor parameters by adding automatic water quality analyzer parameters. The study found that T-P concentrations were higher in the mid-lower reaches of the Nakdong River basin compared to the upper reaches. Pearson correlation analysis identified water quality parameters highly correlated with T-P at each site, which were then used in multiple linear regression analysis to predict T-P. The analysis was conducted with and without the inclusion of Tur, and the performance of models incorporating automatic water quality analyzer parameters was compared with those using only direct reading sensor parameters. The results confirmed the significance of Tur in T-P prediction, suggesting that it can be used as a foundational element in the development of measures to prevent eutrophication.

Keywords: T-P (Total Phosphorus), Turbidity, Eutrophication, Correlation analysis, Multiple linear regression analysis

직독식 센서 측정 항목을 활용한 낙동강 유역의 총인(T-P) 예측 연구

강유민^a · 남수한^b · 김영도^{c*}

^a명지대학교 토목환경공학과 석사과정, ^b명지대학교 토목환경공학과 박사과정, ^c명지대학교 토목환경공학과 교수

요 지

본 연구는 다양한 인간의 활동으로 인해 공공수역에 영양물질이 유입됨에 따라 발생하는 부영양화를 초기에 대응하기 위해 T-P (Total Phosphorus) 예측을 진행하였다. 기존의 T-P 모니터링 시스템은 인력 및 시간이 많이 소요되는 단점이 존재해 직독식 센서를 활용한 측정이 국내외를 막론하고 많이 시도되고 있는 추세이다. 따라서 직독식 센서를 통해 얻을 수 있는 수질항목을 활용하여 T-P 예측을 진행하였으며, 두 단계로 나누어 진행하였다. T-P 예측에 있어 Turbidity (Tur)의 중요성에 대해 살펴보았으며, 자동수질분석기 분석항목을 추가한 분석을 통해 직독식 센서 측정 항목만으로 T-P 예측이 가능한지 살펴보았다. 본 연구의 연구 대상 지점인 낙동강 유역 내 T-P 현황을 살펴본 결과, T-P 농도가 상류 지역 대비 중·하류 지역에 높게 나타났다. Pearson 상관분석을 통해 지점별로 T-P와 상관성이 높은 수질항목을 파악하였으며, 이를 활용하여 다중선형회귀 분석을 진행하여 T-P를 예측하였다. Tur의 유무에 따른 분석을 진행하였으며, 자동수질분석기 분석항목이 포함된 분석을 통해 직독식 센서 측정 항목과의 성능을 비교하였다. 결과적으로 Tur 활용의 중요성을 확인하였으며, 이는 부영양화 개선 대책 수립을 위한 보조 자료로 활용할 수 있을 것으로 판단된다.

핵심용어: T-P (Total Phosphorus), 탁도, 부영양화, 상관분석, 다중선형회귀분석

*Corresponding Author. Tel: +82-31-330-6407

E-mail: ydkim@mju.ac.kr (Kim, YoungDo)

1. 서론

도시화 및 산업화와 같은 다양한 인간의 활동으로 인해 질소, 인 등의 영양물질이 공공수역에 유입되어 부영양화 등을 포함한 수환경에 부정적인 영향을 미치고 있다. 부영양화는 과잉의 질소(N), 인(P) 및 기타 무기 영양소가 상대적으로 폐쇄되고 유속이 느린 수역에 유입되어 생물의 증식을 촉진하는 현상이다(Zhang *et al.*, 2021). 이러한 부영양화는 수환경에 부정적인 영향을 미치고 있으며, 이에 따라 정부는 질소 및 인의 관리를 위한 정책 개발과 환경 기준 설정 등 지속적인 노력을 기울이고 있다. 특히, 많은 호소와 하천에서 수중의 인이 부영양화의 주 원인물질로 인식되고 있기 때문에 수질 정책이 인의 관리에 집중되어 있다(Kim *et al.*, 2015).

과도한 인의 유입으로 부영양화된 수생태계는 시아노박테리아가 생성하는 마이크로시스틴과 같은 독소를 생성할 수 있으며, 이는 인체 건강에 부정적인 영향을 미칠 수 있다. 또한 정수처리장 기능 악화와 수중의 용존 산소 감소로 수중 생태계 파괴 및 악취 등을 야기할 수 있다(Jung *et al.*, 2020; Lim *et al.*, 2015). 따라서 하천에서 부영양화가 발생하지 않게 하기 위해서는 영양물질의 관리가 중요하며, 특히 강우 시 T-P 부하량을 정확하게 파악하는 것이 어렵기 때문에 효과적인 수질관리가 더욱 필요하다. 하지만 기존의 T-P 모니터링은 현장에서 채취한 물 샘플을 바탕으로 실험실 기반의 분석이 이루어지고 있으며, 이 과정은 시간적, 재정적, 인력적 소요가 크다는 단점이 발생한다. 따라서 하천의 영양 상태를 체계적으로 모니터링하기 어려운 실정이다(Baik *et al.*, 2020). 이러한 단점을 보완하기 위해 최근 직독식 센서를 활용한 측정이 활발히 이루어지고 있다. 센서를 활용할 경우 단기간에 많은 양의 데이터를 획득할 수 있으며, 인력 및 시간 소요가 적게 발생한다는 장점이 존재해 국내외를 막론하고 많이 시도되고 있는 추세이다. 특히, 수질항목 중 탁도는 많은 연구에서 T-P의 대체항목으로 사용되었다. 선행연구에 따르면, T-P와 탁도 사이에는 유의미한 상관관계가 있었으며, 탁도를 T-P의 대체항목으로 사용할 수 있다고 보고되고 있다(Valkama and Ruth, 2017). Kämäri *et al.* (2020)은 탁도 센서를 활용하여 탁도가 T-P 농도를 잘 파악할 수 있는지 살펴보았으며, 결과적으로 탁도가 T-P 예측에 도움이 되는 수질항목으로 나타났다고 보고하였다. Villa *et al.* (2019)은 108개의 하천 중 84개 하천에서 탁도를 T-P의 대체항목으로 평가한 결과, T-P와 탁도의 선형회귀분석을 통해 평균 R² 값이 0.62로 나타나, 탁도가 대부분의 하천 유형에서 T-P를 예측하는 데 유용하다고 보고하였다.

따라서, 본 연구에서는 센서 기반의 데이터 확보를 위해 물

환경정보시스템의 자동측정망 일자료를 활용하여 T-P 예측에 대한 연구를 다음과 같이 두 단계로 나누어 진행하였다. 우선, T-P 예측에 있어 탁도의 중요성에 대해 살펴보았으며, 그 다음으로는 자동수질분석기 분석항목을 추가한 분석을 통해 직독식 센서 측정 항목만으로 T-P 예측이 가능한지 살펴보았다. 분석 방법으로는 상관분석과 다중선형회귀분석이 사용되었다. T-P와 수질항목 간의 상관성을 확인하기 위해 Pearson 상관분석을 진행하였으며, 상관계수를 토대로 T-P와 상관성이 높은 수질항목을 하나씩 추가하여 분석하는 Stepwise 변수선택법을 사용하여 다중선형회귀분석(Multiple Linear Regression)을 진행하였다. 또한 도출된 회귀식의 시각화를 통해 T-P 예측 패턴의 경향성을 살펴보았으며, 이를 활용하여 부영양화 개선 대책 수립을 위한 보조 자료로 활용하고자 한다.

2. 연구방법

2.1 연구범위 및 연구 대상 지점

본 연구에서는 대한민국 4대강 중 연중 부영양화 발생 빈도가 높은 낙동강 유역을 연구 대상 지점으로 선정하였다. 낙동강 유역은 대구·경북, 부산·경남권의 식수, 농·공업용수로 아주 중요한 역할을 하고 있으며, 대구지방환경청과 낙동강유역환경청, 주변 지자체 등에서 수질보전을 위해 많은 노력을 기울이고 있다(Ahn *et al.*, 2016). 낙동강 유역의 자동측정망을 기준으로 12지점을 연구 대상 지점으로 선정하였다(Fig. 1).

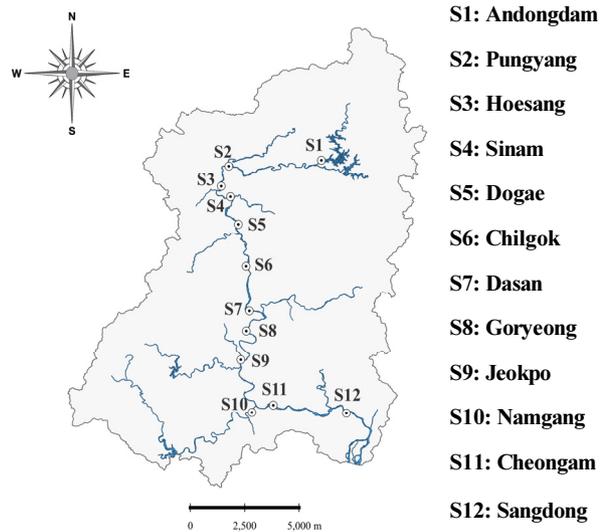


Fig. 1. Description of the sampling sites in Nakdong River basin

2.2 자료 수집

본 연구에서는 낙동강 유역 내 12지점인 안동댐하류(Andongdam), 풍양(Pungyang), 회상(Hoesang), 신암(Sinam), 도개(Dogae), 칠곡(Chilgok), 다산(Dasan), 고령(Goryeong), 적포(Jeokpo), 남강(Namgang), 청암(Cheongam), 상동(Sangdong)지점의 수질자료를 환경부에서 운영하는 ‘물환경정보시스템(<http://water.nier.go.kr>)’에서 수집하였다. 2012년 7월부터 2022년 12월까지 자동측정망 일자료를 수집하였으며, 물환경정보시스템에서 제공되는 다양한 수질자료 중 직독식 센서로 측정 가능한 수질항목과 자동수질분석기 분석항목을 구분하여 사용하였다. 직독식 센서로 측정 가능한 수질항목인 Water Temperature (WT), pH, Electric Conductivity (EC), Dissolved Oxygen (DO), Turbidity (Tur), Total Phosphorus (T-P), Chl-*a* 자료를 수집하였으며, 자동수질분석기 분석항목인 Total Organic Carbon (TOC), Total Nitrogen (T-N), Phosphate-Phosphorus (PO₄-P) 자료를 수집하여 연구에 활용하였다. 지점별로 사용된 수질항목은 Table 1에 나타냈다.

2.3 분석 방법

2.3.1 Pearson 상관분석

낙동강 유역 내 12지점을 대상으로 T-P와 높은 상관성을 가진 수질항목을 파악하기 위해 Pearson 상관분석을 진행하였다. 상관계수는 -1에서 +1까지의 값을 가지며, 절대값이 1에 가까울수록 상관성이 강하다는 것을 의미한다. Pearson 상관계수(*r*)는 Eq. (1)에 따라 계산되며, 변수가 여러 개인 두 데이터를 비교할 때, 각각 *x*와 *y*변수에 대입하여 계산한다(Jung *et al.*, 2019).

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{1}{n} \sum_i z_{x_i} z_{y_i} \quad (1)$$

여기서 z_{x_i} 와 z_{y_i} 는 각각 $(x_i - \bar{x}) / \sqrt{\sum_i (x_i - \bar{x})^2 / n}$ 과 $(y_i - \bar{y}) /$

$\sqrt{\sum_i (y_i - \bar{y})^2 / n}$ 으로 정의한다. 이러한 변형식을 이용하면 Pearson 상관계수를 보다 간명하게 표시할 수 있으며, 이는 두 변수의 표준점수를 서로 곱한 값들의 평균으로 정의될 수 있다(Lee *et al.*, 2018).

2.3.2 다중선형회귀분석(Multiple Linear Regression)

다중선형회귀분석은 한 개의 종속변수에 대해 두 개 이상의 독립변수 사이의 관계를 규명하고 모형화하기 위해 사용되는 통계적 기법이다. 자료들이 선형적인 관계에 있을 때, *n*개의 독립변수 $x_i (i = 1, 2, \dots, n)$ 에 대한 관측치 *y*의 다중선형회귀모형은 다음과 같이 표현할 수 있다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (2)$$

여기서, ϵ 와 β_0 는 각각 다중선형회귀분석에 대한 오차와 *y*축 절편을 나타내며, β_n 은 독립변수 x_n 에 대한 종속변수의 변화량을 나타내는 회귀계수이다(Bong and Kim, 2017). 다중선형회귀모형의 성능 비교 및 평가를 위해 평균 절대 오차(mean absolute error, MAE), 평균 절대 비율 오차(mean absolute percentage error, MAPE), 평균 제곱근 오차(root mean square error, RMSE)와 R²를 사용하였다(Choi and Kim, 2021). MAE, RMSE, MAPE, R²는 관측값과 모델 예측값의 회귀 및 상관 지표들 다룰 때 사용되며, 모델의 정확도를 확인하는 데 사용된다. R²는 Eq. (3)과 같이 계산되며, 회귀분석 모델이 두 값 간의 데이터를 얼마나 잘 설명하는지 나타내는 지표로, 0에서 1 사이의 값을 가지며 1에 가까울수록 설명력이 높음을 의미한다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

Table 1. Water quality parameters evaluated in this study from monitoring sites

Sites	Water Quality Parameters
Andongdam, Pungyang, Namgang	WT (°C), pH, EC (μS/cm), DO (mg/L), T-N (mg/L), T-P (mg/L), TOC (mg/L), Chl- <i>a</i> (mg/m ³)
Hoesang, Sinam, Dogae, Chilgok, Goryeong, Jeokpo, Cheongam, Sangdong	WT (°C), pH, EC (μS/cm), DO (mg/L), Tur (NTU), T-N (mg/L), T-P (mg/L), TOC (mg/L), Chl- <i>a</i> (mg/m ³)
Dasan	WT (°C), pH, EC (μS/cm), DO (mg/L), Tur (NTU), T-N (mg/L), T-P (mg/L), TOC (mg/L), Chl- <i>a</i> (mg/m ³), PO ₄ -P (mg/L)

위 식에서 n 은 총 데이터의 개수, y_i 는 관측값, \hat{y}_i 는 예측값, \bar{y} 는 관측값의 평균을 의미하며, i 는 모델 입력값에 따른 순서이다. RMSE는 Eq. (4)와 같이 오차에 대한 제곱값을 평균하기 때문에 오차가 크면 클수록 기하급수적으로 증가하는 결과를 나타내므로 MAE보다는 오차의 크기에 민감하게 반응한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

MAPE는 Eq. (5)와 같이 상대오차의 절대값을 100분율로 변환한 것으로, 관측값 대비 모델 예측 오차에 대한 편향이 존재한다.

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

MAE는 Eq. (6)과 같이 모델 예측값과 관측값의 차이에 대한 절대값을 평균한 값이며, 예측값의 오차가 선형적으로 반영되므로 이상치가 많은 경우 값이 커진다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

MAE, RMSE, MAPE는 값이 작을수록 모델의 정확도가 높음을 의미하며, 값이 클수록 정확도가 다소 떨어짐을 의미한다. 반면에 R^2 는 값이 클수록 모델의 설명력이 높고, 값이 작을수록 설명력이 낮음을 의미한다(Kim et al., 2023).

3. 결과 및 고찰

3.1 낙동강 유역 내 T-P 현황

앞서 수집한 자료를 활용하여 T-P 농도 분포를 Table 2에 제시하였다. 지점별 T-P 농도 분석 결과, 낙동강 중·하류 지역의 T-P 농도가 낙동강 상류 지역 대비 높은 T-P 농도를 나타냈으며, 중·하류 지역은 평균적으로 하천 생활환경 기준인 T-P 수질기준(1a 등급)을 초과하였다. 대부분 0.02 mg/L (1a 등급) 이상 0.04 mg/L (1b 등급) 이하 수준이었다(Table 2). 또한, 모든 지점의 T-P 최대치는 0.1 mg/L (II 등급) 이상 0.5 mg/L (V 등급) 이하 수준으로 조사되었다. 높은 수준의 T-P 농도는 수계의 부영양화를 유발할 수 있어 T-P에 대한 관리가 필요할 것으로 판단된다. 낙동강 중·하류 지역의 T-P 농도가 낙동강

Table 2. Descriptive Statistics of T-P data

Sites	Number of Data	AVE	STD	MAX	MIN
Andongdam	2,687	0.010	0.017	0.309	0.002
Pungyang	1,945	0.016	0.014	0.103	0.003
Hoesang	2,493	0.018	0.017	0.195	0.002
Sinam	2,628	0.022	0.018	0.142	0.002
Dogae	2,895	0.020	0.020	0.455	0.002
Chilgok	2,432	0.023	0.019	0.116	0.002
Dasan	2,311	0.028	0.018	0.136	0.003
Goryeong	2,391	0.036	0.025	0.161	0.003
Jeokpo	2,737	0.033	0.027	0.195	0.003
Namgang	2,761	0.026	0.021	0.219	0.002
Cheongam	2,618	0.022	0.021	0.115	0.002
Sangdong	2,693	0.025	0.018	0.127	0.003

상류 지역 대비 높은 T-P 농도를 나타내는 것은 하천 상류구간에서 유출되고 있는 축산폐수, 생활하수와 농경지 유출수 등의 비점오염원과 중·하류부에 집중되어 있는 구미, 대구와 같은 대도시와 산업단지로부터 배출되는 많은 오염물질이 분류로 직접 유입되거나 지류를 통해 하천에 유입되고 있어 T-P 농도에 큰 영향을 준 것으로 판단된다(Jung and Kim, 2017).

3.2 상관분석 결과

연구 대상 지점인 낙동강 유역 내 12지점에 대해 Pearson 상관분석을 진행하였으며, T-P와 다른 수질항목 간의 Pearson 상관관계를 Table 3에 제시하였다. 물환경정보시스템에서 수집한 수질항목 중 WT, pH, EC, DO, TOC, T-N, Chl- α 는 모든 지점에서 측정되었으며, Tur은 Andongdam, Pungyang 및 Namgang 지점을 제외한 모든 지점에서 측정되었다. PO₄-P는 유일하게 Dasan 지점에서만 측정되었다. Andongdam, Pungyang 및 Namgang 지점의 경우, Tur 및 PO₄-P를 측정하지 않는 지점으로 자동수질분석기 분석항목인 TOC가 각각 0.499, 0.493, 0.353으로 다른 항목에 비해 T-P와 높은 상관성을 나타냈으며, Andongdam 지점은 pH가 -0.005로 매우 낮은 상관성을 나타냈다. 또한 Pungyang 및 Namgang 지점의 경우, Chl- α 가 각각 -0.067, -0.031로 매우 낮은 상관성을 나타냈다. Hoesang, Sinam, Dogae, Chilgok, Goryeong, Jeokpo, Cheongam, Sangdong 지점의 경우, PO₄-P가 측정되지 않는 지점으로 T-N 및 Chl- α 를 제외한 수질항목에서 대체적으로 T-P와 다소 높은 상관성을 나타냈다. 특히 Dasan 지점의 경우 모든 수질항목이 측정되는 지점으로 PO₄-P가 0.848로 T-P와 가장 높은 상관성을 나타냈으며, T-N은 -0.018로 매우 낮은 상관성을 나타

냈다. 대체적으로 모든 지점에서 WT, pH, EC, DO, Tur 및 TOC 는 T-P와 다소 높은 상관성을 나타냈으며, T-N 및 Chl-*a*는 낮은 상관성을 나타냈다.

3.3 T-P 예측을 위한 Multiple linear regression analysis

3.3.1 다중선형회귀분석을 활용한 T-P 예측

Han *et al.* (2012)은 다변수 간의 복잡한 상호작용을 고려하여 T-P를 예측하기 위해 다중선형회귀분석을 사용하였다. 이를 통해 각 수질 파라미터가 T-P 농도에 미치는 영향을 정량화 하고, 이들 간의 상관관계를 분석하였다. 본 연구에서도 T-P와 상관성이 높은 수질항목을 활용하여 T-P 농도 변화를 예측하기 위해 다중선형회귀분석을 진행하였다. T-P 예측 모델에

서 사용된 수질항목과 예측 결과인 R² 및 P-value (*P*)를 Table 4에 제시하였다. 제시된 예측 모델의 독립변수들이 모두 유의한 관계(*P*<0.05)를 나타냈다. Hoesang, Sinam, Dogae, Chilgok, Dasan, Goryeong, Jeokpo, Cheongam 및 Sangdong 지점의 다중선형회귀분석 모델의 경우 R² = 0.369~0.565로 분석되었으며, Andongdam, Pungyang 및 Namgang 지점의 경우 R²는 각각 0.064, 0.178, 0.247로 분석되었다. 연구 대상 지점 중 Tur이 측정되는 지점이 Tur이 측정되지 않는 지점에 비해 R²는 높았으나, 모두 0.6 이하로 분석되어 정확성이 높지 않은 것으로 판단된다. 이러한 결과는 T-P가 다양한 수질 인자의 복합적인 영향을 받아 변화되기 때문인 것으로 판단된다(Lee *et al.*, 2022).

Table 3. Correlation coefficient between T-P and water quality parameters

Sites	WT	pH	EC	DO	Tur	TOC	T-N	Chl- <i>a</i>	PO ₄ -P
Andongdam	0.216*	-0.005*	0.040*	-0.093*	-	0.499*	0.344*	0.164*	-
Pungyang	0.251*	-0.235*	-0.162*	-0.241*	-	0.493*	0.206*	-0.067*	-
Hoesang	0.249*	-0.454*	-0.448*	-0.537*	0.429*	0.568*	0.086*	-0.138*	-
Sinam	0.532*	-0.389*	-0.516*	-0.553*	0.441*	0.403*	0.218*	-0.185*	-
Dogae	0.465*	-0.353*	-0.403*	-0.394*	0.135*	0.454*	0.082*	-0.267*	-
Chilgok	0.409*	-0.437*	-0.490*	-0.496*	0.460*	0.360*	-0.030*	-0.171*	-
Dasan	0.355*	-0.532*	-0.506*	-0.392*	0.590*	0.335*	-0.018*	-0.293*	0.848*
Goryeong	0.558*	-0.386*	-0.521*	-0.523*	0.491*	0.279*	-0.326*	-0.066*	-
Jeokpo	0.430*	-0.338*	-0.542*	-0.409*	0.472*	0.370*	-0.179*	-0.095*	-
Namgang	0.336*	-0.328*	-0.306*	-0.436*	-	0.353*	-0.161*	-0.031*	-
Cheongam	0.495*	-0.440*	-0.526*	-0.558*	0.300*	0.290*	-0.099*	-0.140*	-
Sangdong	0.412*	-0.351*	-0.439*	-0.431*	0.485*	0.336*	-0.107*	-0.066*	-

* = P-value < 0.05

Table 4. Multiple linear regressions of T-P using water quality parameters for each sites

Sites	Selected parameters in regression model	R ²	P
Andongdam	EC, WT, DO, pH, Chl- <i>a</i>	0.064	0.000
Pungyang	WT, DO, pH, Chl- <i>a</i>	0.178	0.000
Hoesang	Tur, EC, WT, DO, pH	0.479	0.000
Sinam	Tur, EC, WT, pH	0.534	0.000
Dogae	Tur, EC, WT, DO, pH	0.369	0.000
Chilgok	Tur, EC, WT, DO, pH	0.453	0.000
Dasan	Tur, EC, WT, DO, pH, Chl- <i>a</i>	0.539	0.000
Goryeong	Tur, EC, WT, DO, pH, Chl- <i>a</i>	0.565	0.000
Jeokpo	Tur, EC, WT, DO, pH, Chl- <i>a</i>	0.536	0.000
Namgang	EC, WT, pH, Chl- <i>a</i>	0.247	0.000
Cheongam	Tur, EC, WT, pH, Chl- <i>a</i>	0.531	0.000
Sangdong	Tur, EC, WT, pH, Chl- <i>a</i>	0.372	0.000

3.3.2 Tur 유무에 따른 T-P 예측

앞서 분석된 R² 값을 통해 T-P 다중선형회귀분석에서 Tur의 중요성을 확인하였으며, 이를 시계열 그래프(Fig. 2)를 통해 추가로 검증하였다. Tur의 필요성을 설명하기 위하여, Tur이 존재하는 9개 지점의 다중선형회귀분석 모델에서 Tur을 제외한 분석을 진행하였고, 그 결과를 Table 5에 제시하였다. 성능 비교 결과, Tur의 유무에 따라 R², MAE, MAPE 및 RMSE에서 차이가 나타났다. Tur을 포함한 모델의 R² = 0.369~0.565로 분석되었으며, Tur을 미포함한 모델의 R² = 0.332~0.514로 분석되어 0.003~0.099정도 낮게 분석되었다. 또한 MAE 및 MAPE 모두 Tur을 포함한 모델에서 낮은 값을 보여 오차의 크기가 감소함을 확인하였다. Tur 포함 시에 MAE는 일관되게 낮게 나타나며, MAPE는 0.411에서 0.784로, Tur 미포함 시

0.473~0.811로 분석되어 Tur 포함 시 오차를 감소시키는 데 기여함을 확인하였다. RMSE 또한 Tur이 포함된 모델에서 더 낮은 값을 나타내어 예측의 정확성이 Tur 포함에 의해 향상되었음을 보여주었다. Tur이 측정되는 Hoesang, Sinam, Dogae, Chilgok, Dasan, Goryeong, Jeokpo, Cheongam 및 Sangdong 지점의 다중선형회귀분석 모델의 경우 Tur이 측정되지 않는 Andongdam, Pungyang 및 Namgang 지점에 비해 T-P 예측 정확성이 더 높게 나타나는 것으로 보아, T-P 다중선형회귀분석에서 Tur이 포함된 분석이 중요함을 확인할 수 있었다. Tur을 포함한 다중선형회귀분석이 T-P 예측에 있어 중요한 요소로 작용하였으며, 결과적으로 Tur은 T-P 예측에 도움이 되는 수질항목으로 판단된다(Kämäri *et al.*, 2020).

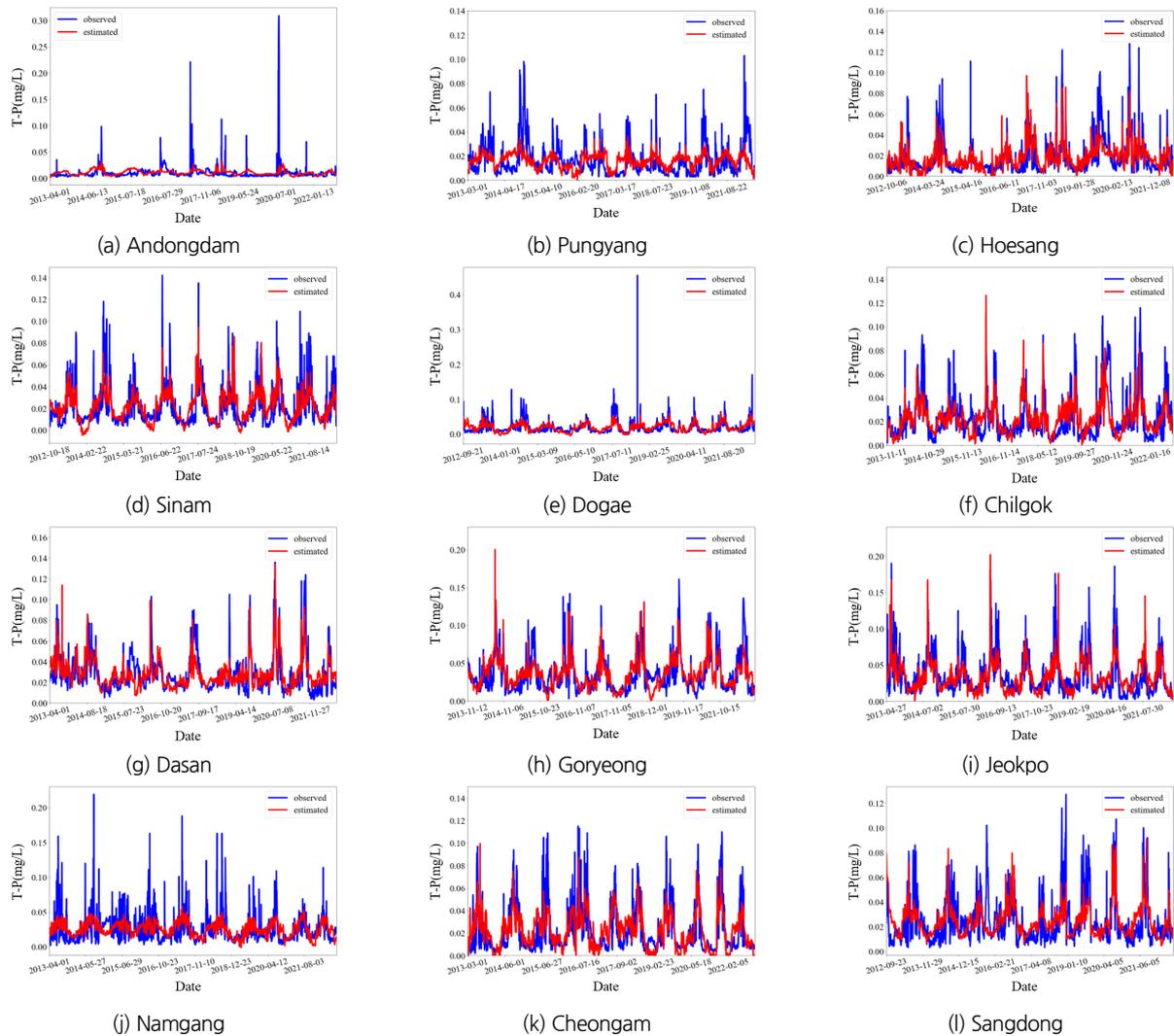


Fig. 2. Regression analysis for each sites based on the various parameters of T-P

3.3.3 고농도 T-P에 대한 변수 중요도 분석 결과

Fig. 2에서 Andongdam, Pungyang 및 Namgang 지점에서는 일정 수준 이상의 T-P 농도 예측에 어려움이 있었으나, 나머지 9개 지점에서는 대체로 고농도의 T-P 예측이 가능함을

확인하였다. 이를 통해 Tur이 고농도의 T-P 예측에 중요한 역할을 한다는 것을 확인하였다. 따라서 고농도의 T-P 예측에서 Tur의 영향을 확인하기 위해 변수 중요도 분석을 추가로 수행하였다. 분석에서는 전체 데이터 중 상위 20%에 해당하는 고

Table 5. Comparison of Performance with Turbidity Existence and Nonexistence: (a) Tur Existence, (b) Tur Nonexistence

Sites	R ²		MAE		MAPE		RMSE	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Andongdam	-	0.064	-	0.007	-	0.757	-	0.017
Pungyang	-	0.178	-	0.009	-	0.802	-	0.012
Hoesang	0.479	0.444	0.008	0.009	0.690	0.736	0.012	0.012
Sinam	0.534	0.514	0.009	0.009	0.630	0.656	0.012	0.013
Dogae	0.369	0.366	0.009	0.009	0.706	0.706	0.016	0.016
Chilgok	0.453	0.395	0.010	0.011	0.672	0.726	0.014	0.015
Dasan	0.539	0.440	0.009	0.010	0.411	0.473	0.013	0.014
Goryeong	0.565	0.474	0.012	0.014	0.486	0.558	0.016	0.018
Jeokpo	0.536	0.441	0.013	0.015	0.663	0.724	0.018	0.020
Namgang	-	0.247	-	0.012	-	0.727	-	0.018
Cheongam	0.531	0.514	0.011	0.011	0.784	0.811	0.014	0.015
Sangdong	0.372	0.332	0.011	0.011	0.639	0.677	0.015	0.015

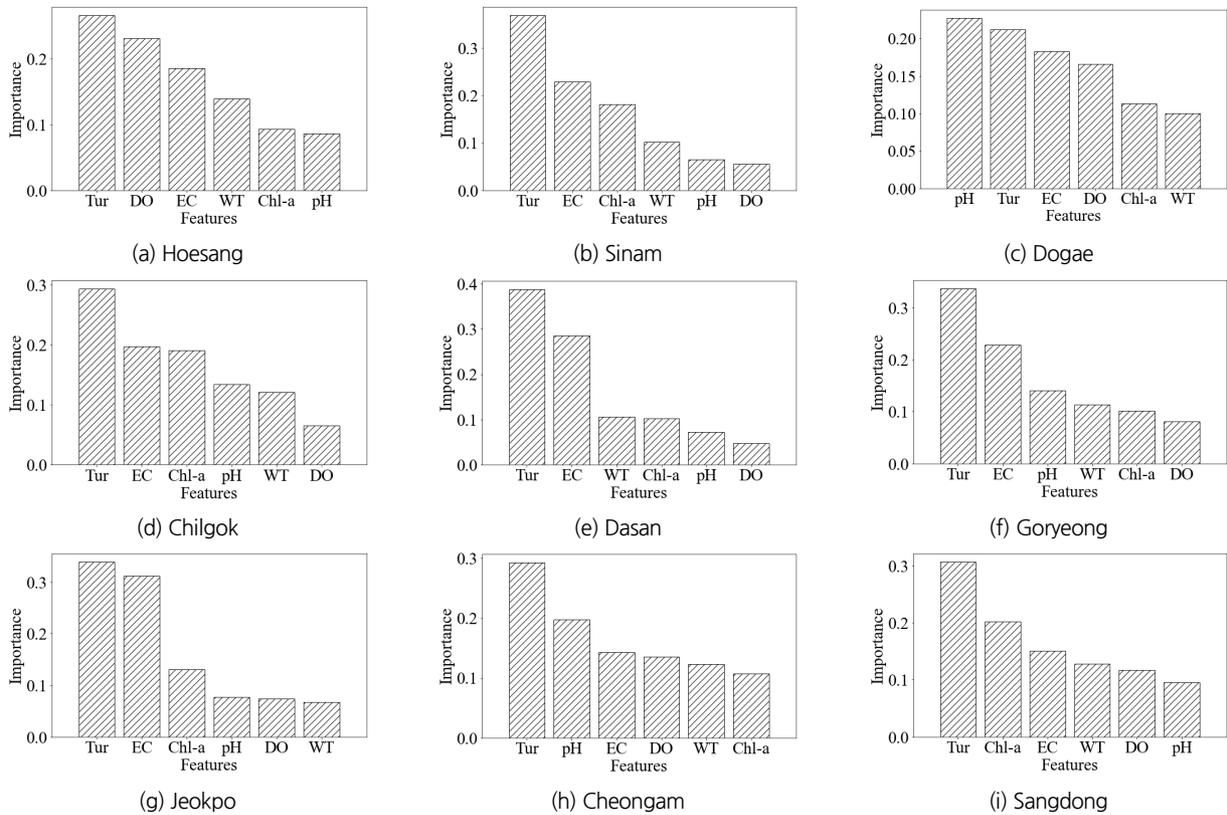


Fig. 3. Importance of Different Water Quality Parameters in Predicting Total Phosphorus (T-P) Concentration. The bars represent the relative importance of each feature, with Turbidity (Tur) showing the highest importance

농도 T-P 데이터를 활용하였다. 변수 중요도 분석은 표준화 회귀계수를 사용하여 평가하였다. Fig. 3에 나타난 바와 같이, 대부분의 지점에서 Tur이 가장 높은 변수 중요도를 보였다. 이는 Tur이 고농도의 T-P 예측에 있어 가장 중요한 변수로 작용함을 의미한다. 따라서 Tur은 T-P 예측 모델에서 중요한 수질항목으로 판단되며, Tur 센서의 중요성 또한 확인하였다. 이 결과는 T-P 예측 모델에 Tur을 포함시키는 것이 유용할 수 있음을 시사하며, 향후 수질관리 및 부영양화 개선 대책 수립을 위한 보조 자료로 활용될 수 있을 것이다.

3.3.4 자동수질분석기 분석항목을 포함한 T-P 예측

앞선 연구를 통해 T-P 예측에 있어 Tur의 중요성에 대해 살펴 보았으며, 자동수질분석기 분석항목을 추가한 분석을 통해 직독식 센서 측정 항목만으로 T-P 예측이 가능한지 조사하였다. 각 지점별 T-P 예측 모델에서 사용된 수질항목에 자동수질분석기 분석항목인 TOC, T-N 및 PO₄-P를 각각 추가하여 분석을 진행하였으며, 분석 결과에 따른 성능 비교를 Table 6에 제시하였다. 제시된 것과 같이 (a)는 직독식 센서 측정 항목인 WT (°C), pH, EC (μ S/cm), DO (mg/L), Tur (NTU), Chl-*a*

Table 6. Performance comparison according to various analyzes: (a) Sensor Measurement Items, (b) Addition of TOC, (c) Addition of T-N, (d) Addition of PO₄-P

Sites		R ²	MAE	MAPE	RMSE
Hoesang	(a)	0.479	0.008	0.690	0.012
	(b)	0.534	0.008	0.685	0.011
	(c)	0.509	0.008	0.651	0.012
Sinam	(a)	0.534	0.009	0.630	0.012
	(b)	0.560	0.009	0.629	0.012
	(c)	0.570	0.008	0.582	0.012
Dogae	(a)	0.369	0.009	0.706	0.016
	(b)	0.462	0.009	0.706	0.013
	(c)	0.399	0.009	0.688	0.015
Chilgok	(a)	0.453	0.010	0.672	0.014
	(b)	0.486	0.010	0.614	0.014
	(c)	0.497	0.010	0.644	0.014
Dasan	(a)	0.539	0.009	0.411	0.013
	(b)	0.553	0.009	0.404	0.012
	(c)	0.530	0.009	0.403	0.012
	(d)	0.776	0.006	0.240	0.008
Goryeong	(a)	0.565	0.012	0.486	0.016
	(b)	0.609	0.012	0.473	0.016
	(c)	0.561	0.012	0.487	0.016
Jeokpo	(a)	0.536	0.013	0.663	0.018
	(b)	0.570	0.013	0.634	0.018
	(c)	0.565	0.013	0.653	0.018
Cheongam	(a)	0.531	0.011	0.784	0.014
	(b)	0.554	0.011	0.783	0.014
	(c)	0.559	0.011	0.778	0.014
Sangdong	(a)	0.372	0.011	0.639	0.015
	(b)	0.420	0.011	0.621	0.014
	(c)	0.376	0.011	0.626	0.015

(mg/m³)을 의미하며, (b)는 직독식 센서 측정 항목에 TOC를 추가한 WT (°C), pH, EC (μS/cm), DO (mg/L), Tur (NTU), Chl-*a* (mg/m³), TOC (mg/L)를 의미한다. 그리고 (c)는 직독식 센서 측정 항목에 T-N을 추가한 WT(°C), pH, EC (μS/cm), DO (mg/L), Tur (NTU), Chl-*a* (mg/m³), T-N (mg/L)을 의미하며, (d)는 직독식 센서 측정 항목에 PO₄-P를 추가한 WT (°C), pH, EC (μS/cm), DO (mg/L), Tur (NTU), Chl-*a* (mg/m³), PO₄-P (mg/L)를 의미한다. 분석 결과, 9개 지점 (a), (b), (c)의 평균 R²는 각각 0.486, 0.528, 0.507로 분석되어 자동수질분석기 분석항목이 포함된 다중선형회귀분석이 직독식 센서 측정 항목보다 높은 정확도를 나타냈다. 9개 지점 (a), (b), (c)의 평균 MAE는 각각 0.01, 0.01, 0.01로 분석되었으며, 9개 지점 (a), (b), (c)의 평균 MAPE의 경우 각각 0.6312, 0.6166, 0.6124로 분석되었다. 또한 9개 지점 (a), (b), (c)의 평균 RMSE의 경우 각각 0.0144, 0.0138, 0.0142로 분석되어 직독식 센서 측정 항목에 비해 TOC 및 T-N이 추가된 분석의 MAE, MAPE, RMSE가 낮게 나타났다. 또한 PO₄-P가 추가된 Dasan 지점의 R²는 0.776으로 분석되어 직독식 센서 측정 항목에 비해 높게 나타났으며, MAE, MAPE 및 RMSE는 각각 0.006, 0.240, 0.008로 분석되어 직독식 센서 측정 항목에 비해 낮게 나타났다. 결과적으로 직독식 센서 측정 항목에 비해 자동수질분석기 분석항목이 포함된 다중선형회귀분석의 R²는 더 높은 성능으로 분석되었으며, MAE, MAPE 및 RMSE는 낮은 성능으로 분석되었다. 하지만 분석 결과에 따르면, 직독식 센서 측정 항목만을 사용한 경우와 자동수질분석기 항목을 추가한 경우 사이의 R², MAE, MAPE 및 RMSE 차이는 미미했다. 이는 직독식 센서 측정 항목만으로도 T-P의 변동성을 일정 수준 이상 예측할 수 있음을 시사한다.

4. 결론

본 연구는 직독식 센서 측정 항목을 활용하여 T-P 예측 분석을 진행하였다. 물환경정보시스템에서 직독식 센서를 통해 얻을 수 있는 수질자료 수집하여 낙동강 유역 내 12지점에 대한 분석을 진행하였으며, 2012년 7월부터 2022년 12월까지의 자동측정망 일자료를 활용하였다. 수집한 자료를 활용하여 분석한 결과 다음과 같은 결론을 도출하였다.

- 1) 낙동강 중·하류 지역의 T-P 농도는 낙동강 상류보다 높게 나타났으며, 중·하류 지역은 평균적으로 하천 생활환경 기준인 Ia 등급을 초과하였다. 대부분 0.02 mg/L (Ia 등급) 이상 0.04 mg/L (Ib 등급) 이하 수준이었으며, 모든 지점의 T-P 최대치는 0.1 mg/L (II 등급) 이상 0.5 mg/L (V 등급) 이하 수준으로 조사되었다. 낙동강 중·하류 지역의 T-P 농도가 높게 나타나는 것은 하천 상류구간에서 유출되고 있는 축산폐수, 생활하수, 농경지 유출수 등의 비점오염원과 중·하류부에 집중되어 있는 대도시와 산업단지로부터 배출되는 많은 오염물질이 본류로 직접 유입되거나 지류를 통해 하천에 유입되고 있어 T-P 농도에 큰 영향을 준 것으로 판단된다.
- 2) 연구 대상 지점인 낙동강 유역 내 12지점의 T-P와 다른 수질항목 간의 Pearson 상관관계를 분석하였다. 수질항목 중 WT, pH, EC, DO, TOC, T-N, Chl-*a*은 모든 지점에서 측정되었으며, Tur 및 PO₄-P는 일부 지점에서 측정되었다. T-P와 다른 수질항목 간의 상관분석 결과, PO₄-P가 T-P와 상관계수가 0.848로 분석되어 가장 높게 나타났으며 WT, pH, EC, DO, Tur 및 TOC는 대체적으로 모든 지점에서 높은 상관성을 나타냈다.
- 3) 상관분석 결과를 바탕으로 T-P 예측을 위한 다중선형회귀분석을 진행하였다. Hoesang, Sinam, Dogae, Chilgok, Dasan, Goryeong, Jeokpo, Cheongam 및 Sangdong 지점의 예측 정확도는 R² = 0.369~0.565로 분석되었으며, Andongdam, Pungyang 및 Namgang 지점의 경우 각각 R² = 0.064, 0.178, 0.247로 분석되어 연구 대상 지점 중 Tur이 측정되는 지점이 Tur이 측정되지 않는 지점에 비해 예측 성능은 높았으나, 모두 0.6 이하로 분석되어 정확성이 높지 않은 것으로 나타났다. 또한 Tur의 유무에 따른 분석을 통해 Tur이 T-P를 예측하는 데 중요한 역할을 한다는 것을 확인하였다. Andongdam, Pungyang, Namgang 지점에서는 일정 수준 이상의 T-P 농도 예측에 어려움이 있었으나, 나머지 9개 지점에서는 어느 정도 고농도의 T-P 예측이 가능했다. 이를 확인하기 위해 고농도 T-P의 변수 중요도를 분석한 결과, 대부분의 지점에서 Tur이 가장 높은 중요도를 보였다. 이는 Tur이 T-P 예측 모델에서 중요한 수질항목으로 작용함을 의미하며, 향후 수질 관리 및 부영양화 개선 대책 수립을 위한 보조 자료로 활용될 수 있을 것으로 판단된다.
- 4) 자동수질분석기 분석항목을 추가한 분석을 통해 직독식 센서 측정 항목만으로 T-P 예측이 가능한지 살펴보았다. 직독식 센서 측정 항목인 WT, pH, EC, DO, Tur, Chl-*a* 항목에 자동수질분석기 분석항목인 TOC, T-N 및 PO₄-P를 각각 추가하여 분석을 진행하였다. 결과적으로 직독식 센서 측정 항목에 비해 자동수질분석기 분석항목이 포함된

분석의 R²는 성능이 더 높게 분석되었으며, MAE, MAPE 및 RMSE는 성능이 더 낮게 분석되었다. 그러나 자동수질 분석기 분석항목을 추가한 분석과 직독식 센서 측정 항목만을 사용한 분석 간의 성능 차이는 크지 않았다. 따라서 시간적, 재정적, 인력적 소요가 적게 발생하는 직독식 센서 측정 방식이 실질적인 이점이 있을 것으로 판단된다.

감사의 글

본 연구는 환경부 수생태계 건강성 확보 기술개발사업의 지원(2021003030005)에 의해 수행되었으며, 이와 같은 지원에 감사드립니다.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Ahn, J.M., Lee, I.J., Jung, K.Y., Kim, J.E., Lee, K.C., Cheon, S.U., and Lyu, S.W. (2016). "Characteristics of trend and pattern for water quality monitoring networks data using Seasonal-kendall, SOM and RDA on the Mulgeum in the Nakdong River." *Journal of Environmental Science International*, Vol. 25, No. 3, pp. 361-371.
- Baik, J.J., Park, J.M., and Choi, M.H. (2020). "Estimation and assessment of total phosphorus and trophic state index estimated from landsat 8 dataset on 2018 at Daecheong reservoir." *Journal of Korean Society of Environmental Engineers*, Vol. 42, No. 9, pp. 415-423.
- Bong, T.H., and Kim, B.I. (2017). "Prediction of ultimate bearing capacity of soft soils reinforced by gravel compaction pile using multiple regression analysis and artificial neural network." *Journal of the Korean Geotechnical Society*, Vol. 33, No. 6, pp. 27-36.
- Choi, J.H., and Kim, Y.K. (2021). "Prediction of the Land price based on deep learning and residual kriging." *Journal of the Korean Data Information Science Society*, Vol. 32, No. 3, pp. 475-485.
- Han, S.H., Lee, E.H., and Kim, H.O. (2012). "Development of software sensor for total phosphorus and nitrogen in various water-body" *Journal of the Korean Society for Environmental Analysis*, Vol. 15, No. 3, pp. 215-223.
- Jung, D.G., Kim, S.D., Kwon, S.Y., Lee, J.Y., Kim, Y.S., Lee, J.Y., Kim, J.K., Kim, S.W., Kong, S.H., and Jung, D.W. (2020). "Development of a portable total-phosphorus monitoring system for preventing eutrophication in Advance." *Journal of Sensor Science and Technology*, Vol. 29, No. 5, pp. 342-347.
- Jung, J.M., Park, S.H., Lee, Y.S., and Gim, J.H. (2019). "The development of infrared thermal imaging safety diagnosis system using pearson's correlation coefficient." *Journal of the Korean Solar Energy Society*, Vol. 39, No. 6, pp. 55-65.
- Jung, S.Y., and Kim, I.K. (2017). "Analysis of water quality factor and correlation between water quality and Chl-a in middle and downstream weir section of Nakdong River." *Journal of Korean Society of Environmental Engineers*, Vol. 39, No. 2, pp. 89-96.
- Kämäri, M., Tarvainen, M., Kotamäki, N., and Tattari, S. (2020). "High-frequency measured turbidity as a surrogate for phosphorus in boreal zone rivers: appropriate options and critical situations." *Environmental Monitoring and Assessment*, Vol. 192, No. 366, pp. 1-20.
- Kim, H.K., Jeong, H.S., and Bae, S.J. (2015). "Deriving water quality criteria of total nitrogen for nutrient management in the stream." *Journal of the Korean Society of Agricultural Engineers*, Vol. 57, No. 3, pp. 121-127.
- Kim, Y.L., Lee, K.H., and Park, S.H. (2023). "Application and evaluation of machine learning techniques for real-time short-term prediction of air pollutants." *Journal of Korean Society for Atmospheric Environment*, Vol. 39, No. 1, pp. 107-127.
- Lee, D.H., Kang, E.T., Joo, J.C., Go, H.W., Ahn, C.M., Bae, Y.H., and Song, K.D. (2022). "Water quality analysis and Chl-a prediction of 15 large-scale freshwater lakes in Korea by multivariate statistical analysis." *Journal of Korean Society of Environmental Engineers*, Vol. 44, No. 12, pp. 589-602.
- Lee, S.I., Cho, D.H., and Lee, M.P. (2018). "Spatializing the pearson's correlation coefficient: An experimental comparison of three relevant techniques." *Journal of the Korean Geographical Society*, Vol. 53, No. 5, pp. 761-776.
- Lim, J.S., Kim, Y.W., Lee, J.H., Park, T.J., and Byun, I.G. (2015). "Evaluation of correlation between Chlorophyll-a and multiple parameters by multiple linear regression analysis." *Journal of Korean Society of Environmental Engineers*, Vol. 37, No. 5, pp. 253-261.
- Valkama, P., and Ruth, O. (2017). "Impact of calculation method, sampling frequency and hysteresis on suspended solids and total phosphorus load estimations in cold climate." *Hydrology Research*, Vol. 48, No. 6, pp. 1594-1610.
- Villa, A., Fölster, J., and Kyllmar, K. (2019). "Determining suspended solids and total phosphorus from turbidity: Comparison of high-frequency sampling with conventional monitoring methods." *Environmental Monitoring and Assessment*, Vol. 191, No. 605, pp. 1-16.
- Zhang, Y., Li, M., Dong, J., Yang, H., van Zwieten, L., Lu, H., Alshameri, A., Zhan, Z., Chen, X., Jiang, X., Xu, W., Bao, Y., and Wang, H. (2021). "A critical review of methods for analyzing freshwater eutrophication." *Water*, Vol. 13, No. 225, pp. 1-20.