

ORIGINAL ARTICLE

소규모 공공하수처리시설의 수질오염물질 배출특성 및 기계학습 기반 TOC 예측

양윤철* · 서희정 · 김선정 · 김하람 · 박지영 · 박주현 · 이기원 · 김성준¹⁾ · 이윤국

광주광역시보건환경연구원 환경연구부, ¹⁾전남대학교 환경에너지공학과

Discharge Characteristics of Water Pollutants and Machine Learning-Based TOC Prediction in Small-Scale Public Sewage Treatment Plants

Yoon-Cheol Yang*, Hee-Jeong Seo, Sun-Jung Kim, Ha-Ram Kim, Ji-Young Park, Ju-Hyun Park, Ki-Won Lee, Seong-Jun Kim¹⁾, Youn-Goog Lee

Department of Environment Research, Health and Environment Research Institute of Gwangju, Gwangju 61954, Korea

¹⁾Department of Environmental and Energy Engineering, Chonnam National University, Gwangju 61954, Korea

Abstract

From 2021 to June 2025, the inflow and effluent of small-scale public sewage treatment plants (Site 1 and Site 2) in the Gwangju area were analyzed. A statistical analysis was conducted to examine the correlation between TOC and related parameters. Machine learning was applied to predict the TOC of the influent, and model selection and result verification were carried out. The average treatment efficiency of the sewage treatment plant Site 1 was 98.9% for BOD, 96.6% for TOC, 99.3% for SS, 85.5% for T-N, 86.9% for T-P, and 99.7% for Total Coliforms. Similarly, Site 2 showed treatment efficiencies of 98.8% for BOD, 98.3% for TOC, 98.2% for SS, 90.9% for T-N, 95.0% for T-P, and 99.8% for Total Coliforms. Both facilities demonstrated overall very high treatment efficiencies. The correlation analysis between TOC and major water pollutants showed that, for Site 1, the inflow had the highest correlations with BOD (0.60) and SS (0.62), while for Site 2, the highest correlations were with BOD (0.52) and SS (0.59). The machine learning model was evaluated by comparing RF, MLR, and SVM. The comparison results showed that RF exhibited the best performance in predicting the inflow TOC for both facilities ($R^2=0.809$ for Site 1, $R^2=0.737$ for Site 2).

Key words: Small scale sewage treatment plant, Water pollutants, Correlation analysis, TOC prediction, Machine learning, Random forest

1. 서론

공공하수처리시설은 생활하수 내 포함된 오염물질을 물리·화학 및 생물학적 공정을 통해 처리하여 수계

로 방류하는 핵심적인 환경 기초시설이다(Shin, 2020). 하수처리 효율을 평가하는 지표로 생물화학적 산소요구량(BOD, biochemical oxygen demand)과 화학적산소요구량(COD, chemical oxygen demand)

Received 24 February, 2026; Revised 23 April, 2026;

Accepted 27 April, 2026

*Corresponding author : Yoon-Cheol Yang, Department of Environment Research, Health and Environment Research Institute of Gwangju, Gwangju 61954, Korea

Phone : +82-62-613-7611

E-mail : ycyang83@korea.kr

© The Korean Environmental Sciences Society. All rights reserved.
© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

은 지난 수십 년간 수질기준의 근간이 되어왔다. 그러나 기존의 COD_{Mn} 방식은 산화률이 약 60~80% 수준에 머물러 난분해성 유기물질을 충분히 반영하지 못하고, 수동 분석에 따른 시간이 많이 소요되며, COD_{Cr} 방식은 높은 산화율에도 분석 과정에서 다량의 유해 폐수가 발생하는 한계가 있다. 반면, 총유기탄소(TOC, total organic carbon)는 시료 내 유기물을 고온으로 연소시켜 발생하는 이산화탄소를 측정함으로써 유기물 내 탄소량을 직접적으로 정량화하는 방식이다. 이는 산화율이 90% 이상으로 매우 높고, 분석 시간이 15분 내외로 짧아 신속한 수질 모니터링 및 실시간 공정 제어가 가능한 장점이 있다. 이러한 장점을 바탕으로 물환경보전법 시행규칙에 따른 공공하수처리시설의 방류수 수질기준 항목 중 COD를 2021년 1월부터 TOC로 변경하였고, 이를 통해 난분해성 유기물질까지 포괄하는 정밀한 수질관리 체계를 구축하였다(Cho et al., 2014; Lee et al., 2021).

현재 우리나라의 경우, 공공하수처리시설은 시설 용량에 따라 체계적으로 관리되고 있으나, 처리 용량 $500 m^3/일$ 미만의 소규모 공공하수처리시설은 물환경보전법상 수질자동측정기기 부착 의무 대상에서 제외되어 상대적으로 관리 사각지대에 놓여 있다. 이러한 소규모 공공하수처리시설은 주로 도심 외곽 및 농어촌 지역이나 하수 관로 미연결 지구에 설치되어 마을 단위의 하수를 직접 처리하여 하천으로 방류하고 있는데 유입수의 성상 및 유량의 시간적·계절적 변동폭이 매우 커 안정적인 처리가 어렵고, 지리적 여건상 전문인력의 상주가 불가능하여 대부분 간헐적 점검에 의존하고 있는 실정이다(Kim, 2020). 또한 대부분의 하수처리시설에서 사용하는 수질분석 장비(TOC, T-N, T-P 등)는 설치 및 유지관리 비용이 높으며, BOD, COD, 총대장균수 등은 실제 운영과 측정시간의 차이가 발생하기 때문에 실시간 공정제어에 즉각적으로 대응이 어렵다(Joo and Lim, 2023).

그리고 하수처리 공정은 유입 유량, pH, 수온, 영양염류 농도 등 수많은 인자가 복합적으로 작용하는 시스템으로 전통적인 선형 회귀 방식으로는 TOC와 같은 동적인 변화를 정확히 예측하기 어려운 특성이 있어 즉각적인 대응이 어려운 환경 데이터의 비선형적인 관계를 규명하고자 최근 기계학습과 딥러닝 기술을 많이 이용하고 있다(Lee et al., 2020; Lee et al., 2022). 특히 기계학습 모델은 각 수질 인자 간의 복잡한 상관관

계를 고도화된 알고리즘을 통해 학습함으로써 물리적 메커니즘이 불확실한 경우에도 높은 예측 정밀도를 나타내는 것으로 알려져 있다(Joo and Lim, 2023). 최근에는 저수지 유해 남조류 발생(2021), 정수장 수질 예측(2024), 하천 탁도변화(2021) 등 수질예측 전반에 걸쳐 기계학습이 도입되고 있으나, 소규모 하수처리시설의 TOC를 대상으로 2021년 이후 축적된 장기 데이터를 분석한 사례는 부족한 실정이다(Lim et al., 2023; Lee et al., 2024).

본 연구에서는 광주지역 소규모 공공하수처리시설에서 수집된 데이터를 바탕으로 최적의 TOC 예측 모델을 구축하고, 그 적용 가능성을 검토하는 것을 목적으로 하였다. 이를 위해 2021년 1월부터 2025년 6월까지 소규모 공공하수처리시설 2개소의 유입수 및 방류수 수질 분석 자료를 이용, 상관 분석을 통해 TOC 농도 변화와 밀접한 관계를 맺고 있는 주요 수질오염물질을 규명하고 통계적 유의성을 고찰하였다. 그리고 다중선형회귀(MLR, multiple linear regression), 랜덤 포레스트(RF, random forest), 서포트벡터머신(SVM, support vector machine) 3개의 기계학습 알고리즘을 적용하여 TOC 예측 모형을 구축하고, 실측값과의 비교를 통해 모델의 예측 성능을 평가하였다. 본 연구의 결과는 향후 소규모 공공하수처리시설 운영 시 에너지 및 약품 비용 절감으로 효율적인 수계 환경 관리 방안 마련에 기여할 수 있을 것으로 기대된다.

2. 재료 및 방법

2.1. 조사대상

본 연구는 광주지역에 위치한 2개의 소규모 공공하수처리시설을 대상으로 수행하였으며, 위치는 Fig. 1과 같다. Site 1은 2014년에 준공되어 운영 중인 시설로 처리 용량은 $260 m^3/일$ 이다. 해당 시설은 처리수를 인근 수계인 석곡천으로 방류하고 있으며, 하수 처리를 위해 FNR (ferrous nutrient removal) 공법을 적용하고 있다. FNR 공법은 생물학적 고도처리 공정에 화학적인(P) 제거 기법을 접목한 기술로, 철염(ferrous salt)계 응집제를 활용하여 하수 내 영양염류를 효과적으로 제거한다. Site 2도 2014년에 준공되어 운영 중인 시설로 처리 용량은 $430 m^3/일$ 이다. 해당 시설은 처리수를 대촌천으로 방류하고 있으며, 하수 처리를 위해 APB-SBR (anoxic process in biofilm-sequencing

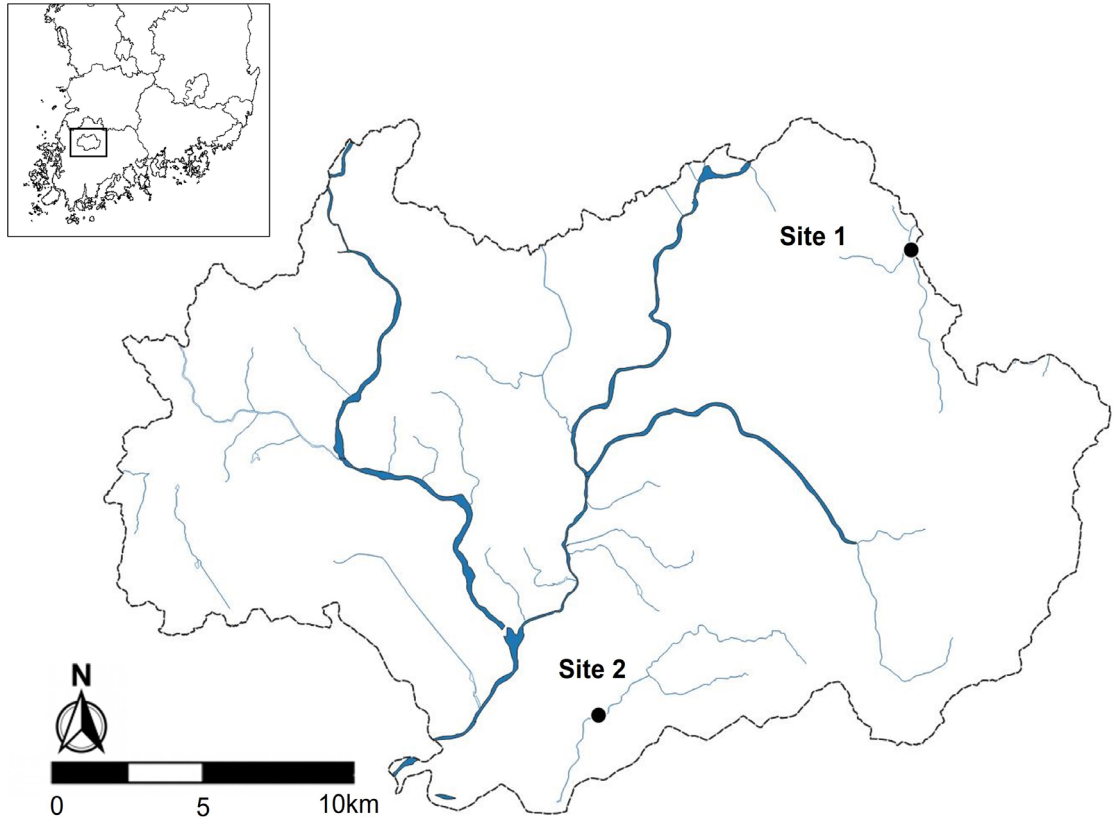


Fig. 1. Sampling sites in sewage treatment plant of Gwangju.

batch reactor) 공법을 적용하고 있다. APB-SBR 공법은 연속 회분식 반응조 내부에 고정상 또는 유동상의 생물막(biofilm)을 형성하여 유기물과 영양염류를 제거하는 고도처리 기술이다.

2.2. 조사항목 및 분석방법

본 연구에 사용한 자료는 2021년 1월부터 2025년 6월까지 광주지역 소규모 공공하수처리시설의 방류수를 대상으로 매주 1회 이상 시료를 채취하여 분석한 결과를 이용하였다. 시료는 각 공공하수처리시설의 유입수와 최종 방류수를 채취하여 분석하였고, 채취된 시료는 멸균된 폴리에틸렌 용기에 담아 현장에서 4℃ 이하의 냉암소에 보관한 후 실험실로 운반하였다. 각 분석항목은 「수질오염공정시험기준」에 따라 TOC는 고온연소산화법을 적용한 TOC 분석기(TOC-L, Shimadzu, Japan)를 이용하여 측정하였으며, BOD는 5일간 배양 후 용존

산소(MutiLab 4010-1Q, YSI, USA)를 측정하는 방식으로 산출하였다. T-N과 T-P는 자외선 흡광광도법 및 분광광도계(QuAAtro39, Bttec, Germany)를 이용하여 정량하였다. 부유물질(SS, suspended solids)는 유리섬유 여과지를 사용한 중량법으로 분석하였고, 총대장균군은 막여과법으로 시험하였다. 항목별로 분석시마다 정밀도(precision), 정확도(accuracy), 방법검출한계(MDL, method detection limit), 정량한계(LOQ, limit of quantitation)에 대하여 조사하였고, 공시료(blank)와 검정곡선검증(CCV, calibration curve verification)까지 수행하여 검증한 결과를 Table 1에 나타내었다.

2.3. 통계분석 및 기계학습

2.3.1. 통계분석

각 항목에 대해 지점별 연평균, 표준편차 등을 산출

하였고, 수질오염물질 항목 간 상관분석을 실시하였다. 데이터의 정규분포 여부와 변수 간 비선형적 관계를 고려하여 스피어만 상관분석(spearman correlation analysis)를 적용하였으며, 집단 간의 유의미한 차이를 정밀하게 비교하고자 비모수검증법(non-parametric test)을 병행하였다. 모든 통계분석 및 시각화는 python으로 수행하였다.

2.3.2 기계학습

기계학습은 주어진 데이터에서 통계적 패턴과 규칙성을 학습하여 예측 모델을 구축하는 기법으로, 전체 데이터를 학습용과 평가용으로 분리하여 학습 데이터를 통해 알고리즘을 훈련시키고 평가데이터를 통해 성능을 검증한다. 이러한 과정을 여러 알고리즘에 반복 적용한 뒤, 가장 우수한 예측 성능을 확보한 모델을 최종 선택하는 방식으로 활용된다.

MLR은 두 개 이상의 독립변수와 하나의 종속변수 간 선형적 관계를 규명하기 위한 통계적 기법이다. 이는 단순회귀분석의 확장형으로, 여러 독립변수의 영향을 동시에 고려하여 종속변수의 변화를 설명하고 예측하는 데 효과적이다. 각 독립변수의 선형적 결합을 통해 최적의 회귀 계수를 추정하며, 다양한 변인이 복합적으로 작용하는 수질 데이터 분석 및 예측 모델 구축에 널리 활용된다.

RF는 다수의 의사결정나무를 개별적으로 학습시킨 후, 그 예측치들을 결합하여, 최종 결과를 산출하는 앙상블 학습 기법이다. 모델은 각 트리 생성 시 변수를 무작위로 선택함으로써 트리 간의 상관성을 최소화하며, 이를 통해 단일 모델에서 발생하기 쉬운 과적합 문제를 효과적으로 방지한다. RF는 비선형적 관계나 변수 간의 복잡한 상호작용을 분석할 때 별도의 가정이 필요치 않아 수질 데이터와 같은 복합적인 변인 분석에 유리하다. 특히 배깅(bagging) 기술을 통해 모델의 분산을 감소시켜 일반화 성능을 높이며, 주요 하이퍼파라미터인 mtry 등을 조절하여 편향과 분산 간의 균형을 최적화할 수 있다. 이러한 특성 덕분에 데이터 전처리 복잡도가 낮으면서도 높은 예측 정확도를 확보할 수 있다는 장점이 있다(Kim et al., 2021; Lee et al., 2024).

SVM은 데이터 간의 마진(margin)을 최대화하는 초평면(hyperplane)을 설정하여 데이터의 패턴을 분석하는 지도학습 알고리즘이다. 회귀분석에서는 설정된 오차 허용 범위 내에서 최적의 회귀함수를 결정하

며, 전체 데이터 중 경계 형성에 기여하는 핵심 샘플인 서포트 벡터만을 활용하여 모델을 구축함으로써 계산 효율성과 일반화 성능을 동시에 확보한다(Hyun, 2025).

본 연구에서는 소규모 공공하수처리시설에서 TOC 농도를 예측하고 이에 영향을 미치는 주요 변수를 평가하고자 하였다. 이를 위해 BOD, SS, T-N, T-P, 총대장균군을 독립변수, TOC 농도를 종속변수로 선정하였다. 예측 모델 구축을 위해 이상치 제거, 정규화 등 데이터 전처리 후 학습데이터 70%, 평가데이터 30%로 데이터를 구분하였고, 기계학습 기법인 MLR, RF, SVM을 통해 TOC 농도 예측을 수행하였다.

2.3.3. 데이터 전처리 및 검증방법

실험분석을 통해 확보한 수질 데이터 내에 포함된 이상치는 예측모델의 학습 과정에서 왜곡을 일으켜 신뢰도를 저하시키는 주요 요인이 된다. 따라서 모델의 성능과 예측 정확도를 확보하기 위해 IQR (interquartile range)를 활용해 데이터의 이상치를 식별하고 제거하였다. 또한, 수질 항목별로 측정값의 범위와 단위가 상이함에 따라 특정 변수가 모델학습에 지배적인 영향을 미치는 편향성을 방지하고자 Z-score 정규화를 수행하였다(Joo and Lim, 2023).

본 연구에서는 소규모 시설의 TOC 농도를 정확하게 예측하기 위해 MLR, RF, SVM의 기계학습 모델을 구축하였고, 모델의 예측성능을 정량적으로 평가하기 위한 지표로는 평균제곱근오차(RMSE, root mean square error)와 결정계수(R^2 , coefficient of determination)를 사용하였다(Lee and Kim, 2021; Park, 2021). 평가지표의 특성을 살펴보면, RMSE는 실측값과 모델 예측값 간의 차이를 나타내는 잔차의 크기로, 이 값이 0에 가까울수록 모델이 실제 수질변화를 오차 없이 반영하고 있음을 의미한다. 결정계수(R^2)는 모델이 전체 데이터의 분산을 설명하는 정도를 나타내며, 1에 가까울수록 실측값과 예측값의 일치도가 높은 우수한 성능의 모델로 판단할 수 있다.

3. 결과 및 고찰

3.1. 소규모 공공하수처리시설의 수질특성

3.1.1. 연도별 수질오염물질 특성

Table 1. Quality control of measuring equipment

Item	MDL (mg/L)	LOQ (mg/L)	R ²	Precision (%)	Accuracy (%)	
BOD	-	-	-	1.4	96.2	
TOC	TC	0.05	0.15	1.0000	0.5	103.9
	IC	0.02	0.07	0.9998	1.3	99.8
	NPOC	0.05	0.15	1.0000	1.6	100.3
	T-N	0.042	0.132	0.9995	6.5	98.0
T-P	0.002	0.008	0.9982	8.2	100.0	

Precision ± 25%, Accuracy 75 ~ 125%

Table 2. Annual discharge characteristics of sewage treatment plants

		(unit : mg/L, Total Coliforms cfu/mL)					
Type	Year	BOD (Mean±SD)	TOC (Mean±SD)	SS (Mean±SD)	T-N (Mean±SD)	T-P (Mean±SD)	Total Coliforms (Mean±SD)
Site 1 inf.	2021	198.3±304.5	115.2±112.6	299.4±420.4	41.20±15.64	4.530±1.825	2.0×10 ⁵ ±1.5×10 ⁵
	2022	114.7±84.6	128.1±58.1	261.1±150.5	47.31±14.77	5.186±1.618	2.4×10 ⁵ ±1.9×10 ⁵
	2023	160.9±85.5	127.5±105.2	500.2±511.1	41.70±19.45	4.789±2.351	3.1×10 ⁵ ±3.7×10 ⁵
	2024	187.7±137.1	120.4±88.7	318.5±345.9	34.22±13.20	4.274±1.908	3.3×10 ⁵ ±2.1×10 ⁵
	2025	283.6±203.7	133.6±103.2	497.4±665.7	33.42±16.49	5.146±2.544	2.5×10 ⁵ ±1.7×10 ⁵
Site 1 eff.	2021	3.2±2.1	4.0±1.7	2.7±1.7	7.71±3.63	0.711±0.770	2.0×10 ³ ±6.0×10 ³
	2022	2.5±1.4	4.3±1.7	3.0±1.4	5.99±2.76	0.627±0.743	5.0×10 ² ±5.4×10 ²
	2023	1.6±1.5	4.0±1.7	2.6±1.6	5.02±2.53	0.490±0.566	9.5×10 ² ±2.0×10 ³
	2024	1.3±0.8	4.5±1.9	2.3±0.9	5.11±2.26	0.688±0.508	2.2×10 ² ±2.6×10 ³
	2025	2.0±1.3	4.8±1.4	1.7±0.9	5.01±2.33	0.439±0.280	1.8×10 ² ±2.0×10 ²
Removal avg. (%)		98.9	96.6	99.3	85.5	86.9	99.7
Site 2 inf.	2021	231.6±194.7	153.0±164.8	366.2±550.5	36.88±9.75	4.554±1.570	1.1×10 ⁵ ±1.2×10 ⁵
	2022	170.5±87.4	119.3±32.7	155.1±63.7	39.30±5.59	4.599±1.149	1.2×10 ⁵ ±1.7×10 ⁵
	2023	186.8±72.1	112.0±36.8	173.2±128.2	42.18±16.95	4.677±2.289	9.7×10 ⁴ ±7.6×10 ⁴
	2024	153.8±42.7	101.3±28.5	121.5±59.5	32.66±11.69	4.070±1.815	1.4×10 ⁵ ±1.3×10 ⁵
	2025	168.8±36.5	93.3±18.4	98.2±33.4	25.95±7.63	3.912±0.955	1.2×10 ⁵ ±3.8×10 ⁴
Site 2 eff.	2021	2.0±1.0	1.4±0.7	3.3±2.4	3.92±1.35	0.286±0.489	4.0×10 ¹ ±5.3×10 ¹
	2022	2.3±1.2	1.3±0.8	3.5±2.1	4.05±1.62	0.265±0.389	3.1×10 ² ±3.9×10 ²
	2023	2.8±2.5	2.5±2.1	3.7±3.2	3.39±1.60	0.224±0.151	2.8×10 ² ±6.3×10 ²
	2024	1.3±1.2	1.7±0.5	2.1±1.1	2.64±1.07	0.172±0.232	9.0×10 ¹ ±1.6×10 ²
	2025	2.3±1.6	3.2±1.2	1.7±0.8	4.90±2.67	0.073±0.040	4.3×10 ¹ ±5.3×10 ¹
Removal avg. (%)		98.8	98.3	98.2	90.9	95.0	99.8

※ Year average precipitation : 2021 108.6 mm, 2022 64.2 mm, 2023 176.3 mm, 2024 107.6 mm, 2025 80.7 mm

Table 2는 2021년부터 2025년 6월까지 광주지역 소규모 공공하수처리시설 두 지점(Site 1, Site 2)의 연도별 평균 수질 분석결과를 나타낸 것이다. 조사기간 내내 유입수의 수질 변동폭이 매우 컸음에도 불구하고 연구기간 동안 방류수 수질은 공공하수처리시설 방류수 수질기준을 안정적으로 준수하며 높은 처리 효율을 유지한 것으로 조사되었다.

Site 1의 유입수 수질 특성을 연도별로 살펴보면, 2025년에 BOD 283.6 mg/L, TOC 133.6 mg/L로 조사 기간 중 가장 높은 평균 농도를 기록하였다. 이는 해당 연도의 연평균 강수량이 80.7 mm로, 강수량이 풍부했던 2023년(176.3 mm) 대비 약 54% 감소함에 따라 유입 오염물질의 희석 효과가 크게 저하되었기 때문으로 판단된다. 특히 SS는 2023년 500.2 mg/L와 큰 표준편차는 해당연도의 강수량이 높아 시설 내 불규칙한 유입 변동으로 보여지며, 2022년에는 T-N 47.31 mg/L, T-P 5.186 mg/L로 농도가 가장 높았다.

이러한 높은 수질 변동성에도 불구하고 방류수의 TOC는 4.0~4.8 mg/L이며, SS 1.7~3.0 mg/L로 좁은 편차 내에서 안정적으로 처리되고 있어 시설이 안정적으로 운영되고 있음을 확인하였다.

Site 2에서는 연도별로 유입수의 수질 편차가 확인되었으나, 유입 TOC는 2021년 153.0 mg/L에서 2025년 93.3 mg/L로 감소하였으며, SS 또한 2021년 366.2 mg/L에서 2025년 98.2 mg/L로 크게 낮아져 기간이 경과함에 따라 전반적인 오염 부하가 점진적으로 낮아지는 경향을 보였다.

연구기간 동안 Site 1의 평균 처리효율은 BOD 98.9%, TOC 96.6%, SS 99.3%, T-N 85.5%, T-P 86.9%, 총대장균군 99.7%였으며, Site 2 또한 BOD 98.8%, SS 98.2%, T-N 90.9%, T-P 95.0%, 총대장균군 99.8%로 두 시설 모두 전반적으로 매우 높은 처리효율을 보였다.

Kim(2020)의 강원도 소재 250 m³/일 이하의 소규모 공공하수처리시설의 경우 BOD 96~98%, SS 96~99%, T-N 60~76%, T-P 66~87%, 총대장균군 99% 이상의 제거효율을 나타내었고, Kim(2020)의 함양군 소규모 하수처리시설도 평균 BOD 97.4%, SS 97.0%, T-N 75.3%, T-P 74.1%로 나타났다. 또한 Cho(2021)의 연구에서 Site 1과 동일공법(FNR)으로 설치한 시설에서는 BOD 92.3%, SS 95.8~99.5% T-N 87.4%, T-P 76.0~94.7%로 Site 2와 유사공법(SBR)으로 적용된 시

설에서 BOD 98.2%, SS 98.2%, T-N 75.0%, T-P 81.0%로 나타났다. 본 연구의 대상 시설의 검사항목 결과는 기존 연구결과와 비슷하거나 일부 항목에서는 더 높은 처리효율로 조사되어 연구 대상시설이 잘 운영되고 있는 것으로 판단되었다.

Fig. 2는 연구기간 동안 두 지점에서 측정된 TOC 농도의 연도별 수질오염물질 특성을 나타낸 것이다. 유입수 TOC는 연도별로 변동을 보였으며, 특히 Site 1은 2021년 115.2±112.6 mg/L, Site 2는 153.0±164.8 mg/L로 높은 변동성을 보였고, 이후 점차 감소하였다. 방류수 TOC는 두 지점 모두 14 mg/L 이하로 유지되었으며, 이는 하수처리공정을 통해 유기물이 안정적으로 처리되고 있음을 보여준다.

또한 2017~2019년 전국 공공하수처리시설 방류수 기준 50~500 m³/일 미만 90개소의 평균 TOC 농도는 7.5(±10.5) mg/L로 평균 Site 1은 4.3 mg/L, Site 2는 2.1 mg/L로 대상 시설보다 우수한 성능을 나타냈다(Son et al., 2021).

3.1.2. 계절별 수질오염물질 특성

공공하수처리시설의 유입수는 유기물 및 오염물질 농도의 범위가 크고 다양하며, 사계절에 따른 온도 변화와 강수량 등의 외부요인에 영향을 받는다(Kim, 2019). 본 연구에서는 2021년부터 2025년 6월까지 데이터를 통합하여 계절별 유입수 및 방류수 수질 특성을 분석하였다. 연구 기간 내 연도별 강수량 편차에 따른 유입 농도의 변동성에도 불구하고, 전 기간에 걸쳐 계절적 변동 패턴을 파악하고자 하였다. 계절 구분은 봄(3~5월), 여름(6~8월), 가을(9~11월), 겨울(12~2월)로 하였으며, 계절별 유입수 및 방류수 수질오염물질 특성을 Table 3에 나타냈다. Site 1의 유입수는 겨울철에 BOD 194.2 mg/L, TOC 139.7 mg/L로 가장 높은 평균값을 보였으며, 여름철에는 BOD 117.9 mg/L, TOC 94.4 mg/L로 상대적으로 낮게 나타났다. Site 2는 또한 겨울철에 BOD 219.7 mg/L, TOC 114.1 mg/L로 다른 계절에 비해 높게 측정되었다. 이는 강수량, 기온 등 외부 환경요인에 따른 계절적 영향으로 판단된다(Kim 2019). 방류수의 경우 두 지점 모두 전 계절에 걸쳐 BOD 평균 1.6~2.6 mg/L, TOC 평균 1.7~4.6 mg/L로 측정되었으며, 계절 간 변동 폭이 크지 않은 것으로 조사되었다. Kim(2019)의 연구와 동일하게 두 지점 모두 유입수 BOD 농도는 여름에 낮게 측정되었으며, 강수량의 증

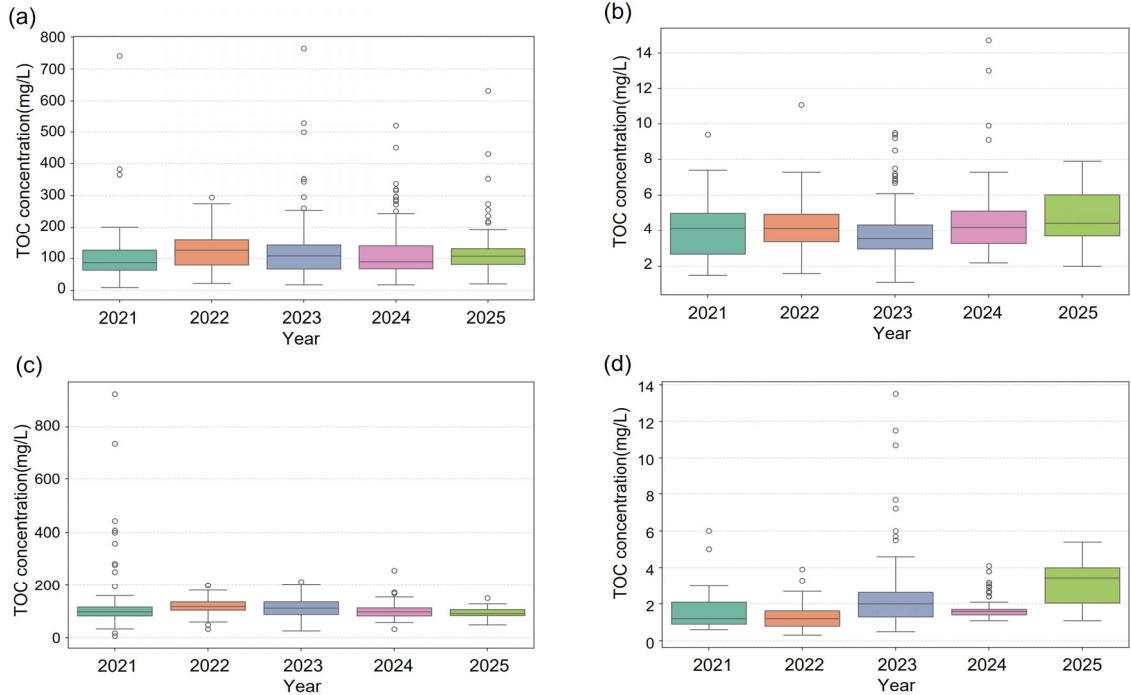


Fig. 2. Annual variations in TOC concentrations at sewage treatment plants (a) Site 1 influent (b) Site 1 effluent (c) Site 2 influent (d) Site 2 effluent.

Table 3. Seasonal discharge characteristics of sewage treatment plants

		(unit : mg/L, Total Coliforms cfu/mL)					
Type	Season	BOD (Mean±SD)	TOC (Mean±SD)	SS (Mean±SD)	T-N (Mean±SD)	T-P (Mean±SD)	Total Coliforms (Mean±SD)
Site 1 inf.	Spring	200.1±232.2	126.4±86.1	391.2±477.0	42.92±15.44	4.891±1.905	2.6×10 ⁵ ±1.9×10 ⁵
	Summer	117.9±99.2	94.4±75.7	285.7±344.7	30.01±13.39	3.594±1.755	2.5×10 ⁵ ±1.9×10 ⁵
	Autumn	161.9±114.6	133.2±110.5	443.9±464.5	37.25±16.56	4.432±1.739	3.8×10 ⁵ ±3.9×10 ⁵
	Winter	194.2±157.3	139.7±98.6	348.5±325.6	50.01±14.53	5.662±2.154	2.3×10 ⁵ ±2.2×10 ⁵
Site 1 eff.	Spring	2.2±1.6	4.6±1.9	3.1±1.5	6.15±2.82	0.514±0.342	1.4×10 ³ ±2.8×10 ³
	Summer	1.7±1.6	4.2±2.0	2.4±1.4	5.51±2.77	0.697±0.711	1.3×10 ³ ±4.6×10 ³
	Autumn	1.6±1.3	3.9±1.2	2.2±0.9	4.81±1.87	0.682±0.694	3.3×10 ² ±6.1×10 ²
	Winter	2.2±1.6	3.9±1.5	2.8±1.6	6.27±3.60	0.572±0.666	8.5×10 ² ±2.0×10 ³
Site 2 inf.	Spring	184.8±82.9	126.6±97.8	213.0±302.5	39.03±8.17	4.094±1.523	1.3×10 ⁵ ±1.1×10 ⁵
	Summer	157.7±81.0	123.0±100.9	237.0±371.9	34.60±13.77	4.112±1.716	7.8×10 ⁴ ±6.7×10 ⁴
	Autumn	162.0±64.1	103.7±46.1	151.6±160.5	36.07±15.14	4.426±1.839	9.6×10 ⁴ ±6.7×10 ⁴
	Winter	219.7±154.7	114.1±34.6	141.3±56.5	41.37±13.87	5.194±2.216	1.9×10 ⁵ ±2.0×10 ⁵
Site 2 eff.	Spring	2.6±2.3	2.1±1.1	4.5±3.1	3.48±1.49	0.176±0.163	2.3×10 ² ±3.0×10 ²
	Summer	1.6±2.2	1.7±1.1	2.1±3.1	3.10±1.49	0.220±0.125	1.2×10 ² ±3.0×10 ²
	Autumn	1.6±2.1	1.8±1.0	2.3±2.7	3.25±1.48	0.267±0.126	6.6×10 ¹ ±2.9×10 ²
	Winter	2.5±2.1	2.1±1.0	3.2±2.7	3.17±1.51	0.193±0.128	3.8×10 ² ±2.9×10 ²

* Seasonal average precipitation : Spring 79.8 mm, Summer 256.7 mm, Autumn 84.0 mm, Winter 30.3 mm

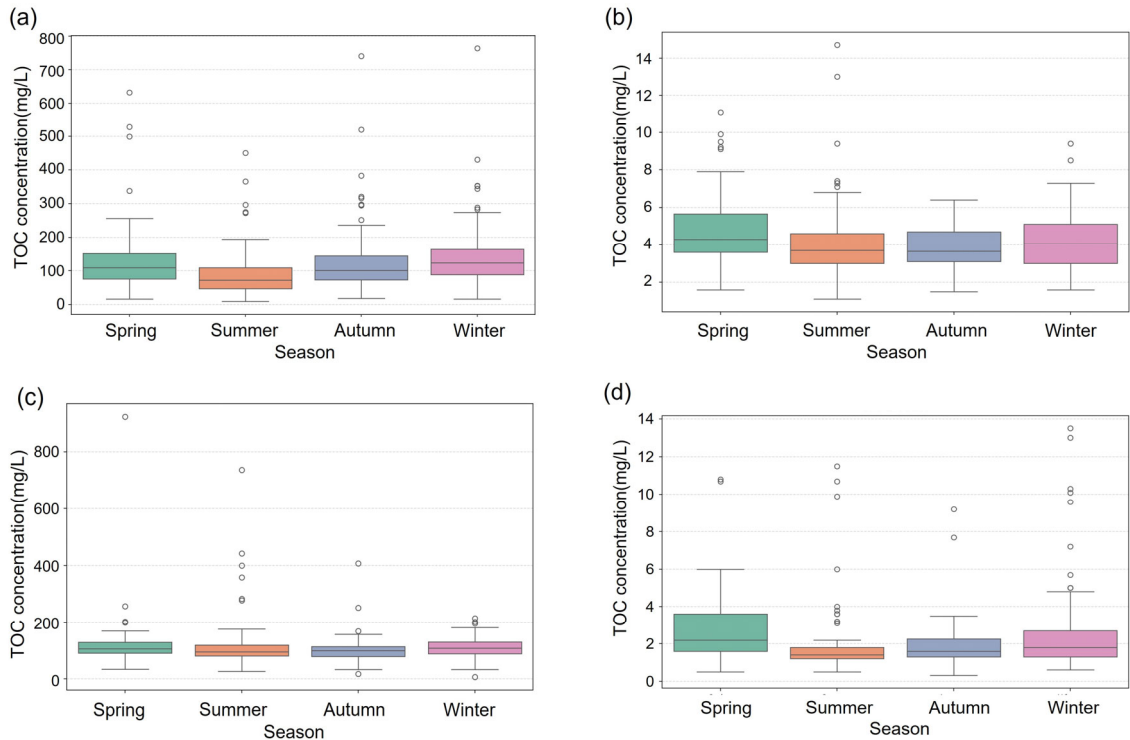


Fig. 3. Seasonal variations in TOC concentrations at sewage treatment plants (a) Site 1 influent (b) Site 1 effluent (c) Site 2 influent (d) Site 2 effluent.

가로 인한 외부 유입수가 증가한 영향으로 판단된다. 특히, 총대장균군은 유입수에서 Site 1이 가을에 최대 3.8×10^6 개/mL로 계절적으로 차이가 있었으나, 방류수에서는 $6.6 \times 10^1 \sim 1.4 \times 10^3$ 개/mL 범위로 수질기준 이하로 조사되어 미생물 처리효율이 지속적으로 유지되는 것으로 나타났다. 이러한 결과로 유입수의 경우 계절적 요인에 의해 오염물질 농도가 크게 변하더라도, 방류수에서는 처리 공정이 안정적으로 운영되고 있음을 확인할 수 있었다.

연구기간 중 Site 1 및 Site 2에서 관측된 TOC 농도의 계절별 특성은 Fig. 3과 같다. Site 1의 유입수 TOC 농도는 계절에 따라 뚜렷한 변동성을 보였으나, 다른 계절에 비해 여름에 평균 94.4 mg/L로 가장 낮게 측정되었다. Site 1의 방류수 TOC 농도는 계절별로 큰 차이가 없이 약 3.9~4.6 mg/L로 조사되어 공정이 일정 수준의 처리 효율을 유지하는 것으로 나타났다.

Site 2 유입수의 TOC 농도는 평균 103.7~126.6 mg/L로 계절별 농도 차이가 크지 않았으나, 주로 봄과

여름에 고농도로 일부 측정되기도 하였다. Site 2 방류수의 TOC 농도는 전 계절에서 평균 1.7~2.1 mg/L로 Site 1에 비해 낮은 농도를 유지하여 전반적으로 안정적으로 처리되고 있는 것으로 조사되었다.

3.2. TOC와 수질오염지표 간 상관성 분석

TOC 농도에 영향을 미치는 주요 영향인자를 파악하기 위해 항목별 스피어만(spearman) 상관성 분석을 하였고, 그 결과를 Fig. 4에 나타내었다.

Site 1에서 TOC와 주요 수질오염 항목 간의 상관성 분석 결과 유입수의 경우, BOD (0.60), SS (0.62), T-N (0.52), T-P (0.57)로 양의 상관관계를 나타냈으며, 이는 유기물과 영양염류가 동시에 증가하는 하수 특성을 나타낸다고 할 수 있다. 반면, 총대장균군 (0.17)과는 낮은 상관성을 보여, 미생물 지표가 유기물 농도나 영양염류 농도와 상관성이 낮은 것으로 분석되었다. 방류수에서는 BOD (0.56), SS (0.32), T-N

(0.39), T-P (0.38), 총대장균군 (0.38)로 유입수에 비해 상관관계수가 낮게 나타났다.

Site 2에서는 유입수의 경우 BOD (0.52), SS (0.59), T-N (0.48), T-P (0.39)로 양의 상관관계를 나타내어 Site 1과 유사한 경향을 나타냈으며, 총대장균군 (0.06)과는 낮은 상관성을 보였다. 반면에 방류수에서는 BOD (0.38), T-N (0.46) 양의 상관관계를 나타냈으며, SS (0.16), T-P (0.15), 총대장균군 (0.09)로 낮은 상관성을 보였다.

Son et al.(2021)의 연구에서는 2017~2019년 전국 공공하수처리시설 방류수 기준 50~500 m³/일 미만 100개소에서 TOC와 BOD (0.68)와 상관성이 있다고 발표하였고, Joo and Lim(2023)의 연구에서도 공공하수처리시설의 COD 항목과 SS (0.37), T-N (0.56), T-P (0.18)로 나타내고 있어 본 연구 결과와 유사하였다.

다만, 항목 간 상관성은 유입수 대비 방류수에서 각 인자 간의 상관관계수가 전체적으로 낮게 나타났으며, 이는 공공하수처리시설의 고도처리 공법(FNR 및 APB-SBR) 모두 유기물이나 영양염류를 적정하게 제거할 수 있는 공법으로 확인할 수 있었으나, 방류수의 낮은 농도 분포로 인해 데이터의 변별력이 감소한 점이 상관관계수가 낮아진 것으로 보인다(Fig. 4). 그리고 본 연구에 활용된 고도처리 공법(FNR 및 APB-SBR)의 방류수 특성을 분석하였는데 Site 1은 철염 투입을 통한 화학적 응집 공정의 영향으로 T-P과 유기물 간의 상관성이 유입수 대비 크게 낮아졌으며, Site 2 또한 생물학적 고도처리 과정을 통해 입자성 물질과 T-P 성분이 크게 저감 된 것으로 보여진다.

그리고 유입수 수질항목 중 T-N과 T-P의 상관관계는 Site 1 (0.84), Site 2 (0.70)으로 양의 상관성이 나타났으며, Jun(2020)의 연구에서 금강수계로 방류되는 5개의 소규모 공공하수처리시설에서의 T-N과 T-P의 상관성이 0.65~0.75로 본 연구와 유사하게 나타났다.

3.3. 기계학습 모델을 활용한 TOC 예측성능 비교

공공하수처리시설의 경우 다양한 계절적 영향을 받기 때문에 2021~2025년 6월까지의 매주 시험한 결과인 360개 이상의 자료로 활용하였으며, 데이터를 학습시키는 과정에서 각각의 알고리즘에서 반복 시행 횟수에 따른 RMSE 결과와 R²값을 산출하였다. Fig. 5는 각

지점별 기계학습 모델의 TOC 예측 결과이며, 모델별 예측 정밀도를 검증하기 위한 평가지표(R², RMSE)의 분석 결과를 Table 4에 나타냈다.

공공하수처리시설 유입수 TOC 농도의 예측을 위해 기계학습을 적용한 결과 Site 1의 경우 RF는 R²=0.809, RMSE=0.409, MLR은 R²=0.385, RMSE=0.733, SVM은 R²=0.424, RMSE=0.709로 나타나 RF 모델이 다른 두 모델보다 더 높은 정확도를 보였다. Site 2의 경우 RF는 R²=0.737, RMSE=0.385, MLR은 R²=0.295, RMSE=0.625, SVM은 R²=0.250, RMSE=0.645로 마찬가지로 RF에서 다른 두 모델보다 더 높은 정확도를 보였다.

공공하수처리시설 유입수 예측은 대부분 COD 결과를 이용하였고, Joo and Lim(2023)의 검증결과에서 평가지표(RMSE)는 0.3~0.6의 비율로 나타나 Barahi et al.(2025)의 연구에서 R²은 0.83으로 본 연구의 모델과 유사하였다.

TOC 농도 예측을 위해 세 가지 모델의 성능을 비교 분석한 결과 RF 모델이 가장 우수한 예측 정밀도와 최적 적합도를 나타내었다. 이에 따라 최종 선정된 RF 모델을 활용하여 TOC 예측에 기여하는 주요 수질인자의 중요도를 분석하였으며, 그 결과는 Fig. 6과 같다.

Site 1과 Site 2 모두에서 SS와 T-N이 가장 높은 기여도를 나타내는 것으로 확인되었다. Site 1에서는 SS (0.263), T-N (0.255), BOD (0.193) 순으로 중요도가 높게 나타났으며, Site 2 역시 SS (0.275), T-N (0.266), BOD (0.208) 순으로 유의미한 영향력을 보였다. 그러나 두 시설 모두에서 총대장균군수의 중요도는 0.1 이하로 산출되어 TOC 예측에 미치는 상대적 영향력은 낮은 것으로 분석되었다. 이와 같은 결과는 향후 기계학습 기반의 TOC 모니터링 모델 구축 시 입자성 물질과 영양염류 인자를 우선적인 독립변수로 고려하는 것이 예측 정밀도 향상에 효과적이라고 판단하였고, 두 시설 모두에서 SS와 T-N가 높은 TOC 예측 인자로 도출되었으며, 비선형적 예측 기여도의 차이로 보여진다. 또한 TOC와 BOD 항목은 유의미한 상관관계가 확인되었으며, RF 모델은 이러한 단순 선형 관계를 넘어 변수 간의 복잡한 상호작용을 학습함으로써 BOD보다 높은 데이터 분별력을 나타냈다. 이는 RF 모델이 TOC의 비선형적인 변동폭을 설명하는 데 있어 더욱 효과적이라 할 수 있다고 판단하였고, 수질검사 항목끼리의 다차원적이고 비선형적인 관계를 효과적으로

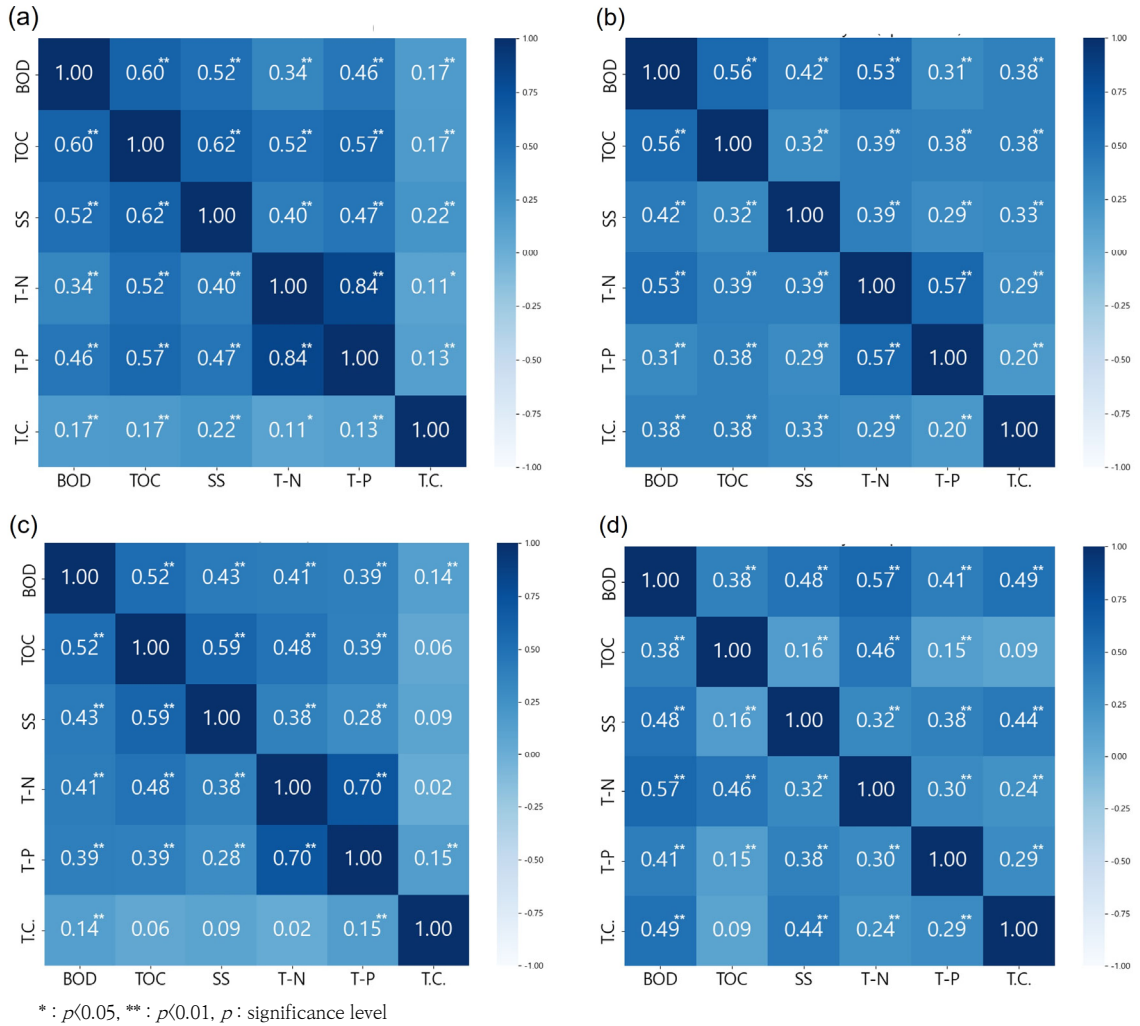


Fig. 4. Correlation analysis of influent and effluent in sewage treatment plants (a) Site 1 influent (b) Site 1 effluent (c) Site 2 influent (d) Site 2 effluent.

Table 4. Model based machine learning results for influent at sewage treatment plants

Type	RF		MLR		SVM	
	R ²	RMSE	R ²	RMSE	R ²	RMSE
Site 1	0.809	0.409	0.385	0.733	0.424	0.709
Site 2	0.737	0.385	0.295	0.625	0.250	0.645

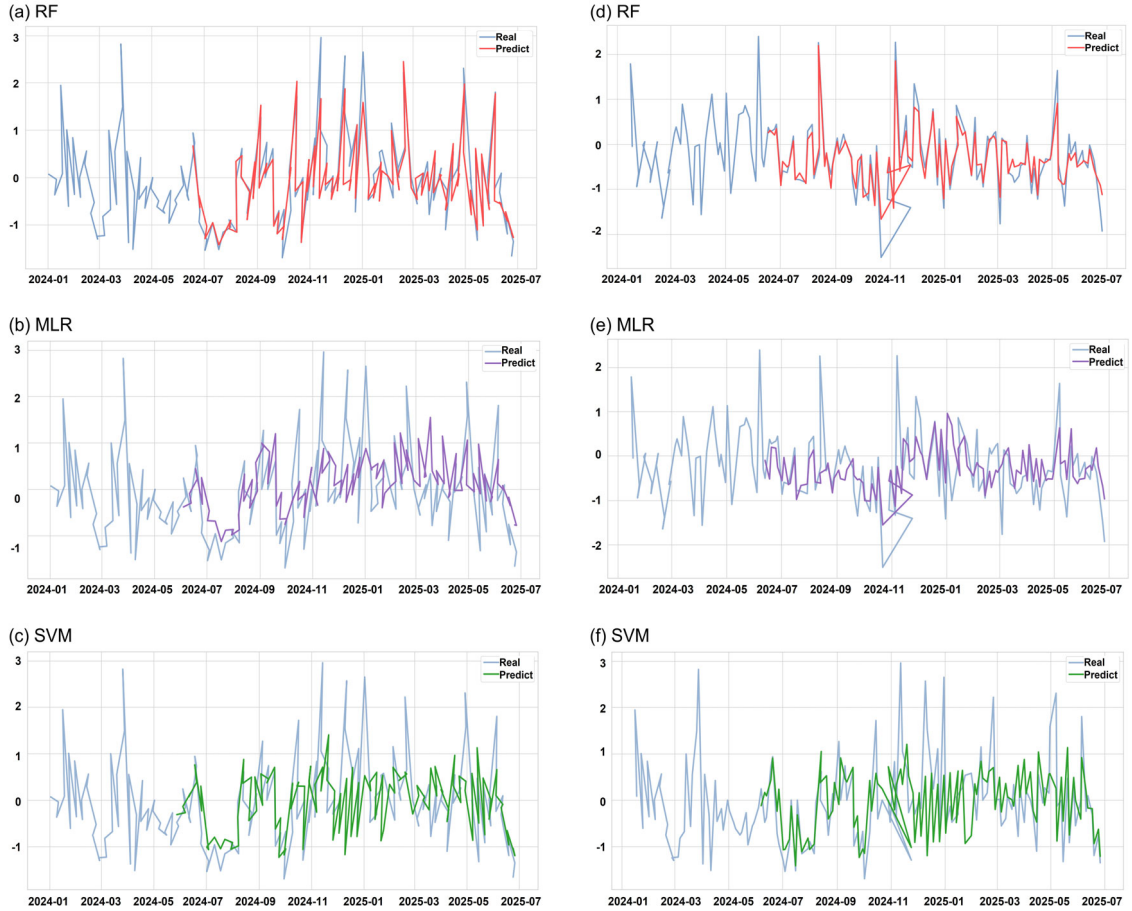


Fig. 5. Machine learning-based TOC prediction results for sewage treatment plants (a) RF Site 1 (b) MLR Site 1 (c) SVM Site 1 (d) RF Site 2 (e) MLR Site 2 (f) SVM Site 2.

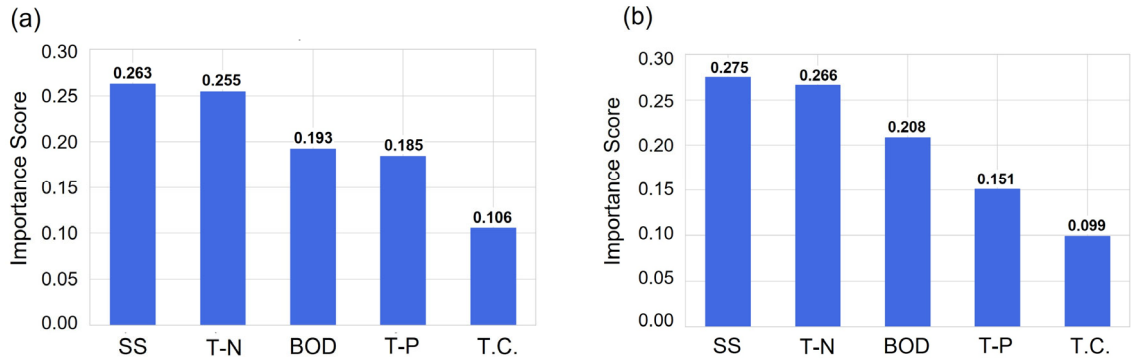


Fig. 6. Feature importance analysis for TOC prediction using RF (a) Site 1 (b) Site 2.

반영할 수 있어 본 연구 대상과 유사한 소규모 공공하수처리시설의 수질 예측에 적용 가능성이 있음을 시사한다고 할 수 있다.

그리고 MLR은 RF와 비교하면 상대적으로 낮은 예측 성능을 보였는데 유입수의 TOC 변화가 선형적인 구조로 설명되기 어렵다는 점을 보여준다. 또한 최종적으로 수질 데이터와 같이 데이터 분산이 크고 차원 및 규모의 차이가 클 때 RF와 같은 의사결정 구조의 적용이 적절하다고 판단된다.

4. 결론

본 연구는 2021년부터 2025년 6월까지 광주지역 소규모 공공하수처리시설 2개소(Site 1, 2)를 대상으로 수질 배출 특성을 고찰하고, 기계학습 기반의 TOC 예측 모델을 구축하여 다음과 같은 결론을 얻었다.

- 1) 조사기간 동안 두 시설 모두 유기물(BOD, TOC), 부유물질(SS), 영양염류(T-N, T-P) 및 미생물 항목에서 85.5~99.8%의 높은 평균 처리 효율을 기록하였다. 특히 계절에 따른 유입 수질 변동에도 불구하고 방류수질은 연중 안정적으로 수질 기준 이내를 유지하여 공법의 신뢰성을 확인하였다.
- 2) 검사항목의 상관성 분석 결과로는 유입수의 TOC가 BOD 및 SS와 높은 상관관계를 보였으나, 처리 공정을 거친 방류수에서는 농도 저감에 따라 항목 간 변동성이 축소되며 상관성이 낮아지는 경향을 나타냈다.
- 3) 기계학습을 통한 모델의 평가는 RF 모델이 두 시설의 유입수 TOC 예측에서 가장 우수한 성능을 보였으며(Site 1: $R^2=0.809$, Site 2: $R^2=0.737$), 비선형성을 효과적으로 반영하는 모델로 확인되었다. 또한 RF 모델에서 변수 중요도 분석에서 SS와 T-N이 TOC 변동의 핵심 인자로 도출되었다.

본 연구는 광주지역 소규모 공공하수처리시설의 배출특성을 규명하고, RF 모델의 높은 TOC 예측 성능과 주요 영향인자를 확인하였다. 향후 지속적인 모니터링을 통해 수질 데이터뿐만 아니라 처리장의 조건 및 운전 특성, 유입수의 특징, 측정 유량, 온도, pH, 기상 자료 등의 외부 변수를 추가로 확보하여 모델에 적용한다면, TOC 예측 성능을 개선할 수 있을 것으로 판단된다. 이러한 데이터 기반의 지능형 운영 체계 구축은 소규모

하수처리 공정의 효율성을 극대화하여 운영 비용을 절감하는 동시에, 방류수 수질 기준을 안정적으로 준수함으로써 공공수역의 수질 보전에 크게 기여할 수 있을 것으로 판단된다.

감사의 글

본 논문은 2025년 환경부의 재원으로 국립환경과학원의 지원(과제번호 NIER-2025-01-03-002)과 광주광역시 보건환경연구원의 지원(연구지원 및 역량강화)을 받아 수행하였습니다.

REFERENCES

- Barahi, S., Azizi, A., Taky S., Belhamidi S., 2025, Machine learning algorithms in wastewater technology machine learning algorithms in wastewater technology, J. Water Land Dev., 66, 137-151.
- Cho, N. K., 2021, A Study on the effective management of small-scale village sewage treatment process : Based on the results of test run, M.S. Thesis, Mokpo National University, Jeollanam-do, Korea, <https://www.riss.kr/link?id=T15915446>
- Cho, Y. B., Oh, Y. K., Shin, D. C., Park, C. H., 2014, Distribution of total organic carbon and correlations between organic matters of sewage treatment plants, J. Korean Soc. Environ. Anal., 17, 207-214.
- Hyun, J. W., 2025, Study on machine learning algorithms for turbidity prediction in non-point source pollution monitoring networks based on water quality factors and rainfall event characteristics, Ph.D. Dissertation, Inha University, Incheon, Korea, <https://www.riss.kr/link?id=T17198029>
- Joo, H. G., Lim, J. M., 2023, Machine learning prediction model of water quality factors in sewage treatment facilities, J. Korean Inst. Ind. Eng., 49, 95-106.
- Jun, D. H., 2020, A Study of forecasting the effluent quality for small sewage treatment facilities using ARIMA model, Ph.D. Dissertation, Korea National University of Transportation, Chungcheongbuk-do, Korea, <https://www.riss.kr/link?id=T15481722>
- Kim, J. G., 2019, A Study on the analysis of operational factors of sewage treatment plant, M.S. Thesis, Ajou University, Gyeonggi-do, Korea, <https://www.riss.kr/link?id=T15095306>
- Kim, S. H., Park, J. H., Kim, B. H., 2021, Prediction of cyanobacteria harmful algal blooms in reservoir using machine learning and deep learning, J. Korea Water

- Resour. Assoc., 54, 1167-1181.
- Kim, S. J., 2020, Comparison of village sewage treatments by advanced treatment, M.S. Thesis, Semyung University, Chungcheongbuk-do, Korea, <https://www.riss.kr/link?id=T15542069>
- Kim, S. Y., 2020, A Study on improvement plans of small-scale sewage treatment facilities for watershed sewerage management, M.S. Thesis, Gyeongnam National University of Science and Technology, Gyeongnam, Korea, <https://www.riss.kr/link?id=T15465901>
- Kim, T. W., 2019, Case study for artificial intelligence application technology at sewage treatment plant, M.S. Thesis, Korea National University of Transportation, Chungcheongbuk-do, Korea, <https://www.riss.kr/link?id=T15048823>
- Lee, G. C., Park, Y. J., Kang, K. H., Jung, M. O., Ryu, D. H., Jung, S. S., Lee, W. T., 2021, Characteristics of organic matters in influents and effluents of sewage treatment plants in Gyeongsanbuk-do, J. Korean Soc. Environ. Eng., 43, 367-376.
- Lee, J. W., Kim, Y. H., Gil, K. G., 2022, A Study on operation control technology required for introduction of intelligent sewage treatment plant, J. Wetlands Res., 24, 38-43.
- Lee, S. M., Kang, Y. J., Song, J. W., Kim, J. H., Kim, H. S., Kim, S. J., 2024, Study on water quality prediction in water treatment plants using AI techniques, J. Korea Water Resour. Assoc., 57, 151-164.
- Lee, S. M., Kim, I. K., 2021, A Study on applying random forest and gradient boosting algorithm for Chl-a prediction of Daecheong lake, J. Korean Soc. Water Wastewater, 35, 507-516.
- Lee, S. M., Park, K. D., Kim, I. K., 2020, Comparison of machine learning algorithms for Chl-a prediction in the middle of Nakdong river, J. Korean Soc. Water Wastewater, 34, 277-288.
- Lim, Y. W., Eom, J. Y., Kwahk, K. Y., 2023, Development of a water quality prediction model for mineral springs in the metropolitan area using machine learning, J. Intell. Inform. Syst., 29, 307-325.
- Ministry of environment, 2025, Sewerage act, Sejong, Korea, <https://www.law.go.kr>
- Ministry of environment, 2025, Standard methods for water and wastewater examination, Sejong, Korea, <https://www.law.go.kr>
- Oh, W. K., Jang, H. N., Shin, S. G., 2022, Application of machine learning in water industry, J. Korean Soc. Water Wastewater, 36, 9-21.
- Park, J. S., 2021, Comparative characteristic of ensemble machine learning and deep learning models for turbidity prediction in a river, J. Korean Soc. Water Wastewater, 35, 83-87.
- Park, J. S., 2022, Development of a model to predict water quality using an automated machine learning algorithm, J. Korean Soc. Water Wastewater, 36, 329-337.
- Shin, J. G., 2020, A Study on the treatment characteristics of public sewage treatment facilities and management method : Focusing on the Boseong-gun, M.S. Thesis, Mokpo National University, Jeollanam-do, Korea, <https://www.riss.kr/link?id=T15703495>
- Son, D. J., Jeong, D. H., Park, K. H., 2021, Correlation analysis between TOC and organic matter indices in influent and effluent of public sewage treatment facilities, J. Conver. Inf. Technol., 11, 122-129.
-
- Researcher. Yoon-Cheol Yang
Department of Environment Research, Health and Environment Research Institute of Gwangju
ycyang83@korea.kr
 - Senior Researcher. Hee-Jeong Seo
Department of Environment Research, Health and Environment Research Institute of Gwangju
hjseo@korea.kr
 - Researcher. Sun-Jung Kim
Department of Environment Research, Health and Environment Research Institute of Gwangju
ksj0762@korea.kr
 - Researcher. Ha-Ram Kim
Department of Environment Research, Health and Environment Research Institute of Gwangju
khr0821@korea.kr
 - Researcher. Ji-Young Park
Department of Environment Research, Health and Environment Research Institute of Gwangju
jiyoung730@korea.kr
 - Researcher. Ju-Hyun Park
Department of Environment Research, Health and Environment Research Institute of Gwangju
jhpark7066@korea.kr
 - Researcher. Ki-Won Lee
Department of Environment Research, Health and Environment Research Institute of Gwangju
yabez2714@korea.kr
 - Professor. Seong-Jun Kim
Department of Environmental and Energy Engineering, Chonnam National University
seongjun@jnu.ac.kr
 - Director of Industrial Wastewater Division. Youn-Goo Lee
Department of Environment Research, Health and Environment Research Institute of Gwangju
inglyg@korea.kr