

**다중선형회귀경험식과 머신러닝모델의 2차원 횡 분산계수  
예측성능 평가**  
**Performance Evaluation of Multilinear Regression Empirical Formula  
and Machine Learning Model for Prediction of Two-dimensional  
Transverse Dispersion Coefficient**

이선미\*, 박인환\*\*  
 Sun Mi Lee, Inhwan Park

.....  
 요 지

분산계수는 하천에서 오염물질의 혼합능을 파악할 수 있는 대표적인 인자이다. 특히 하수처리장 방류수 혼합예측과 같이 횡 방향 혼합에 대한 예측이 중요한 경우, 하천의 지형적, 수리학적 특성을 고려한 2차원 횡 분산계수의 결정이 필요하다. 2차원 횡 분산계수의 결정을 위해 기존 연구에서는 추적자실험결과로부터 경험식을 만들어 횡 분산계수 산정에 사용해왔다. 회귀분석을 통한 경험식 산정을 위해서는 충분한 데이터가 필요하지만, 2차원 추적자 실험 건수가 충분치 않아 신뢰성 높은 경험식 산정이 어려운 상황이다. 따라서 본 연구에서는 SMOTE기법을 이용하여 횡분산계수 실험데이터를 증폭시켜 이로부터 횡 분산계수 경험식을 산정하고자 한다. 또한 다중선형회귀분석을 통해 도출된 경험식의 한계를 보완하기 위해 다양한 머신러닝 기법을 적용하고, 횡 분산계수 산정에 적합한 머신러닝 기법을 제안하고자 한다.

기존 추적자실험 데이터로부터 하폭 대 수심비, 유속 대 마찰유속비, 횡 분산계수 데이터 셋을 수집하였으며, SMOTE 알고리즘의 적용을 통해 회귀분석과 머신러닝 기법 적용에 필요한 데이터 그룹을 생성했다. 새롭게 생성된 데이터 셋을 포함하여 다중선형회귀분석을 통해 횡 분산계수 경험식을 결정하였으며, 새로 제안한 경험식과 기존 경험식에 대한 정확도를 비교했다. 또한 다중선형회귀분석을 통해 결정된 경험식은 횡 분산계수 예측범위에 한계를 보였기 때문에 머신러닝기법을 적용하여 다중선형회귀분석에 대한 예측성능을 평가했다. 이를 위해 머신러닝 기법으로서 서포트 벡터 머신 회귀(SVR), K근접이웃 회귀(KNN-R), 랜덤 포레스트 회귀(RFR)를 활용했다. 세 가지 머신러닝 기법을 통해 도출된 횡 분산계수와 경험식으로부터 결정된 횡 분산계수를 비교하여 예측 성능을 비교했다. 이를 통해 제한된 실험데이터 셋으로부터 2차원 횡 분산계수 산정을 위한 데이터 전처리 기법 및 횡 분산계수 산정에 적합한 머신러닝 절차와 최적 학습기법을 도출했다.

**핵심용어 : 분산계수, 머신러닝, 서포트 벡터 머신, K근접이웃, 랜덤포레스트**

**감사의 글**

본 결과물은 환경부의 재원으로 한국환경산업기술원의 미세플라스틱 측정 및 위해성평가 기술개발사업의 지원을 받아 연구되었습니다.(과제번호: 2021003110003).

\* 정회원 · 서울과학기술대학교 건설시스템공학과 석사과정 · E-mail : [sunmi619@seoultech.ac.kr](mailto:sunmi619@seoultech.ac.kr)

\*\* 정회원 · 서울과학기술대학교 건설시스템공학과 조교수 · E-mail : [ihpark@seoultech.ac.kr](mailto:ihpark@seoultech.ac.kr)