

머신러닝을 활용한 팔당호 유해남조 세포수 예측 Prediction of harmful algal cell density in Lake Paldang using machine learning

변서현*, 이한규**, 김진휘***, 신재기****, 박용은*****
Seohyun Byeon, Hankyu Lee, Jin Hwi Kim, Jae-Ki Shin, Yongeun Park

요 지

유해 남조 대발생(Harmful Algal blooms, HABs)이 담수호에 발생하면 마이크로시스틴과 같은 독성물질과 맛·냄새 물질을 생성하여 상수원이용과 친수활동을 방해한다. 그래서 유해 남조 대발생 전 유해남조 세포수를 예측하여 선제적 대응하는 것은 중요하다. 따라서 본 연구는 머신러닝 기반 Random Forest(RF)를 활용하여 팔당호 앞의 유해남조 세포수를 예측하는 모델을 개발하고 성능을 평가하고자 한다.

모델 구축을 위해 2012년 4월부터 2021년 12월까지의 팔당호(삼봉리, 경안천) 및 남북한강(의암댐~이포보)권역의 조류, 수질, 수리/수문, 기상 자료를 수집하여 입력 및 출력 자료로 이용하였다. 수집된 데이터에는 다양한 입력변수들이 있어 남조 세포수 예측 성능 비교를 위한 전체 26개 변수 적용과 통계학적으로 상관관계가 높은 12개 변수 적용을 통해 모델을 구축하였다. 입력, 출력 자료로 이용한 유해남조 세포수는 로그변환된 값으로 사용하였으며 일반적인 조류 시료 채취 기간이 7일이므로 7일 후를 예측하기 위한 모델을 구축하였다. 구축한 모델의 성능은 실측데이터와 예측데이터의 R^2 로 산출하여 평가하였다.

전체 26개 입력변수로 모델 구축 후 학습 및 검증 수행 결과 R^2 의 학습 0.803, 검증 0.729로 나타났다. 유해남조 세포수와 유의미한 상관관계를 보이는 12개 입력변수로 모델 구축 후 학습 및 검증 수행 R^2 은 학습 0.784, 검증 0.731로 나타났다. 두 모델의 성능을 살펴본 결과 입력변수 개수의 변화에 따른 성능차이는 크지 않은 것으로 나타났으며, 남조세포수 예측을 위한 모델로서 활용 가능성을 알 수 있었다. 향후 연구에서는 Random Forest 외 다른 기계학습 모델들과 딥러닝 모델을 통해 남조세포수 예측 성능이 높은 모델을 구축해볼 필요성이 있다.

핵심용어 : 유해남조 세포수 예측, Random Forest, 상관성

감사의 글

본 연구는 한강수계관리위원회 환경기초조사사업의 지원을 받아 수행되었습니다.

* 정희원 · 건국대학교 대학원 사회환경플랜트공학과 석사과정 · E-mail : shbyeon1@gmail.com

** 정희원 · 건국대학교 대학원 사회환경플랜트공학과 박사수료 · E-mail : haeckel@konkuk.ac.kr

*** 정희원 · 건국대학교 공과대학 사회환경공학부 박사 후 연구원 · E-mail : jinhwi25@gmail.com

**** 일반회원 · 한국수자원공사 부산권지사 수석연구원 · E-mail : jaekishin@kwater.or.kr

***** 정희원 · 건국대학교 공과대학 사회환경공학부 부교수 · E-mail : yepark@konkuk.ac.kr