creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사 학위논문

기계 학습을 이용한 퇴적물 주입 활성탄의 소수성 유기오염물질 겉보기 흡착능 예측

Machine learning-based prediction of apparent adsorption capacity of sediment-amended activated carbon for hydrophobic organic contaminants

2023 년 08 월

서울대학교 대학원

건설환경공학부 환경공학 전공

이현민

기계 학습을 이용한 퇴적물 주입 활성탄의 소수성 유기오염물질 겉보기 흡착능 예측

지도 교수 최 용 주

이 논문을 공학석사 학위논문으로 제출함
2023 년 08 월

서울대학교 대학원
건설환경공학부 환경공학 전공
이 현 민

이현민의 공학석사 학위논문을 인준함
2023 년 08 월

위 원 장 _____ 남 경 필 _____ (인)

부위원장 _____ 최 용 주 _____ (인)

위 원 _____ 최 정 권 _____ (인)

초 록

소수성 유기오염물질로 오염된 토양은 인간과 생태계에 심각한 위험을 가한다. 활성탄을 이용한 원위치 안정화 기술은 높은 정화 효율과 적은 생태 교란 때문에 각광받고 있는 정화기술의 일종이다. 그러나 퇴적물-물-활성탄 시스템의 복잡성에 의해 활성탄이 퇴적물로 주입되었을 때의 흡착능을 예측하는 것은 여전히 과제로 남아있으며, 아직까지 흡착능 예측 모델이 존재하지 않는 상황이다. 본 연구에서는 순수 분배 계수(물-활성탄 시스템) 뿐만 아니라 겉보기 분배 계수(퇴적물-물-활성탄 시스템)를 예측하는 기계학습 기반 모델을 개발하였다. 기존 연구에서 얻은 186개의 데이터를 이용하여 모델을 학습시켰으며, 변수 선정과 하이퍼파라미터 선정을 통해 모델을 최적화하였다. 순수 분배 계수에 영향을 주는 요소는 흡착제의 비표면적, 흡착질의 분자량과 옥탄올-물 분비 계수이며, 겉보기 분배 계수에 영향을 주는 요소는 퇴적물 특성인 총유기탄소량과 black carbon양, 흡착제의 특성인 활성탄 입자 크기, 흡착질의 특성인 분자량, 그리고 실험 조건인 실험 기간으로 나타났다. 다른 구조를 가진 세 가지 모델을 비교하여 순수 분배 계수와 겉보기 분배 계수를 분리하여 학습시키는 모델 구조가 가장 좋은 성능을 보이는 것을 확인하였으며, 이를 통해 순수 분배 계수와 겉보기 분배 계수에 영향을 미치는 요소가 다르다는 것을 확인하였다. 최종 예측 모델은 모두 support vector regressor를 기계학습 알고리즘으로 활용하였으며, 순수 분배 계수에 대해서는 결정 계수 0.91과 평균 제곱근 오차 0.38, 겉보기 분배 계수에 대해서는 결정 계수 0.86과 평균 제곱근 오차 0.40을 달성하였다.

본 연구는 활성탄의 흡착능 예측에 있어서 기계 학습의 적용 가능성을 제시하고, 순수 분배 계수 예측 모델과 겉보기 분배 계수 예측 모델의 차이에 대해 집중적으로 논의하였다. 이를 통해 퇴적물 주입 활성탄에서 발생하는 흡착 과정에 대해 더욱 폭넓게 이해하여 원위치 안정화 기술을 효율적으로 적용 및 관리하는 데에 기여하였다.

주요어 : 퇴적물 오염, 활성탄, 흡착능, 기계 학습, 소수성 유기오염물질, 예측 모델

학 번 : 2021-22740

목 차

제 1 장 서 론 1

제 1 절 연구의 배경 1

제 2 절 연구의 목적 및 내용 4

제 2 장 실험 재료 및 방법 5

제 1 절 데이터 수집 5

제 2 절 모델 개발 및 성능 평가 6

제 3 절 변수 선정 10

제 4 절 변수 분석 11

제 3 장 실험 결과 13

제 1 절 모델 예측 결과 13

제 2 절 변수 중요도 분석 21

1. 흡착질 특성의 영향 26

2. 퇴적물 특성의 영향 27

3. 흡착제 특성의 영향 29

4. 실험 조건의 영향 31

제 4 장 결론 및 제언 33

제 1 절 한계 및 향후 발전 방향 33

제 2 절 결과 및 의의 35

참고문헌 37

부록 46

Abstract 58

표 목차

[표 1] 변수 선정 전 모델 예측 결과 14

[표 2] 변수 선정 후 모델 예측 결과 17

[표 3] 여섯 개의 입력 변수들 간의 피어슨 상관계수 22

[표 4] 최종 모델의 변수 조합 변화에 따른 예측 결과 23

그림 목차

[그림 1] 예측 모델 구조 도식 8

[그림 2] 최종 모델의 순수 분배 계수(Clean)와 겉보기 분배 계수 (Apparent)에 대한 실제 값 대비 예측 값 그래프 20

[그림 3] 최종 모델의 부분 의존도 그래프 24

[그림 4] 최종 모델의 순열 특성 중요도 25

제 1 장 서 론

제 1 절 연구의 배경

폴리염화비페닐(Polychlorinated biphenyls, 이하 PCBs)이나 다환방향족탄화수소 (Polycyclic aromatic hydrocarbons, 이하 PAHs)와 같은 소수성 유기오염 물질들은 유전독성과 발암성을 가지고 있기 때문에 소수성 유기오염 물질로 오염된 토양은 인간과 주변 생태계에 심각한 위험을 초래한다[①]. 다양한 정화기술들 중, 활성탄을 이용한 원위치 안정화 기술은 정화 효율이 높고 준설이나 피복 기술에 비해 기술 적용 시 생태 교란이 비교적 적다는 점에서 각광받는 기술이다[②]. 활성탄은 넓은 비표면적과 소수성 표면을 가지고 있기 때문에 소수성 유기오염물질에 대한 흡착능이 매우 높다[③]. 활성탄을 오염 퇴적물에 주입하게 되면 활성탄은 퇴적물 내 공극수에 존재하는 소수성 유기오염물질을 흡착시키는 흡착제로 작용한다. 이 과정을 통해 퇴적물 및 공극수 내 오염물질의 농도가 감소하고, 따라서 오염물질의 생물학적 이용 가능성이 감소하여 주변 생태계에 미치는 위험성이 줄어들게 된다.

원위치 안정화 기술의 정화 효율은 사용되는 흡착제의 흡착능에 큰 영향을 받는다. 많은 연구들이 물-활성탄 시스템

---

① Drwal et al., 2019; Dai et al., 2016
② Lofrano et al., 2017; Majone et al., 2015.
③ Cho et al., 2007.

내에서의 활성탄의 흡착 특성에 대한 실험을 진행하였으며, 흡착능에 영향을 주는 다양한 인자들을 파악하였다. 비표면적이나 표면 작용기와 같은 흡착제의 특성은 흡착질인 오염물질이 흡착제에 흡착될 수 있는 공간과 오염물질의 표면에 대한 친화도를 결정하기 때문에 흡착능 결정에 있어 중요한 지표로 활용된다[④]. 또한, 오염물질의 분자량과 소수성 같은 흡착질의 특성은 흡착제 내 공극으로의 접근성이나 표면에 대한 친화도를 결정하기 때문에 중요한 지표로 작용된다[⑤].

그러나 물- 활성탄 시스템에서 계산한 활성탄의 흡착능과 퇴적물-물-활성탄 시스템에서 계산한 흡착능은 차이를 가지기 때문에, 물-활성탄 시스템에서 결정된 활성탄의 흡착능은 안정화 기술에 활성탄이 적용되어 퇴적물 내로 주입되었을 때의 활성탄의 흡착능을 정확하게 표현하기 힘들다. 이러한 차이는 흙에 존재하는 용존유기물질과 입자성 물질들이 소수성 유기오염물질의 활성탄으로의 흡착을 방해하는 현상 때문에 발생한다[⑥]. 예를 들어, Cornelissen 외 (2006)는 퇴적물 내에서 활성탄의 흡착능이 물에서의 흡착능보다 6배 이상 감소했다고 보고하였다. 또한, Werner 외 (2006)는 퇴적물-물-활성탄 시스템에서 흡착 실험을 진행한 결과, 기존 문헌 값에 비해 흡착능이 16배가량 감소한 것으로 확인하였다. 이러한 차이를 고려하였을 때, 퇴적물-물-활성탄 시스템에서 결정된 겉보기 흡착능 (apparent adsorption capacity)을 사용하면 원위치 안정화 기술의 효율을 더욱

---

④ Zhang et al., 2010.; Gomez-Eyles et al., 2013
⑤ Jamil et al., 2019; Wang et al., 2010
⑥ Koelmans et al., 2009.; Rakowska et al., 201

직접적으로 표현할 수 있다[⑦]. 그러나 퇴적물 내 용존 유기물질과 입자성 물질들이 오염물질 및 활성탄과 복잡한 상호작용을 하기 때문에 퇴적물이 활성탄의 흡착능에 미치는 영향을 정확히 파악하는 것에 한계가 존재하며, 따라서 퇴적물 주입 활성탄의 겉보기 흡착능을 예측하는 것은 여전히 과제로 남아있다[⑧]. 이러한 이유로 인해 원위치 안정화 기술의 오염 저감 효율을 예측하는 것에 어려움을 겪고 있으며, 나아가 소수성 유기오염물질로 오염된 퇴적물이 야기하는 위험성을 효과적으로 관리하는 것에 제한이 있다.

최근 들어 기계학습은 복잡한 대규모 데이터로부터 학습하여 예측 모델을 개발하기 위한 도구로써 활발히 사용되고 있다[⑨]. 기계학습은 복잡한 데이터 내에 숨겨진 관계성과 패턴을 파악하기 위한 유용한 수단으로 사용될 수 있으며, 흡착능 관련 연구에서 역시 예측 모델 개발이나 영향 인자 분석에 기계 학습을 적용한 사례들이 여럿 보고되었다[⑩]. 기계 학습은 기존의 선형 회귀모델 모다 다른 변수들 간의 관계를 효율적으로 학습할 수 있다는 점에서, 원위치 안정화기술에 적용된 활성탄의 겉보기 흡착능을 예측하기 위한 효과적인 수단으로 사용될 수 있다.

---

⑦ Choi et al. 2013.; Hale et al.; 2010.
⑧ Pignatello et al., 2006; Cornelissen and Gustafsson, 2006; Corwin and Summers, 2011
⑨ Zhu et al., 2019
⑩ Zhu et al., 2022; Zhang et al., 2020

제 2 절 연구의 목적 및 내용

이 연구는 기계학습을 이용하여 활성탄의 소수성 유기오염물질에 대한 겉보기 흡착능을 예측하는 모델을 개발하는 것에 목표를 두고 있다. 이를 위해 가장 범용적으로 사용되는 세 기계 학습 알고리즘과 세 가지의 다른 구조를 활용하여 예측 모델을 개발하였으며, 가장 좋은 성능을 보이는 모델을 최종 모델로 선정하여 예측을 진행하였다. 나아가 모델 학습에 사용된 인자들이 활성탄의 겉보기 흡착능에 어떤 영향을 끼치는지를 다양한 지표를 이용하여 평가하였다. 본 연구는 물-활성탄 시스템에서의 활성탄 흡착능이 아닌, 퇴적물-물-활성탄 시스템에서 활성탄의 겉보기 흡착능을 예측한다는 점에서 기존 연구들과 차이점을 가지며, 기계 학습을 이용하여 활성탄의 겉보기 흡착능에 대한 예측 모델을 개발한 첫 연구이다.

제 2 장 실험 재료 및 방법

제 1 절 데이터 수집

연구에서 예측 모델의 학습 및 평가에 사용된 데이터는 기존 문헌들로부터 수집되었으며, 3.28 ~ 8.09 사이의 로그 옥탄올-물 분배계수 ($\log K_{ow}$) 값을 가지는 PAHs와 PCBs를 연구 대상 흡착질로 한정하였다. 데이터의 총 개수는 186개이며, 사용된 데이터는 부록의 부표 2에 첨부하였다.

본 연구에서는 활성탄의 흡착능에 영향을 준다고 알려진 여덟 가지 입력변수를 사용하였으며, 변수 전체 혹은 일부를 이용하여 기계학습 모델들을 훈련시켰다. 이 변수들은 네 종류로 분류될 수 있다. (1) 흡착질 특성: 분자량 (MW, [$\frac{g}{mol}$]), 로그 옥탄올-물 분배계수 ($\log K_{ow}$, [-]), (2) 흡착제 특성: 입자 크기 (AC size, [cm]), 비표면적 (AC SSA, [$\frac{m^2}{kg}$]), (3) 퇴적물 특성: 퇴적물 내 총유기탄소의 무게 비 (TOC, [%]), Black carbon의 무게 비 (BC, [%]), (4) 실험 조건: 실험 기간 (duration, [days]), 퇴적물 대비 활성탄의 무게 기준 투입비 (AC dose, [%]). 위 여덟 가지 변수 외에도 활성탄의 흡착능에 영향을 준다고 알려져 있는 변수들이 존재하지만, 본 연구에서 개발한 예측모델의 적용 가능성과 데이터 활용 가능성을 고려하여 입력 변수를 여덟 가지로 한정하였다.

활성탄의 흡착능은 활성탄-물 분배계수($\log K_{AC}$)를 이용하여 나타내었다. 활성탄-물 분배계수는 물-활성탄 실험 혹은 퇴적물-

물-활성탄 실험의 마지막에 측정된 오염물질의 활성탄 내 농도를 수중 농도로 나눈 값이다.

$$K_{AC} = C_{AC}/C_w \tag{1}$$

물-활성탄 실험을 통해 결정된 활성탄-물 분배계수는 순수 분배 계수(Clean distribution coefficient, $\log K_{AC_clean}$)로 표기하며, 퇴적물-물-활성탄 실험을 통해 결정된 분배계수는 겉보기 분배 계수 (Apparent distribution coefficient, $\log K_{AC,apparent}$)로 표기한다. 연구에 사용된 분배계수들은 28일부터 365일 간 진행한 실험의 종료 시점에서 측정한 실험 값이며, 이는 흡착 평형을 이루기에는 부족한 시간이기 때문에 사용된 분배 계수들이 평형 상태에서의 분배계수가 아니라는 점에 유의해야 한다[11].

제 2 절 모델 개발 및 성능 평가

순수 분배 계수와 겉보기 분배 계수를 예측하기 위해 세 가지의 모델 구조를 구상하였으며 그림 1에 각 구조에 대한 모식도를 첨부하였다. 앞서 언급한 여덟 가지 변수 전체 혹은 일부를 사용하여 모델을 학습시켰으며, 변수 선정 과정을 통해 최고의 성능을 보이는 변수 조합을 적용한 모델을 최종 모델로 선정하였다.

Model 1은 일곱 가지의 변수 (MW, $\log K_{ow}$, AC size, AC SSA,

[11] Hale et al., 2010.; Ahn et al., 2008.

TOC, BC, 그리고 duration)를 사용하여 순수 분배 계수와 겉보기 분배 계수를 예측하는 단일 모델이다. 활성탄의 투입량 (AC dose)의 경우, 물-활성탄 실험에서는 해당 값을 계산할 수 없기 때문에 입력 변수에서 제외하였다. 또한, 물-활성탄 실험에서는 TOC와 BC의 값은 0으로 설정되었다. Model 2는 순수 분배 계수를 예측하는 모델과 겉보기 분배 계수를 예측하는 모델을 분리하여 각각 학습시킨다. 전자는 Model 2_clean이며, 퇴적물 특성과 활성탄의 투입량을 제외한 다섯 가지 변수(MW, $log\,K_{ow}$, AC size, AC SSA, 그리고 duration)를 입력 변수로 사용하였다. 후자는 Model 2_apparent이며, 여덟 가지 변수를 모두 사용하였다. Model 3은 두 가지 과정으로 진행된다. 첫 번째 과정에서는 세 변수 (MW, $log\,K_{ow}$, 그리고 AC SSA) 및 알고리즘의 파라미터가 최적화된 예측 모델을 사용하여 순수 분배 계수를 예측한다. 두 번째 과정에서는 여덟 가지 변수를 첫 번째 과정에서 예측된 순수 분배 계수를 입력 변수로써 포함하여 총 아홉 가지의 입력 변수를 이용하여 겉보기 분배 계수를 예측한다.

모델 구축에 사용된 기계 학습 알고리즘은 multiple linear regression (MLR), support vector regression (SVR), and random forest regression (RFR) 총 세 가지이다. 이전 연구들에서 SVR, RFR 알고리즘이 탄소계 흡착제의 흡착능을 예측하는 데에 성공적인 결과를 보여주었기 때문에 본 연구에서도 학습 알고리즘으로 사용하였다[⑫]. MLR의 경우 가장 간단한 방식의 전통적인 선형 회귀 모델로서, SVR, RFR과 예측 성능을 비교하기

[⑫] Palansooriya et al., 2022; Lateef et al., 2022.

위해 적용되었다.

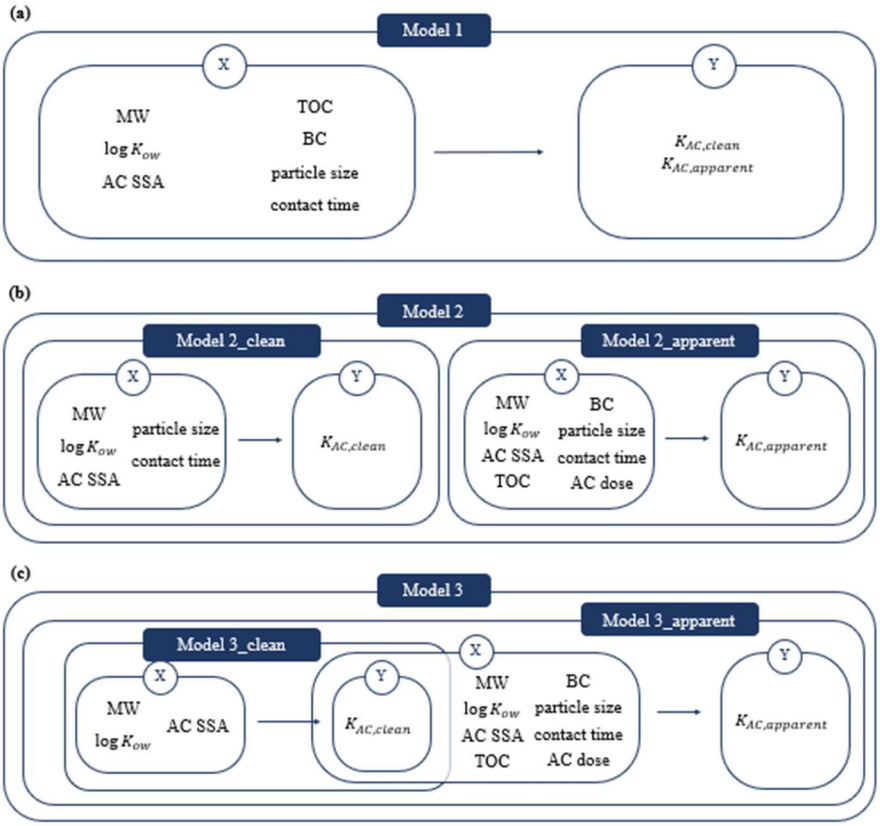

그림 1 예측 모델 구조 도식 (a) Model 1, (b) Model 2, (c) Model 3

입력 변수들은 변수들의 크기를 조정하기 위해 정규화 하여 모델에 적용하였다.

$$x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

x_i^*는 x_i의 정규화 된 값을, x_{min}과 x_{max}는 변수 x 중 최소값과 최대값을 일컫는다.

데이터는 80:20의 비율로 훈련 데이터 세트와 평가 데이터

세트로 나누어졌다. 이후 훈련 데이터 세트에 대해 five-fold 교차검증을 진행하여 최적의 하이퍼 파라미터 (Hyperparameter)를 선정하였다. Five-fold 교차검증에서는 훈련 데이터 세트를 같은 크기의 다섯 데이터 세트로 다시 나눈 뒤 네 데이터 세트는 모델 훈련에 사용하고 나머지 한 데이터 세트는 성능을 검증하기 위한 데이터 세트로 사용한다. 이 과정을 다섯 번 반복하여 다섯 개의 데이터 세트가 모두 검증 데이터로 활용되며, 이를 통해 모델의 성능을 최적화하는 하이퍼 파라미터를 구할 수 있다. Five-fold 검정을 통해 하이퍼 파라미터가 선정되 이를 모델에 적용한 후 평가 데이터 세트를 통해 모델의 예측 정확도를 평가한다.

모델의 성능을 평가하기 위한 지표로는 결정 계수(R^2 , coefficient of determination)와 평균 제곱근 편차 (RMSE, Root mean squared error)가 사용되었다. 결정계수는 실제 결과값 중 입력 변수를 이용하여 예측한 값으로 설명할 수 있는 비율을 나타내는 계수이다. R^2가 1에 가까울 수록 예측을 이용해 설명할 수 있는 비율이 높다는 뜻이므로 모델의 예측 정확도가 높다고 할 수 있다. 반면 평균 제곱근 편차는 입력 변수를 통해 예측한 값과 실제 결과값 사이의 차이를 평균적으로 나타낸 수치이므로, 값이 작을 수록 예측 정확도가 높다고 할 수 있다. 두 지표를 계산한 식은 아래와 같다.

$$R^2 = 1 - \frac{\Sigma_{i=1}^{N}(\hat{y_i} - y_i)^2}{\Sigma_{i=1}^{N}(y_i - \bar{y})^2} \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\hat{y_i} - \bar{y})^2} \quad (3)$$

$\hat{y_i}$는 실제 결과값 y_i에 대한 모델의 예측 값을, $\bar{y}$는 y_i값들의 평균을 일컫는다.

모델의 성능을 정확하게 보고하기 위해 각기 다른 평가 데이터 세트에 대해 평가를 최소 100번 이상 반복한 후 결정 계수의 평균값과 표준편차 값, 그리고 제곱근 편차의 평균값을 보고하였다. 이 지표들을 통해 모델의 예측 성능을 표현하여 모델의 신뢰성을 전반적으로 평가할 수 있다.

제 3 절 변수 선정

예측 모델의 정확성과 효율성을 향상시키기 위하여 전체 입력 변수들 중 결과 값 예측에 불필요한 변수를 제거하는 변수 선정 과정을 진행하였다. 입력 변수의 총 개수가 여덟 개이므로 만들 수 있는 조합의 가짓수가 많지 않기 때문에, 세 가지 이상의 변수를 포함하는 가능한 모든 변수 조합을 모델 학습에 사용하여 가장 높은 성능을 보이는 변수 조합을 최종 모델에 적용하였다. 최적의 변수 조합 선정 이후에는 앞서 언급한 five-fold 교차 검증을 이용하여 하이퍼 파라미터를 적용하여 최종 모델을 결정하였다.

제 4 절 변수 분석

변수 선정과 하이퍼 파라미터를 선정하여 최종 모델을 결정한 후, 각 변수들이 결과값인 분배 계수에 미치는 영향을 분석하기 위해 세 가지 지표를 이용하였다.

첫 번째 지표는 피어슨 상관계수 (Pearson correlation coefficient)이다. 피어슨 상관계수는 두 변수의 선형적 관계성을 정량화 하는 지표이다. 상관계수가 1 (혹은 −1)에 가까울 경우, 두 변수가 강한 양의 (음의) 상관관계를 가진다는 것을 나타낸다.

$$\mathrm{r_{xy}} = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})\Sigma_{i=1}^{n}(y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

$\bar{x}$과 $\bar{y}$는 각각 x, y 값의 평균값을 일컫는다

피어슨 상관계수는 예측 모델의 성능에 상관없이 두 변수의 값을 직접 계산에 적용하여 변수들 간의 관계를 계산한다. 반면 이후 다룰 지표인 부분 의존도 (Partial dependence)와 순열 특성 중요도 (Permutation feature importance)는 모델 기반 지표이다. 이는 해당 지표들이 훈련된 모델을 기반으로 계산되기 때문에 어떤 예측 모델을 사용했는지에 따라 값이 달라진다는 것을 의미한다. 부분 의존도는 아래와 같은 식을 통해 계산한다.

$$\widehat{f_{x_v}}(x_v) = \int \hat{f}\,(x_v, x_C)\,d\boldsymbol{P}(x_C) \tag{4}$$

$\widehat{f_{x_v}}$는 부분 의존도 함수를, $\hat{f}$는 예측 모델을 나타내며, x_v는 관심 변수를, x_C는 관심 변수를 제외한 모든 변수를 일컫는다[13]. 부분 의존도는 보통 그래프로 표현이 되는데, 그래프는 관심 변수를 제외한 모든 변수를 고정한 후, 관심 변수를 전체 구간에 대해 변화시키면서 나타나는 결과값의 변화를 그래프 상에 나타내는 방식으로 작성된다. 이를 통해 특정 변수의 변화가 모델의 결과값에 어떤 영향을 주는지 나타낼 수 있다.

순열 특성 중요도는 기존 변수 구성에서 보이는 모델의 성능에 대비하여 특정 변수를 무작위로 섞었을 때 모델의 성능이 얼마나 감소하는지를 측정함으로써 계산할 수 있다. 이를 통해 모델이 특정 변수가 제공하는 정보에 얼마나 많이 의존하고 있는지를 평가할 수 있다[14].

---

[13] Gusenbauer et al., 2020
[14] Breiman, 2001

제 3 장 실험 결과

제 1 절 모델 예측 결과

우선, SVR, RFR, 그리고 MLR 총 세 가지의 알고리즘을 모든 입력변수와 각각의 하이퍼 파라미터를 이용하여 모델을 훈련시켰다. 예측 결과는 표 1과 같다.

표 1 변수 선정 전 모델 예측 결과

		Regression model	R^2 average (standard deviation)	RMSE average
Model 1	$\log K_{AC,clean}$	SVR	0.88 ± 0.08	0.45
		RFR	0.90 ± 0.06	0.42
		MLP	Invalid	1.73
	$\log K_{AC,apparent}$	SVR	0.82 ± 0.11	0.46
		RFR	0.82 ± 0.14	0.45
		MLP	Invalid	1.69
Model 2	$\log K_{AC,clean}$	SVR	0.91 ± 0.05	0.40
		RFR	0.89 ± 0.06	0.45
		MLP	0.60 ± 0.22	0.83
	$\log K_{AC,apparent}$	SVR	0.84 ± 0.10	0.44
		RFR	0.86 ± 0.11	0.43
		MLP	Invalid	1.34
Model 3	$\log K_{AC,apparent}$	SVR	0.83 ± 0.10	0.45
		RFR	0.82 ± 0.14	0.46
		MLP	Invalid	1.20

MLR 알고리즘을 이용해 학습시킨 다섯 개의 모델들 중 네 모델이 R^2 평균 값에서 Invalid, 즉 음수 값을 나타냈다. R^2는 식 (2)로 계산되는데, 이를 계산할 때 식 (6)과 같이 전체 제곱합이 잔차 제곱합과 회귀 제곱합의 합으로 표현이 될 수 있다는 전제를

내포하고 있다.

$$Total\ sum\ of\ squares = Explained\ sum\ of\ squares + Residual\ sum\ of\ squares \tag{6}$$

$$\Sigma_{i=1}^{N}(\bar{y} - y_i)^2 = \Sigma_{i=1}^{N}(\hat{y_i} - \bar{y})^2 + \Sigma_{i=1}^{N}(\hat{y_i} - y_i)^2$$

그러나 예측 모델이 항상 결과값의 평균($\bar{y}$)을 예측 값으로 제시하는 상수함수보다 부정확할 경우, 잔차 제곱합이 전체 제곱합을 능가하면서 R^2의 기본 전제인 식 (6)을 만족하지 못하게 된다. 이로 인해 R^2값이 음수가 될 수 있으며, 이는 모델이 변수 간의 상관 관계를 파악하여 예측 모델을 만드는 데에 실패하였다는 것을 의미한다. 이 결과를 통해 활성탄의 흡착능을 예측하는 데에 있어 전통적인 선형회귀식을 적용하기 힘들며, 따라서 다른 기계 학습 알고리즘처럼 다양한 형태를 가진 데이터를 유연하게 학습할 수 있는 알고리즘을 적용해야 한다는 것을 알 수 있다.

SVR, RFR 이 적용된 모델은 항상 MLR이 적용된 모델보다 우수한 성능을 보인다. 이는 SVR과 RFR이 MLR보다 데이터 변수가 가진 비선형성이나 변수 간의 복잡한 관계를 파악하는데 유리하며, 결과적으로 두 기계학습 모델이 분배 계수 예측에 더욱 효과적임을 나타낸다.

Model 1,2,3 중 Model 2가 순수 분배 계수와 겉보기 분배 계수 모두에 대해 가장 좋은 성능을 보였다. 두 분배 계수에 대해 같은 입력 변수를 사용한 Model 1과는 달리, Model 2_clean과 Model 2_apparent는 순수 분배 계수와 겉보기 분배 계수에 대해

각기 다른 입력 변수를 이용하여 분리되어 훈련되었기 때문에 입력 변수와 분배 계수 결과값 간의 서로 다른 관계성을 모델이 학습하였을 수 있다. Model 3_apparent는 Model 2_apparent와 달리 첫 번째 과정에서 예측된 순수 분배 계수 값을 입력 변수로 활용하였으나 이것이 예측 정확도 향상에 기여하지 못했다. 이는 별개의 모델로 예측된 순수 분배 계수 값이 가지는 불확실성 때문인 것일 수 있다.

다음으로, SVR, RFR을 알고리즘으로써 적용하고 변수 선정을 진행하여 변수 조합 중 최고의 예측성능을 보이는 모델을 선별하였다. Model 1,2,3에 대한 변수 선정 결과는 표 2에 첨부하였다. 부록의 부표 2에 표 1, 2에 포함된 모든 모델의 하이퍼 파라미터와 선정 변수를 첨부하였다.

표 2 변수 선정 후 모델 예측 결과

		Regression model	R^2 average	RMSE	Selected features	Excluded features
Model 1	$\log K_{AC,clean}$	SVR	0.90	0.40	$\log K_{ow}$, AC size, AC SSA, TOC, and Duration	MW and BC
		RFR	0.90	0.40	MW, $\log K_{ow}$, AC SSA, TOC, and BC	AC size and Duration
	$\log K_{AC,apparent}$	SVR	0.82	0.46	MW, $\log K_{ow}$, AC SSA, AC size, TOC, BC, and Duration	-
		RFR	0.83	0.43	MW, $\log K_{ow}$, AC SSA, TOC, and BC	AC size and Duration
Model 2	$\log K_{AC,clean}$	SVR	0.91	0.38	MW, $\log K_{ow}$, and AC SSA	AC size and Duration

		RFR	0.89	0.43	$\log K_{ow}$, AC size, AC SSA, and Duration	MW
	$\log K_{AC,apparent}$	SVR	0.86	0.40	MW, AC size, TOC, BC, and Duration	$\log K_{ow}$, AC SSA, and AC dose
		RFR	0.86	0.43	MW, $\log K_{ow}$, AC SSA, AC size, TOC, BC, Duration, and AC dose	-
Model 3	$\log K_{AC,apparent}$	SVR	0.86	0.42	Predicted $\log K_{AC,clean}$, MW, AC size, TOC, BC, and Duration	$\log K_{ow}$, AC SSA, and AC dose
		RFR	0.83	0.44	Predicted $\log K_{AC,clean}$, MW, AC size, AC SSA, TOC, Duration, and AC dose	$\log K_{ow}$ and BC

모든 모델 구조 및 기계학습 알고리즘에 대하여, 순수 분배 계수를 예측하는 모델이 겉보기 분배 계수를 예측하는 모델보다 더 높은 정확도를 보였다. 이는 순수 분배 계수를 얻기 위해 진행된 물-활성탄 실험에 수반된 흡착 과정이 겉보기 분배 계수를 얻기 위해 진행된 퇴적물-물-활성탄 실험에 수반된 흡착 과정에 비해 덜 복잡하고 예측하기 쉽다는 것을 나타낸다. 이러한 정확도의 차이는 퇴적물의 존재로 인해 입자 간의 상호작용이 더 복잡해졌기 때문인 것으로 추측되며, 이로 인해 겉보기 분배 계수를 예측하는 데에 필요한 인자의 수가 증가하고 모델의 예측 성능이 감소하였다.

변수 선정의 결과로 1-3개의 입력 변수들이 제외되었다 (표 2). 이는 특정 입력 변수들은 예측 모델의 성능에 부정적인 영향을 미치거나 혹은 큰 영향을 주지 않는 다는 것을 나타낸다. 각 입력 변수들의 영향에 대한 논의는 다음 부분에서 다루었다. 변수 선정 이후에는 10 개의 모델 중 8 개의 모델에서 평균 결정 계수 값이 증가한 것을 관찰할 수 있었으며, 이를 통해 모델의 예측 성능이 향상되었다는 것을 확인할 수 있다. 두 모델(Model 1_apparent with SVR algorithm and Model 2_apparent with RFR algorithm)에 대해서는 변수 선정 이후 평균 결정 계수 값이 감소하였기 때문에 모든 변수를 최종 변수 조합으로 선정하였다.

흥미롭게도 표 2에서 순수 분배 계수와 겉보기 분배 계수를 예측하는데 사용된 최적의 변수가 서로 다르다는 것을 알 수 있다. 뿐만 아니라 두 분배 계수를 동시에 학습한 Model 1의 경우에도, 두 분배 계수를 가장 잘 예측하는 변수들의 조합은 서로 다르다는

것을 알 수 있다. 이는 두 분배 계수를 다른 입력 변수를 사용하여 각각 학습시키는 Model 2가 더 좋은 성능을 보이는 것과 맥락을 같이한다. 이러한 결과를 통해 순수 분배 계수와 겉보기 분배 계수를 결정하는 데에 관여하는 인자들이 다르다는 것을 알 수 있다.

Model 2가 순수 분배 계수와 겉보기 분배 계수 모두에 대해 가장 좋은 예측 결과를 보여주었기 때문에 Model 2를 최종 모델로 선정하였다. 그림 2에서는 최종 모델을 사용하여 $\log K_{AC,apparent}$ 와 $\log K_{AC,clean}$ 에 대한 실제 값 대비 예측 값의 그래프를 제시하였다.

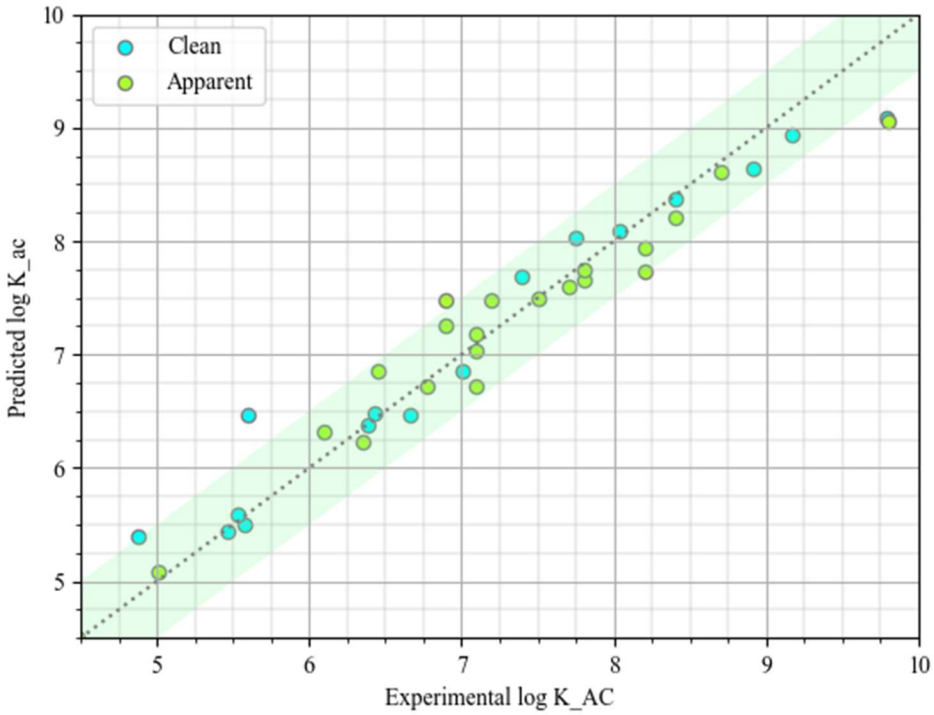

그림 2 최종 모델의 순수 분배 계수(Clean)와 겉보기 분배 계수 (Apparent)에 대한 실제 값 대비 예측 값 그래프. 점선은 실제 값과 예측 값이 같은 선임.

제 2 절 변수 중요도 분석

최종 모델인 Model 2_clean과 Model 2_apparent에 대해 변수들을 (1) 흡착질 특성, (2) 흡착제 특성, (3) 퇴적물 특성, (4) 실험 조건, 총 네 가지 기준으로 나누어 변수의 중요도 및 결과값에 대한 영향을 분석하였다. 최종 모델 결과 (표 2), 입력 변수 간 피어슨 상관 계수 (표 3), 최종 모델 변수 조합 변경 결과 (표 4), 부분 의존도 그래프 (그림 3), 순열 특성 중요도 (그림 4)를 이용하여 전반적인 변수의 영향을 분석하였다. 최종 모델 결과에서 나아가, 최종 모델의 변수 조합을 변경하여 그 결과를 표 4에 제시하였다. 부분 의존도 그래프에서는 활성탄의 비표면적과 흡착질의 옥탄올-물 분배계수, 분자량이 분배 계수와 긍정적 상관 관계를 가지며 BC, TOC, 실험 기간, 활성탄 입자 크기와는 부정적 상관 관계를 가지는 것을 확인하였다. 나아가, 각 입력 변수들의 순열 특성 중요도를 비교하였고, Model 2_clean에서는 활성탄의 비표면적이, Model 2_apparent에서는 흡착질의 분자량이 가장 큰 중중요도를 가지는 것을 확인하였다.

표 3. 여섯 개의 입력 변수들 간의 피어슨 상관계수. 약어는 다음과 같음 BC: Black carbon, TOC: Total organic carbon, SSA: specific surface area, $\boldsymbol{log\, K_{ow}}$: logarithmic octanol−water partitioning coefficient, MW: molecular weight.

Input features	Pearson correlation coefficient
Adsorbate characteristics (MW and $\log K_ow$)	0.92
Sediment characteristics (TOC and BC)	0.84
Adsorbent characteristics (Particle size and AC SSA)	0.23

표 4. 최종 모델의 변수 조합 변화에 따른 예측 결과

	Input features								R^2	RMSE
	BC	TOC	$\log K_{ow}$	MW	AC SSA	Particle size	Duration	AC dose		
Model 2_clean			O*	O	O				0.91	0.38
			O		O				0.90	0.39
				O	O				0.88	0.45
Model 2_apparent	O	O		O		O	O		0.86	0.40
		O		O		O	O		0.86	0.41
	O	O	O	O		O	O		0.86	0.41
		O	O	O		O	O		0.85	0.42
	O	O	O			O	O		0.82	0.47
		O	O			O	O		0.82	0.47
	O			O		O	O		0.80	0.49
	O		O	O		O	O		0.79	0.50
	O		O			O	O		0.76	0.54

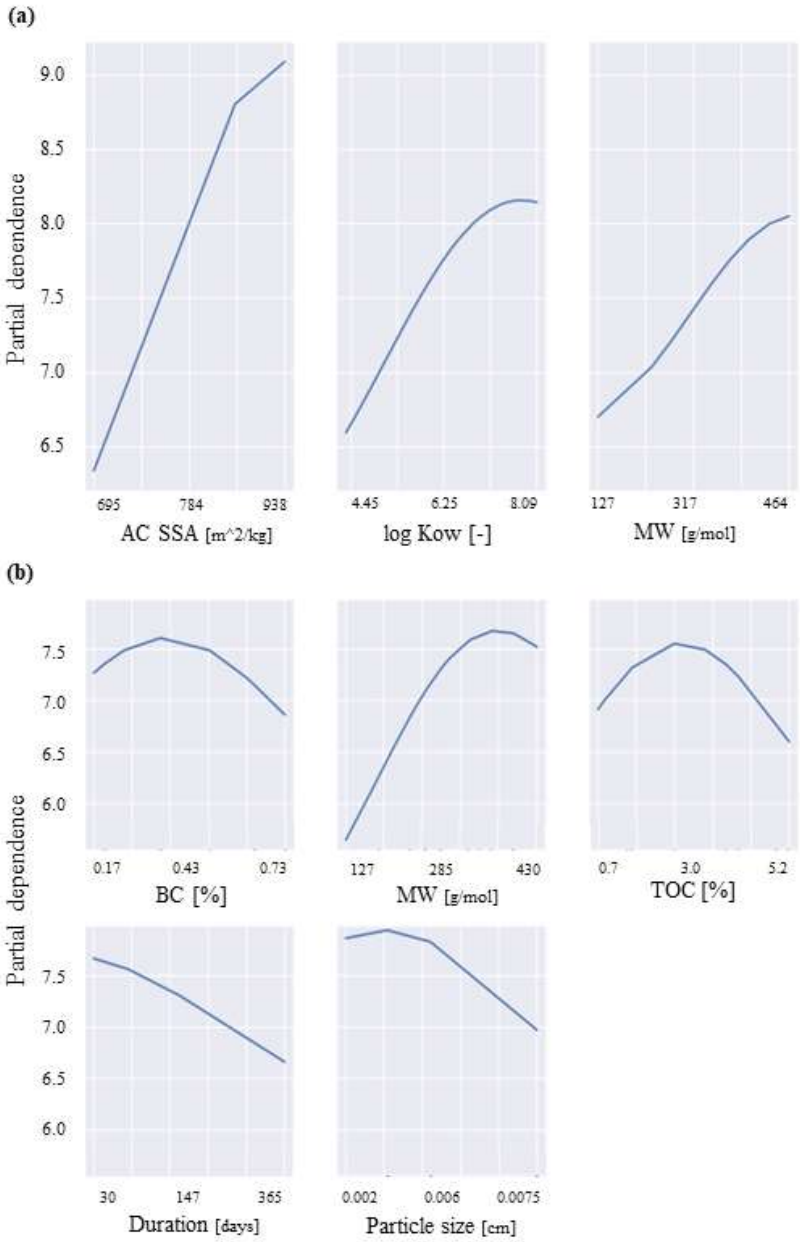

그림 3 최종 모델의 부분 의존도 그래프 (a) Model 2_clean (b) Model 2_apparent. 약어의 뜻은 다음과 같음 BC: Black carbon, MW: molecular weight, TOC: total organic carbon, AC SSA: AC specific surface area, $\boldsymbol{log\,K_{ow}}$: logarithmic octanol-water partitioning coefficient.

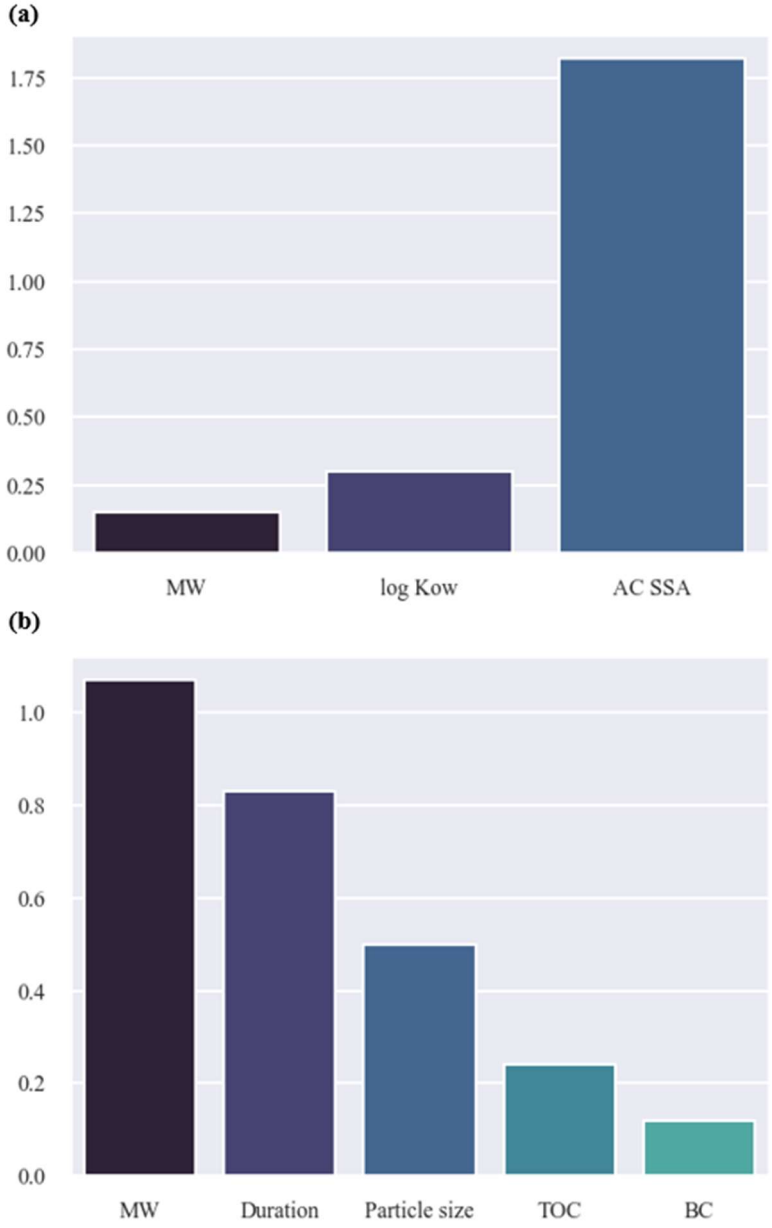

그림 4. 최종 모델의 순열 특성 중요도 (a) Model 2_clean (b) Model 2_apparent. 약어의 뜻은 다음과 같음 BC: Black carbon, MW: molecular weight, TOC: total organic carbon, AC SSA: AC specific surface area, $\boldsymbol{log\, K_{ow}}$: logarithmic octanol−water partitioning coefficient

1. 흡착질 특성의 영향

그림 3을 보면, Model 2_apparent와 Model 2_clean 두 모델 모두에서 분자량과 옥탄올-물 분배 계수가 증가함에 따라 분배 계수가 증가하는 양상을 볼 수 있다. 이는 흡착질이 가진 소수성으로 인해 가지는 활성탄 표면에 대한 친화도로 설명될 수 있다. 비극성이고 물에 잘 녹지 않는 소수성 물질은 소수성을 띤 표면을 가지고 있는 활성탄에 높은 친화도를 보인다. 결과적으로, 높은 소수성을 지닌 물질들이 활성탄에 더 많이 흡착되는 경향을 보인다. 이와 같은 유기 물질의 소수성과 흡착 분배 계수 간의 양의 관계는 Nam 외 (2014)에 의해서도 보고된 바 있다.

또한, 표 3의 Model 2_apparent에서 분자량이 다섯 변수들 중 가장 높은 중요도를 가진 것과 Model 2_clean에 사용된 세 변수들 중 두 가지가 물질의 소수성을 나타내는 지표인 분자량과 옥탄올-물 분배계수임을 통해 물질의 소수성이 분배 계수 예측에 있어 중요한 역할을 한다는 것을 알 수 있다.

입력 변수들의 관계와 각 변수들의 모델에 대한 영향을 분석하기 위하여 피어슨 상관 계수를 계산하였으며 (표 3), 이를 활용하여 최종 모델의 변수 구성을 변경한 후 모델의 성능 변화를 관찰하였다 (표 4).

표 3에 따르면, 흡착질 특성은 0.92의 높은 상관관계를 가진다. 따라서 Model 2_clean과 Model 2_apparent에 대해 두 가지 흡착질 특성 중 입력 변수로써 1) 분자량만 적용, 2) 옥탄올-물

분배계수만 적용, 3) 두 가지 모두 적용하여 모델의 성능을 비교하였다

흡착질 특성 간의 높은 상관관계에도 불구하고, 피어슨 상관계수와 모델의 성능은 연관성이 없는 것을 확인하였다. Model 2_clean에서는 옥탄올-물 분배계수와 분자량 모두를 흡착질 특성으로 적용한 경우 최고의 정확도를 보였으나 Model 2_apparent에서는 분자량만 적용하였을 때 최고의 정확도를 보였다.

기계 학습 알고리즘은 복잡한 데이터를 처리할 수 있는 뛰어난 유연성을 제공하지만, 개별 입력 변수들의 상호작용과 각 변수들의 영향에 대해서는 충분한 정보를 제공하지 못한다. 따라서 특정 입력 변수가 제외되고 다른 변수가 포함된 이유를 정확하게 판단할 수는 없지만, 피어슨 상관 계수와 모델 예측 결과가 직접적으로 연관되어 있지 않은 것을 확인하였다. 피어슨 상관 계수는 두 변수 간의 선형적 상관 관계를 나타내는 지표이며, 높은 상관 계수를 가진 특성들을 합치거나 하나의 변수로 대체할 수 있음을 의미하는 것은 아니다.

2. 퇴적물 특성의 영향

퇴적물의 흡착 저해 효과는 기존 문헌들에서 많이 보고되었으며, 이러한 저해 효과에 기여하는 기작은 여러가지가 있다. 첫번째로, 퇴적물 내의 용존 유기 물질들이 흡착제의 공극을 막아 오염물질 같은 흡착질이 내부 공극으로 접근하는 것을 방해하는 공극 막힘 현상이 있다. 두번째로, 퇴적물 내 유기물질들이 흡착제에 존재하는 흡착 공간을 두고 오염물질과

경쟁하여 먼저 흡착됨으로써 오염물질이 흡착될 수 있는 면적을 감소시키는 직접적 흡착 경쟁이 있다.

그림 3 (b)에서는 TOC가 증가할수록 겉보기 분배계수가 감소하는 경향을 확인할 수 있다. TOC는 퇴적물 내 총유기탄소량을 나타내는 지표로써, 용존유기탄소와 입자성유기탄소를 포함하는 개념이다. 이 중 용존 유기탄소는 공극 막힘과 직접적 흡착 경쟁을 일으켜 흡착 저해 현상을 일으키는 주된 원인으로 지목되어왔다[⑮]. 그러나 퇴적물 내 총유기탄소 중 용존 유기 탄소의 비율은 퇴적물마다 달라진다. 따라서 총유기탄소만으로는 퇴적물에 의한 흡착 저해 현상에 대한 정보를 충분히 제공하지 못하며, 이는 예측 모델 내 오차의 원인으로 작용할 수 있다. 그럼에도 불구하고 예측 모델은 겉보기 분배 계수와 TOC 값 간의 음의 상관관계를 포착하였으며, TOC값을 반영하여 겉보기 분배 계수를 예측하였다. 향후 연구에서는 퇴적물의 구성과 관련된 정보나 용존 유기물질에 대한 정보를 추가하여 퇴적물의 흡착 저해 현상에 대해 보다 심도 깊게 이해할 필요가 있다.

그림 3 (b)에서 BC가 증가할수록 겉보기 분배계수가 감소하는 경향도 확인할 수 있다. Black carbon은 TOC의 1-15 %를 차지하는 구성요소이기 때문에 TOC와 BC는 강한 상관관계를 가지며, 그림 4에서 나타난 피어슨 상관 계수 0.84를 통해 본 연구에서 사용된 데이터에서도 두 변수가 양의 상관관계를 가지는 것을 확인할 수 있다[16]. 이를 통해 TOC와 BC가 그림 3 (b)의 부분

---

⑮ Yu et al., 2011

16 Middelburg et al., 1999

의존도 그래프에서 비슷한 양상을 보이는 것을 이해할 수 있다.

퇴적물 특성의 피어슨 상관계수가 0.84로 높기 때문에 흡착질 특성과 마찬가지로 변수 구성 변화에 따른 모델 성능 변화를 관찰하였다. 흡착질 특성의 경우와 마찬가지로, 피어슨 상관계수와 모델 예측 성능 간의 관계를 발견하지 못하였다. 그러나 상위 6개 모델에 대해서는 항상 TOC가 입력 변수에 포함되어 있으며, 이는 입력 변수 중 TOC의 중요도를 알 수 있는 결과이다.

본 연구에서 개발한 겉보기 분배 계수 예측 모델은 입력 데이터에 TOC와 DOC가 가지는 차이를 반영하지 못했음에도 불구하고 퇴적물의 흡착 저해 효과를 포착하였다는 점에서 의의를 가진다. 또한 활성탄을 이용한 원위치 안정화 기술의 정화 효율 예측 및 평가에 있어 퇴적물의 특징을 고려하는 것이 중요하다는 것을 강조하며, 나아가 퇴적물의 구성을 입력 변수에 추가한다면 안정화 기술의 효율과 관련한 보다 심도 깊은 이해가 가능할 것이다.

3. 흡착제 특성의 영향

Brunauer-Emmett-Teller (BET) 분석으로 측정되는 활성탄의 비표면적은 활성탄 내에서 오염물질이 흡착될 수 있는 표면의 면적을 정량화해주기 때문에 흡착능과 밀접하게 연관되어있는 요소이다. 다양한 연구들이 활성탄의 비표면적과 흡착능

간의 양의 상관관계를 보고하였다[17]. 그림 3도 이러한 결과를 뒷받침하는데, 활성탄의 비표면적이 증가함에 따라 순수 분배 계수가 증가하는 양상을 관찰할 수 있다. 이는 비표면적 값이 높을수록, 즉 활성탄 내 표면적이 넓을 수록 오염물질이 흡착할 수 있는 공간이 더 많아져 흡착능이 증가함을 나타낸다.

반면, Model 2_apparent의 경우 활성탄의 비표면적을 변수에서 제외하였다. 이는 활성탄의 비표면적이 불순물이 없는 상태에서 측정되기 때문에, BET 분석을 통해 측정한 비표면적과 활성탄이 퇴적물로 주입되었을 때 오염물질이 실제 흡착할 수 있는 면적 간에 차이가 생겼기 때문일 수 있다. McDonough 외 (2008)는 용존 유기물질이 존재할 경우 활성탄의 비표면적이 감소하고, 그에 따라 활성탄의 흡착능이 감소함을 보고한 바 있다. 이로 인해 BET 분석을 통해 제시된 활성탄의 비표면적은 퇴적물 주입 활성탄 내의 실제 흡착 면적을 실질적으로 반영하지 못했기 때문에 Model 2_apparent에서 입력 변수로 제외되었을 수 있다.

반면, 활성탄의 입자 크기는 Model 2_apparent에서 입력변수로 포함되었다. 기존의 연구들에서 활성탄이 평형 상태에 도달하지 않았을 경우 활성탄의 입자 크기가 흡착능에 영향을 준다는 결과가 보도된 바 있다[18]. 작은 활성탄 입자는 더 짧은 확산 경로를 가지기 때문에 흡착이 더욱 빠르게 일어난다[19]. 본 연구에서 사용된 순수 분배 계수와 겉보기 분배 계수 모두 평형 상태에서의 분배 계수 값이 아니기 때문에, 활성탄 입자 크기가 커질수록 흡착 속도가

[17] Qu et al, 2019; Teng et al, 1998.
[18] Weber and Morris, 1964; Randtke and Snoeyink, 1983.
[19] Weber et al., 1983

감소하여 실험 종료 시에 측정된 겉보기 분배 계수가 줄어들었을 수도 있다. 또한, 활성탄 입자가 클 경우 용존 유기 물질에 의한 공극 막힘에 더욱 취약한 경향을 나타낸다. Corwin과 Summers (2010)는 활성탄 크기가 클수록 용존 유기물질에 의해 막힌 공극 뒤에 있는 미세 공극의 표면적 또한 커지기 때문에 겉보기 분배계수가 더 큰 폭으로 감소한다는 결과를 보고한 바 있다. 이러한 이유들로 인해 활성탄의 비표면적보다 입자 크기가 겉보기 분배 계수에 더 큰 영향을 미치므로 최종적인 입력 변수로써 선정되었을 수 있다.

4. 실험 조건의 영향

그림 3 (b)에서 실험 기간이 증가할 수록 겉보기 분배 계수가 감소함을 확인할 수 있다. 본 연구에서 참고한 문헌들에서는 겉보기 분배 계수를 결정하는 퇴적물-물-활성탄 실험 동안 반응 용기가 계속해서 교반 되었는데, 이러한 물리적 환경이 오염물질의 탈착에 취약한 환경을 조성하여 시간이 지날수록 겉보기 분배 계수가 감소하였을 수 있다.

그러나 실제 원위치 안정화 기술에서는 퇴적물이 계속해서 교반 되지 않고 주로 확산을 통해 오염물질이 이동하기 때문에, 실험 기간 내내 교반을 진행한 실험과 비교하였을 때 물리적 교란이 확연히 적다. 이러한 물리적 차이로 인해 원위치 안정화 기술에서도 겉보기 분배 계수가 시간에 따라 감소할 것이라고 결론을 내릴 수는 없지만, 비슷한 실험 조건에서는 교반과 같은 물리적 변화가 겉보기 분배 계수에 영향을 끼친다는 것을 알 수

있다.

제 4 장 결론 및 제언

제 1 절 한계 및 향후 발전 방향

본 연구에서는 활성탄의 흡착능에 대한 선행 연구에서 보고된 값을 사용하여 예측 모델을 개발하였지만, 활성탄의 흡착능을 보고한 문헌들 중에서도 일부 변수 데이터가 보고되지 않아 많은 데이터들이 입력 변수에서 제외되었다. 흡착질에 대한 정보는 검색을 통해 쉽게 얻을 수 있는 반면, 실험에 사용된 퇴적물의 총유기탄소나 black carbon 같은 측정값, 그리고 흡착제의 비표면적 등은 저자의 분석과 보고를 통해서만 얻을 수 있다. 연구 결과가 더욱 활발하게 활용되기 위해서는 흡착능과 유관한 변수들을 문헌 내에 모두 보고하여 데이터의 향후 활용과 논의를 용이하게 할 필요가 있다.

흡착제의 공극 크기 분포 (pore size distribution)는 흡착질의 흡착제 내부 공극에 대한 접근성과 흡착제 표면과 흡착질 간의 접촉을 결정하기 때문에 흡착능에 큰 영향을 주는 인자이다[20]. 또한 여러 흡착질 간의 흡착 경쟁에도 영향을 주기 때문에 이는 용존 유기오염물질의 흡착 저해 효과와 긴밀하게 연관 되어있다[21]. 그러나, 활성탄의 공극 크기 분포를 측정하여 보고한 연구의 수가 많지 않았기 때문에 해당 정보를 예측 모델 개발에 활용할 수

[20] Lorenc-Grabowsk., 2016.; Pelekani & Vernon., 2000
[21] (Pelekani & Vernon., 2000

없었다. 공극 크기 분포가 앞서 언급한 이유로 인해 흡착에 큰 영향을 끼치는 것을 감안하면 흡착에 대한 연구들을 진행할 때 이에 관한 데이터를 포함하는 것이 필수적이다[22].

퇴적물 내의 용존 유기물질은 흡착 저해 효과에 큰 영향을 준다. 용존 유기물질은 공극 막힘이나 직접 흡착 경쟁을 통해 흡착질의 흡착을 방해한다[23]. 그러나 퇴적물의 용존유기물질에 관한 정보를 보고한 연구의 수가 충분하지 않아 이 역시 예측 모델 개발을 위한 인자로써 사용하지 못했다. 3장 2절의 퇴적물 특성의 영향에 대한 논의에서 언급하였듯이, 흡착 저해 효과를 반영하여 예측 모델의 정확도를 더욱 높이기 위해서는 퇴적물의 용존유기물질에 대한 정보를 모델에 반영하는 것이 필수적이다.

본 연구에서 사용된 겉보기 분배 계수는 교반기를 이용하여 반응용기를 계속해서 교반 하는 slurry phase 실험을 통해 얻어졌다. 그러나 실제 현장 조건에서는 퇴적물 내 오염물질의 이동은 인위적인 기계적 혼합 없이 확산을 통해서만 진행된다. Slurry phase 실험을 통해 얻은 분배계수는 활성탄의 흡착능을 결정하는 인자들에 대한 중요한 정보를 제공하지만, 오염물질의 확산 영향이나 정화 기간 동안 퇴적물, 공극수, 안정화제 상의 농도 변화에 대한 정보는 제공하지 못한다. 퇴적물 내에서 일어나는 흡착 과정을 총체적으로 이해하기 위해서는 현장 조건에서 진행한 실험 결과와 흡착 속도 실험 결과를 활용하여 모델에 활용하는 것이 필수적이다.

흡착능 예측모델을 발전시키기 위해서는 퇴적물 혹은 흡착제에

---

[22] Dubinin. 1989
[23] Pignatello et al., 2006; Cornelissen and Gustafsson, 2006

대한 추가적인 데이터 확보가 필수적이다. 또한, 현장 규모의 실험에서 도출한 결과값과 흡착 속도에 대한 정보를 반영함으로써 흡착 저해 현상과 기술 적용 기간 동안의 농도 변화 과정을 더욱 폭넓게 이해할 수 있다.

제 2 절 결과 및 의의

본 연구에서는 물-활성탄 시스템에서 활성탄의 흡착능인 순수 분배 계수와 퇴적물-물-활성탄 시스템에서 활성탄의 흡착능인 겉보기 분배 계수에 대한 기계 학습 기반 예측 모델을 개발하였다. 후자의 경우 퇴적물이 존재할 때의 활성탄 흡착능을 나타내 주기 때문에 오염 퇴적물 원위치 안정화 기술의 효율을 보다 직접적으로 나타내 줄 수 있다. 최종 모델은 흡착능 지표인 순수 분배 계수와 겉보기 분배 계수 모두에 대해 높은 정확도를 보였다. 순수 분배 계수를 예측하는 Model 2_clean의 경우 결정 계수 0.91과 평균 제곱근 오차 0.38을 보였고, 겉보기 분배 계수를 예측하는 Model 2_apparent의 경우 결정 계수 0.86과 평균 제곱근 오차 0.40을 보였다.

각 모델에 사용된 변수에 대한 분석을 진행함으로써 변수들의 중요도와 각 변수들이 결과값에 미치는 영향을 파악하였다. 순수 분배 계수에 대해서는 흡착질의 특성인 분자량과 옥탄올-물 분배계수, 그리고 활성탄의 비표면적이 예측에 큰 영향을 주는 것을 확인하였다. 반면, 겉보기 분배 계수에 대해서는 흡착질 특성, 흡착제 특성, 퇴적물 특성, 실험 조건 모두가 분배 계수 예측에 관여하였다. 특히 퇴적물 내 총유기탄소와 black carbon이

증가할수록 겉보기 분배 상수가 감소하는 흡착 저해 현상을 관찰할 수 있었다. 이를 통해 순수 분배 계수와 겉보기 분배 계수 예측에는 서로 다른 입력 인자들이 관여하며, 특히 원위치 안정화 기술에 직접적인 영향을 주는 겉보기 분배 계수에 대해서는 퇴적물 특성을 반영하여야 더욱 정확한 값을 예측할 수 있다는 것을 알 수 있다.

향후 연구는 활성탄의 흡착능과 연관된 데이터인 퇴적물의 용존유기물질이나 흡착제의 공극 분포에 대한 데이터를 더 확보하여 모델의 정확도를 향상시키고, 현장 규모의 실험 결과와 흡착 속도를 측정한 실험의 결과를 확보하여 모델이 실제 현장 조건에서도 적용될 수 있도록 해야 한다.

종합적으로, 본 연구는 기계 학습을 사용하여 퇴적물 주입 활성탄의 겉보기 분배 계수를 효과적으로 예측하였으며, 이를 통해 활성탄을 이용한 원위치 안정화 기술의 효율적이고 안정적인 관리에 기여하였다.

참고 문헌

Ahn, S., Werner, D., & Luthy, R. G. (2008). Modeling PAH mass transfer in a slurry of contaminated soil or sediment amended with organic sorbents. Water Research, 42(12), 2931-2942.

Cho, Y. M., Smithenry, D. W., Ghosh, U., Kennedy, A. J., Millward, R. N., Bridges, T. S., & Luthy, R. G. (2007). Field methods for amending marine sediment with activated carbon and assessing treatment effectiveness. Marine environmental research, 64(5), 541-555.

Choi, Y., Cho, Y. M., Gala, W. R., & Luthy, R. G. (2013). Measurement and modeling of activated carbon performance for the sequestration of parent-and alkylated-polycyclic aromatic hydrocarbons in petroleum-impacted sediments. Environmental science & technology, 47(2), 1024-1032.

Choi, Y., Cho, Y. M., Werner, D., & Luthy, R. G. (2014). In situ sequestration of hydrophobic organic contaminants in sediments under stagnant contact with activated carbon. 2. Mass transfer modeling. Environmental science & technology, 48(3), 1843-1850.

Cornelissen, G., & Gustafsson, Ö. (2006). Effects of added PAHs and precipitated humic acid coatings on phenanthrene sorption to environmental Black carbon. Environmental

Pollution, 141(3), 526-531.

Cornelissen, G., Breedveld, G. D., Kalaitzidis, S., Christanis, K., Kibsgaard, A., & Oen, A. M. (2006). Strong sorption of native PAHs to pyrogenic and unburned carbonaceous geosorbents in sediments. Environmental science & technology, 40(4), 1197-1203.

Corwin, C. J., & Summers, R. S. (2010). Scaling trace organic contaminant adsorption capacity by granular activated carbon. Environmental science & technology, 44(14), 5403-5408.

Corwin, C. J., & Summers, R. S. (2011). Adsorption and desorption of trace organic contaminants from granular activated carbon adsorbers after intermittent loading and throughout backwash cycles. Water research, 45(2), 417-426.

Dai, Q., Min, X., & Weng, M. (2016). A review of polychlorinated biphenyls (PCBs) pollution in indoor air environment. Journal of the Air & Waste Management Association, 66(10), 941-950.

Drwal, E., Rak, A., & Gregoraszczuk, E. L. (2019). polycyclic aromatic hydrocarbons (PAHs)—action on placental function and health risks in future life of newborns. Toxicology, 411, 133-142.

Dubinin, M. M. (1989). Fundamentals of the theory of adsorption in micropores of carbon adsorbents:

Characteristics of their adsorption properties and microporous structures. Carbon, 27(3), 457–467.

Gomez-Eyles, J. L., Yupanqui, C., Beckingham, B., Riedel, G., Gilmour, C., & Ghosh, U. (2013). Evaluation of biochars and activated carbons for in situ remediation of sediments impacted with organics, mercury, and methylmercury. Environmental science & technology, 47(23), 13721–13729.

Gusenbauer, M., Oezelt, H., Fischbacher, J., Kovacs, A., Zhao, P., Woodcock, T.G., Schrefl, T., 2020.

Hale, S. E., & Werner, D. (2010). Modeling the mass transfer of hydrophobic organic pollutants in briefly and continuously mixed sediment after amendment with activated carbon. Environmental science & technology, 44(9), 3381–3387.

Hale, S. E., Kwon, S., Ghosh, U., & Werner, D. (2010). Polychlorinated biphenyl sorption to activated carbon and the attenuation caused by sediment. Global NEST Journal, 12(3), 318–326.

Jamil, S., Loganathan, P., Listowski, A., Kandasamy, J., Khourshed, C., & Vigneswaran, S. (2019). Simultaneous removal of natural organic matter and micro-organic pollutants from reverse osmosis concentrate using granular activated carbon. Water research, 155, 106–114.

Jonker, M. T., & Koelmans, A. A. (2002). Sorption of polycyclic aromatic hydrocarbons and polychlorinated biphenyls to

soot and soot-like materials in the aqueous environment: mechanistic considerations. Environmental science & technology, 36(17), 3725-3734.

Koelmans, A. A., Kaag, K., Sneekes, A., & Peeters, E. T. H. M. (2009). Triple domain in situ sorption modeling of organochlorine pesticides, polychlorobiphenyls, polyaromatic hydrocarbons, polychlorinated dibenzo-p-dioxins, and polychlorinated dibenzofurans in aquatic sediments. Environmental science & technology, 43(23), 8847-8853.

Kwon, S., & Pignatello, J. J. (2005). Effect of natural organic substances on the surface and adsorptive properties of environmental black carbon (char): pseudo pore blockage by model lipid components and its implications for N2-probed surface properties of natural sorbents. Environmental science & technology, 39(20), 7932-7939.

L. Breiman, "Random Forests", Machine Learning, 45(1), 5-32, 2001.

Lateef, S. A., Oyehan, I. A., Oyehan, T. A., & Saleh, T. A. (2022). Intelligent modeling of dye removal by aluminized activated carbon. Environmental Science and Pollution Research, 29(39), 58950-58962.

Lofrano, G., Libralato, G., Minetto, D., De Gisi, S., Todaro, F., Conte, B., ... & Notarnicola, M. (2017). In situ remediation of contaminated marinesediment: an overview.

Environmental Science and Pollution Research, 24, 5189-5206.

Lorenc-Grabowska, E. (2016). Effect of micropore size distribution on phenol adsorption on steam activated carbons. Adsorption, 22(4-6), 599-607.

Majone, M., Verdini, R., Aulenta, F., Rossetti, S., Tandoi, V., Kalogerakis, N., ... & Fava, F. (2015). In situ groundwater and sediment bioremediation: barriers and perspectives at European contaminated sites. New biotechnology, 32(1), 133-146.

McDonough, K. M., Fairey, J. L., & Lowry, G. V. (2008). Adsorption of polychlorinated biphenyls to activated carbon: Equilibrium isotherms and a preliminary assessment of the effect of dissolved organic matter and biofilm loadings. Water Research, 42(3), 575-584.

McDonough, K. M., Fairey, J. L., & Lowry, G. V. (2008). Adsorption of polychlorinated biphenyls to activated carbon: Equilibrium isotherms and a preliminary assessment of the effect of dissolved organic matter and biofilm loadings. Water Research, 42(3), 575-584.

Middelburg, J. J.; Nieuwenhuize, J.; Van Breugel, P. Mar. Chem., 1999, 65, 245.

Gustafsson, O¨ .; Gschwend, P. M. Geochim. Cosmichim. Acta., 1998, 62, 465.

Nam, S. W., Choi, D. J., Kim, S. K., Her, N., & Zoh, K. D.

(2014). Adsorption characteristics of selected hydrophilic and hydrophobic micropollutants in water using activated carbon. Journal of hazardous materials, 270, 144-152.

Oen, A. M., Beckingham, B., Ghosh, U., Krusa°, M. E., Luthy, R. G., Hartnik, T., ... & Cornelissen, G. (2012). Sorption of organic compounds to fresh and field-aged activated carbons in soils and sediments. Environmental science & technology, 46(2), 810-817.

Palansooriya, K. N., Li, J., Dissanayake, P. D., Suvarna, M., Li, L., Yuan, X., ... & Ok, Y. S. (2022). Prediction of soil heavy metal immobilization by biochar using machine learning. Environmental Science & Technology, 56(7), 4187-4198.

Pelekani, C., & Snoeyink, V. L. (2000). Competitive adsorption between atrazine and methylene blue on activated carbon: the importance of pore size distribution. Carbon, 38(10), 1423-1436.

Pignatello, J. J., Kwon, S., & Lu, Y. (2006). Effect of natural organic substances on the surface and adsorptive properties of environmental black carbon (char): attenuation of surface activity by humic and fulvic acids. Environmental science & technology, 40(24), 7757-7763.

Qu, W., Yuan, T., Yin, G., Xu, S., Zhang, Q., & Su, H. (2019). Effect of properties of activated carbon on malachite green adsorption. Fuel, 249, 45-53.

Rakowska, M. I., Kupryianchyk, D., Harmsen, J., Grotenhuis, T.,

& Koelmans, A. A. (2012). In situ remediation of contaminated sediments using carbonaceous materials. Environmental Toxicology and Chemistry, 31(4), 693-704.

Randtke, S. J., & Snoeyink, V. L. (1983). Evaluating GAC adsorptive capacity. Journal-American Water Works Association, 75(8), 406-413.

Sun, X., & Ghosh, U. (2007). PCB bioavailability control in Lumbriculus variegatus through different modes of activated carbon addition to sediments. Environmental science & technology, 41(13), 4774-4780.

Sun, X., & Ghosh, U. (2008). The effect of activated carbon on partitioning, desorption, and biouptake of native polychlorinated biphenyls in four freshwater sediments. Environmental Toxicology and Chemistry: An International Journal, 27(11), 2287-2295.

Tehrany, E. A., Fournier, F., & Desobry, S. (2004). Simple method to calculate octanol–water partition coefficient of organic compounds. Journal of food engineering, 64(3), 315-320.

Teng, H., & Hsieh, C. T. (1998). Influence of surface characteristics on liquid-phase adsorption of phenol by activated carbons prepared from bituminous coal. Industrial & engineering chemistry research, 37(9), 3618-3624.

Wang, X., Liu, Y., Tao, S., & Xing, B. (2010). Relative importance of multiple mechanisms in sorption of organic

compounds by multiwalled carbon nanotubes. Carbon, 48(13), 3721-3728.

Weber Jr, W. J., & Morris, J. C. (1964). Equilibria and capacities for adsorption on carbon. Journal of the Sanitary Engineering Division, 90(3), 79-108.

Weber Jr, W. J., Voice, T. C., & Jodellah, A. (1983). Adsorption of humic substances: the effects of heterogeneity and system characteristics. Journal-American Water Works Association, 75(12), 612-619.

Werner, D., Ghosh, U., & Luthy, R. G. (2006). Modeling polychlorinated biphenyl mass transfer after amendment of contaminated sediment with activated carbon. Environmental science & technology, 40(13), 4211-4218.

Yu, H., Huang, G. H., An, C. J., & Wei, J. (2011). Combined effects of DOM extracted from site soil/compost and biosurfactant on the sorption and desorption of PAHs in a soil–water system. Journal of Hazardous Materials, 190(1-3), 883-890.

Zhang, K., Zhong, S., & Zhang, H. (2020). Predicting aqueous adsorption of organic compounds onto biochars, carbon nanotubes, granular activated carbons, and resins with machine learning. Environmental Science & Technology, 54(11), 7008-7018.

Zhang, S., Shao, T., Kose, H. S., & Karanfil, T. (2010). Adsorption of aromatic compounds by carbonaceous

adsorbents: a comparative study on granular activated carbon, activated carbon fiber, and carbon nanotubes. Environmental science & technology, 44(16), 6377–6383.

Zhou, Y., Miao, D., Gomez-Eyles, J. L., Ghosh, U., Bi, M., Li, J., & Ren, F. (2022). Comparative study on polychlorinated biphenyl sorption to activated carbon and biochar and the influence of natural organic matter. Chemosphere, 287, 132239.

Zhu, X., He, M., Sun, Y., Xu, Z., Wan, Z., Hou, D., ... & Tsang, D. C. (2022). Insights into the adsorption of pharmaceuticals and personal care products (PPCPs) on biochar and activated carbon with the aid of machine learning. Journal of Hazardous Materials, 423, 127060.

Zimmerman, J. R., Ghosh, U., Millward, R. N., Bridges, T. S., & Luthy, R. G. (2004). Addition of carbon sorbents to reduce PCB and PAH bioavailability in marine sediments: Physicochemical tests. Environmental Science & Technology, 38(20), 5458–5464.

Zimmerman, J. R., Werner, D., Ghosh, U., Millward, R. N., Bridges, T. S., & Luthy, R. G. (2005). Effects of dose and particle size on activated carbon treatment to sequester polychlorinated biphenyls and polycyclic aromatic hydrocarbons in marine sediments. Environmental Toxicology and Chemistry: An International Journal, 24(7), 1594–1601.

부록

부표 1. 표 2 내 모델들의 하이퍼 파라미터 및 사용 변수

Model framework	Output value	Regression model	Prediction score		Selected features	Hyperparameter
			R^2	RMSE		
Model 1	$\log K_{AC,clean}$	SVR	0.90	0.40	AC SSA, $\log K_{ow}$, TOC, Duration, Particle size	C= 10, gamma= 1, kernel= 'rbf'
		RFR	0.90	0.40	BC, AC SSA, $\log K_{ow}$, MW, TOC	max_depth= 10, min_samples_split= 2, n_estimators= 26
	$\log K_{AC,apparent}$	SVR	0.82	0.46	MW, $\log K_{ow}$, AC SSA, AC size, TOC, BC, and Duration	C= 10, gamma= 1, kernel= 'rbf'
		RFR	0.83	0.43	BC, AC SSA, $\log K_{ow}$, MW, TOC	max_depth= 10, min_samples_split= 2, n_estimators= 26
Model 2	$\log K_{AC,clean}$	SVR	0.91	0.38	AC SSA, $\log K_{ow}$, MW	C= 10, gamma= 1, kernel= 'rbf'
		RFR	0.89	0.43	AC SSA, $\log K_{ow}$, Duration, Particle size	max_depth= 5, min_samples_split= 4, n_estimators= 26

	$\log K_{AC,apparent}$	SVR	0.86	0.40	BC, MW, TOC, Duration, Particle size	C= 10, gamma= 1, kernel= 'rbf'
		RFR	0.86	0.43	BC, $\log K_{ow}$, MW, TOC, Particle size	max_depth= 7, min_samples_split= 2, n_estimators= 29
Model 3	$\log K_{AC,apparent}$	SVR	0.86	0.42	Predicted $\log K_{AC,clean}$,BC, MW, TOC, Duration, Particle size	C= 10, gamma= 1, kernel= 'rbf'
		RFR	0.83	0.44	Predicted $\log K_{AC,clean}$,AC, AC SSA, MW, TOC, Duration, Particle size	max_depth= 12, min_samples_split= 2, n_estimators= 20

부표 2. 모델 학습에 사용된 전체 데이터

Pollutant name	MW	$log\,K_{ow}$	particle size	BET	TOC	BC	duration	reference	log Kac	구분
Ant	178	4.45	0.0075	938	0	0	365	(1)	6.3	Clean
NAPH	127.1	5.35	0.0075	938	0	0	365	(1)	7.6	Clean
BaA	228	5.75	0.0075	938	0	0	365	(1)	8.4	Clean
PCB101	326	6.38	0.0075	938	0	0	365	(1)	9.1	Clean
PCB153	361	6.92	0.0075	938	0	0	365	(1)	8.9	Clean
PCB180	395	7.35	0.0075	938	0	0	365	(1)	9.2	Clean
PCB19	188.66	5.19	0.0045	875	0	0	183	(2)	7.39	Clean
PCB18	257	5.24	0.0045	875	0	0	183	(2)	8.21	Clean
PCB45	292	5.53	0.0045	875	0	0	183	(2)	8.7	Clean
PCB22	257	5.58	0.0045	875	0	0	183	(2)	8.22	Clean
PCB51	292	5.63	0.0045	875	0	0	183	(2)	8.06	Clean
PCB26	257	5.66	0.0045	875	0	0	183	(2)	8.4	Clean
PCB40	292	5.66	0.0045	875	0	0	183	(2)	8.77	Clean
PCB25	257	5.67	0.0045	875	0	0	183	(2)	8.42	Clean
PCB28	257	5.67	0.0045	875	0	0	183	(2)	8.47	Clean
PCB31	257	5.67	0.0045	875	0	0	183	(2)	8.41	Clean

PCB44	292	5.75	0.0045	875	0	0	183	(2)	8.54	Clean
PCB42	292	5.76	0.0045	875	0	0	183	(2)	8.99	Clean
PCB47	292	5.85	0.0045	875	0	0	183	(2)	8.91	Clean
PCB64	292	5.95	0.0045	875	0	0	183	(2)	8.61	Clean
PCB91	326.4	6.13	0.0045	875	0	0	183	(2)	8.08	Clean
PCB74	292	6.2	0.0045	875	0	0	183	(2)	8.87	Clean
PCB118	326.4	6.26	0.0045	875	0	0	183	(2)	9.01	Clean
PCB83	326	6.26	0.0045	875	0	0	183	(2)	8.64	Clean
PCB97	326	6.29	0.0045	875	0	0	183	(2)	8.58	Clean
PCB101	326	6.38	0.0045	875	0	0	183	(2)	9.05	Clean
PCB99	326	6.39	0.0045	875	0	0	183	(2)	9.59	Clean
Di	222	5.1	0.0075	938	0	0	30	(3)	8.12	Clean
Tri	256	5.5	0.0075	938	0	0	30	(3)	8.3	Clean
Tetra	289	5.9	0.0075	938	0	0	30	(3)	9.17	Clean
Penta	323	6.3	0.0075	938	0	0	30	(3)	9.5	Clean
Hexa	357	6.7	0.0075	938	0	0	30	(3)	9.67	Clean
Hepta	391	7.1	0.0075	938	0	0	30	(3)	9.79	Clean
PCB18	257	5.24	0.0044	695.8	0	0	90	(4)	5.24	Clean
PCB45	292	5.53	0.0044	695.8	0	0	90	(4)	5.07	Clean
PCB46	292	5.53	0.0044	695.8	0	0	90	(4)	4.88	Clean

PCB22	257	5.58	0.0044	695.8	0	0	90	(4)	5.6	Clean
PCB51	292	5.63	0.0044	695.8	0	0	90	(4)	5.58	Clean
PCB26	257	5.66	0.0044	695.8	0	0	90	(4)	5.33	Clean
PCB40	292	5.66	0.0044	695.8	0	0	90	(4)	5.19	Clean
PCB25	257	5.67	0.0044	695.8	0	0	90	(4)	5.47	Clean
PCB28	257	5.67	0.0044	695.8	0	0	90	(4)	5.89	Clean
PCB31	257	5.67	0.0044	695.8	0	0	90	(4)	5.82	Clean
PCB44	292	5.75	0.0044	695.8	0	0	90	(4)	5.58	Clean
PCB42	292	5.76	0.0044	695.8	0	0	90	(4)	5.6	Clean
PCB48	292	5.78	0.0044	695.8	0	0	90	(4)	5.17	Clean
PCB37	257	5.83	0.0044	695.8	0	0	90	(4)	5.53	Clean
PCB47	292	5.85	0.0044	695.8	0	0	90	(4)	5.22	Clean
PCB49	292	5.85	0.0044	695.8	0	0	90	(4)	5.59	Clean
PCB64	292	5.95	0.0044	695.8	0	0	90	(4)	5.6	Clean
PCB74	292	6.2	0.0044	695.8	0	0	90	(4)	5.84	Clean
PCB136	361	6.22	0.0044	695.8	0	0	90	(4)	6.66	Clean
PCB97	326	6.29	0.0044	695.8	0	0	90	(4)	6.93	Clean
PCB85	326	6.3	0.0044	695.8	0	0	90	(4)	6.39	Clean
PCB101	326	6.38	0.0044	695.8	0	0	90	(4)	5.6	Clean
PCB99	326	6.39	0.0044	695.8	0	0	90	(4)	6.43	Clean

PCB134	361	6.55	0.0044	695.8	0	0	90	(4)	7.01	Clean
PCB132	361	6.58	0.0044	695.8	0	0	90	(4)	7.06	Clean
PCB128	361	6.74	0.0044	695.8	0	0	90	(4)	6.9	Clean
PCB141	361	6.82	0.0044	695.8	0	0	90	(4)	7.52	Clean
PCB146	361	6.89	0.0044	695.8	0	0	90	(4)	7.2	Clean
PCB153	361	6.92	0.0044	695.8	0	0	90	(4)	7.57	Clean
PCB177	395	7.08	0.0044	695.8	0	0	90	(4)	7.63	Clean
PCB174	395	7.11	0.0044	695.8	0	0	90	(4)	7.68	Clean
PCB185	395	7.11	0.0044	695.8	0	0	90	(4)	7.58	Clean
PCB175	395	7.17	0.0044	695.8	0	0	90	(4)	7.84	Clean
PCB183	395	7.2	0.0044	695.8	0	0	90	(4)	7.7	Clean
PCB199	430	7.2	0.0044	695.8	0	0	90	(4)	6.74	Clean
PCB193	395	7.52	0.0044	695.8	0	0	90	(4)	7.76	Clean
PCB191	395	7.55	0.0044	695.8	0	0	90	(4)	7.3	Clean
PCB198	430	7.62	0.0044	695.8	0	0	90	(4)	7.71	Clean
PCB201	430	7.62	0.0044	695.8	0	0	90	(4)	8.26	Clean
PCB207	464	7.74	0.0044	695.8	0	0	90	(4)	7.89	Clean
PCB194	430	7.8	0.0044	695.8	0	0	90	(4)	7.75	Clean
PCB205	430	8	0.0044	695.8	0	0	90	(4)	8.04	Clean
PCB206	464	8.09	0.0044	695.8	0	0	90	(4)	7.99	Clean

PCB29	257	5.6	0.0075	938	0.7	0.17	90	(5)	6.3	Apparent
PCB(52+49)	292	5.9	0.0075	938	0.7	0.17	90	(5)	6.05	Apparent
PCB69	292	6	0.0075	938	0.7	0.17	90	(5)	6.1	Apparent
PCB103	326	6.2	0.0075	938	0.7	0.17	90	(5)	5.8	Apparent
PCB136	361	6.22	0.0075	938	0.7	0.17	90	(5)	6.45	Apparent
PCB101	326	6.38	0.0075	938	0.7	0.17	90	(5)	6.7	Apparent
PCB153	361	6.92	0.0075	938	0.7	0.17	90	(5)	6.75	Apparent
PCB180	395	7.4	0.0075	938	0.7	0.17	90	(5)	7.1	Apparent
di-CB	222	5.1	0.0075	938	0.83	0.258	30	(6)	7	Apparent
tri-CB	256	5.5	0.0075	938	0.83	0.258	30	(6)	6.6	Apparent
tetra-CB	289	5.9	0.0075	938	0.83	0.258	30	(6)	7	Apparent
penta-CB	323	6.3	0.0075	938	0.83	0.258	30	(6)	7.3	Apparent
tri-CB	323	6.3	0.0075	938	1.5	0.2	180	(7)	6.9	Apparent
tri-CB	323	6.3	0.0075	938	1.5	0.2	180	(7)	6.9	Apparent
hexa-CB	357	6.7	0.0075	938	1.5	0.2	180	(7)	7.6	Apparent
hexa-CB	357	6.7	0.0075	938	1.5	0.2	180	(7)	7.9	Apparent
tetra-CB	357	6.7	0.0075	938	1.5	0.2	180	(7)	7.5	Apparent
tetra-CB	357	6.7	0.0075	938	1.5	0.2	180	(7)	7.5	Apparent
hepta-CB	391	7.1	0.0075	938	1.5	0.2	180	(7)	7.7	Apparent
hepta-CB	391	7.1	0.0075	938	1.5	0.2	180	(7)	7.8	Apparent

penta-CB	391	7.1	0.0075	938	1.5	0.2	180	(7)	7.6	Apparent
penta-CB	391	7.1	0.0075	938	1.5	0.2	180	(7)	7.6	Apparent
octa-CB	430	7.5	0.0075	938	1.5	0.2	180	(7)	7.7	Apparent
octa-CB	430	7.5	0.0075	938	1.5	0.2	180	(7)	7.8	Apparent
tri-CB	323	6.3	0.0075	938	1.5	0.2	30	(8)	6.7	Apparent
tri-CB	323	6.3	0.0075	938	1.5	0.2	30	(8)	7.2	Apparent
tri-CB	323	6.3	0.0075	938	1.5	0.2	30	(8)	6.9	Apparent
hexa-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	7.8	Apparent
hexa-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	8	Apparent
hexa-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	7.7	Apparent
tetra-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	7.1	Apparent
tetra-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	7.8	Apparent
tetra-CB	357	6.7	0.0075	938	1.5	0.2	30	(8)	7.5	Apparent
hepta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.9	Apparent
hepta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.8	Apparent
hepta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.6	Apparent
penta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.3	Apparent
penta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.7	Apparent
penta-CB	391	7.1	0.0075	938	1.5	0.2	30	(8)	7.4	Apparent
octa-CB	430	7.5	0.0075	938	1.5	0.2	30	(8)	7.7	Apparent

octa-CB	430	7.5	0.0075	938	1.5	0.2	30	(8)	7.7	Apparent
octa-CB	430	7.5	0.0075	938	1.5	0.2	30	(8)	7.1	Apparent
PHEN	178	4.07	0.002	1100	2.5	0.51	30	(5)	8.4	Apparent
PYR	202.25	4.54	0.002	1100	2.5	0.51	30	(5)	8.5	Apparent
ANT-d10	188	4.57	0.002	1100	2.5	0.51	30	(5)	7.8	Apparent
FLUO	166.2	4.97	0.002	1100	2.5	0.51	30	(5)	8.7	Apparent
FLUO-d10	166.2	4.97	0.002	1100	2.5	0.51	30	(5)	8.15	Apparent
PCB32	257	5.4	0.002	1100	2.5	0.51	30	(5)	9.8	Apparent
CHRS-d12	240	5.67	0.002	1100	2.5	0.51	30	(5)	7.85	Apparent
B(a)PYR	252.3	6.05	0.002	1100	2.5	0.51	30	(5)	9.5	Apparent
di-CB	222	5.1	0.0075	938	3.22	0.64	30	(6)	6.6	Apparent
tri-CB	256	5.5	0.0075	938	3.22	0.64	30	(6)	7.6	Apparent
tetra-CB	289	5.9	0.0075	938	3.22	0.64	30	(6)	7.3	Apparent
penta-CB	323	6.3	0.0075	938	3.22	0.64	30	(6)	7.1	Apparent
di-CB	222	5.1	0.0075	938	3.7	0.616	30	(6)	7.2	Apparent
tri-CB	256	5.5	0.0075	938	3.7	0.616	30	(6)	7.6	Apparent
tetra-CB	289	5.9	0.0075	938	3.7	0.616	30	(6)	7.2	Apparent
penta-CB	323	6.3	0.0075	938	3.7	0.616	30	(6)	7.1	Apparent
phenanthrene	178	4.07	0.0075	938	4	0.73	365	(9)	4.5	Apparent

phenanthrene	178	4.07	0.0032	938	4	0.73	365	(9)	4.5	Apparent
anthracene	178.2	4.45	0.0075	938	4	0.73	365	(9)	4.3	Apparent
anthracene	178.2	4.45	0.0032	938	4	0.73	365	(9)	4.4	Apparent
pyrene	202.25	4.54	0.0075	938	4	0.73	365	(9)	5	Apparent
pyrene	202.25	4.54	0.0032	938	4	0.73	365	(9)	5	Apparent
fluorene	166.2	4.97	0.0075	938	4	0.73	365	(9)	4	Apparent
fluorene	166.2	4.97	0.0032	938	4	0.73	365	(9)	4	Apparent
fluoranthene	166.2	5.16	0.0075	938	4	0.73	365	(9)	4.9	Apparent
fluoranthene	166.2	5.16	0.0032	938	4	0.73	365	(9)	4.9	Apparent
benz[a]anthracene	228	5.76	0.0075	938	4	0.73	365	(9)	5.7	Apparent
benz[a]anthracene	228	5.76	0.0032	938	4	0.73	365	(9)	5.7	Apparent
benzo[b]fluoranthene	252.3	5.78	0.0075	938	4	0.73	365	(9)	6.3	Apparent
benzo[b]fluoranthene	252.3	5.78	0.0032	938	4	0.73	365	(9)	6.3	Apparent
chrysene	228.3	5.81	0.0075	938	4	0.73	365	(9)	5.7	Apparent
chrysene	228.3	5.81	0.0032	938	4	0.73	365	(9)	5.7	Apparent
benzo[k]fluoranthene	252.3	6.11	0.0075	938	4	0.73	365	(9)	6.4	Apparent
benzo[k]fluoranthene	252.3	6.11	0.0032	938	4	0.73	365	(9)	6.4	Apparent
benzo[a]pyrene	252.3	6.13	0.0075	938	4	0.73	365	(9)	6.4	Apparent
benzo[a]pyrene	252.3	6.13	0.0032	938	4	0.73	365	(9)	6.4	Apparent
perylene	252	6.25	0.0075	938	4	0.73	365	(9)	6.3	Apparent

perylene	252	6.25	0.0032	938	4	0.73	365	(9)	6.4	Apparent
benzo[e]pyrene	252.3	6.44	0.0075	938	4	0.73	365	(9)	6.2	Apparent
benzo[e]pyrene	252.3	6.44	0.0032	938	4	0.73	365	(9)	6.2	Apparent
benzo[ghi]perylene	276.3	6.7	0.0075	938	4	0.73	365	(9)	6.7	Apparent
benzo[ghi]perylene	276.3	6.7	0.0032	938	4	0.73	365	(9)	6.8	Apparent
indeno[1,2,3-cd]pyrene	276.3	6.7	0.0075	938	4	0.73	365	(9)	6.8	Apparent
indeno[1,2,3-cd]pyrene	276.3	6.7	0.0032	938	4	0.73	365	(9)	6.8	Apparent
dibenz[ah]anthracene	278	6.8	0.0075	938	4	0.73	365	(9)	6.7	Apparent
dibenz[ah]anthracene	278	6.8	0.0032	938	4	0.73	365	(9)	6.8	Apparent
NAPH	127.1	3.28	0.004444	1012	5.17	0.365	30	(10)	5.5	Apparent
FLU	166.2	3.84	0.004444	1012	5.17	0.365	30	(10)	8.2	Apparent
PHEN	178	4.07	0.004444	1012	5.17	0.365	30	(10)	7.8	Apparent
ANTH	178.2	4.19	0.004444	1012	5.17	0.365	30	(10)	8.2	Apparent
FLUA	202.3	4.5	0.004444	1012	5.17	0.365	30	(10)	8.7	Apparent
PYR	202.25	4.54	0.004444	1012	5.17	0.365	30	(10)	8.5	Apparent
CHRY	228.3	5.14	0.004444	1012	5.17	0.365	30	(10)	8.7	Apparent
BaA	228	5.26	0.004444	1012	5.17	0.365	30	(10)	8.7	Apparent
BaP	252.3	5.83	0.004444	1012	5.17	0.365	30	(10)	8.6	Apparent
BkF	252.3	5.91	0.004444	1012	5.17	0.365	30	(10)	8.7	Apparent

BbF	252.3	5.94	0.004444	1012	5.17	0.365	30	(10)	8.9	Apparent
I(1,2,3-cd)P	276.3	6.55	0.004444	1012	5.17	0.365	30	(10)	9.2	Apparent
di-CB	222	5.1	0.0075	938	5.17	0.365	30	(6)	7.2	Apparent
tri-CB	256	5.5	0.0075	938	5.17	0.365	30	(6)	7.4	Apparent
tetra-CB	289	5.9	0.0075	938	5.17	0.365	30	(6)	7.1	Apparent
penta-CB	323	6.3	0.0075	938	5.17	0.365	30	(6)	6.7	Apparent
Di	222	5.1	0.0075	938	5.2	0.365	30	(3)	7.06	Apparent
Tri	256	5.5	0.0075	938	5.2	0.365	30	(3)	6.92	Apparent
Tetra	289	5.9	0.0075	938	5.2	0.365	30	(3)	6.6	Apparent
Penta	323	6.3	0.0075	938	5.2	0.365	30	(3)	6.15	Apparent
Hexa	357	6.7	0.0075	938	5.2	0.365	30	(3)	5.5	Apparent
Hepta	391	7.1	0.0075	938	5.2	0.365	30	(3)	5.14	Apparent

[1] (Choi et al, 2014), [2] (Hale et al, 2010b), [3](Sun et al, 2007), [4](Zhou et al, 2022), [5](Oen et al, 2012), [6](Sun and Ghosh, 2008), [7] (Zimmerman et al., 2004),[8](Zimmerman et al., 2005), [9] (Choi et al, 2014), [10] (Hale et al, 2010a).

Abstract

Machine learning-based prediction of apparent adsorption capacity of sediment-amended activated carbon for hydrophobic organic contaminants

이현민(Lee Hyeonmin)
건설환경공학부 환경공학전공(Civil and environmental engineering)
The Graduate School
Seoul National University

Contamination of sediments by hydrophobic organic compounds (HOCs), such as polychlorinated biphenyls and polycyclic aromatic hydrocarbons, poses significant risks to human health and the environment. In-situ stabilization using activated carbon (AC) has emerged as a promising remediation

method due to its high efficiency and less impact on surrounding ecosystem. However, due to the complexity sediments possess, accurately predicting the adsorption capacity of sediment-amended AC remains challenging, with no existing prediction model available. In this study, we developed machine learning models that can accurately predict two types of adsorption capacity: the clean adsorption capacity of AC ($\log K_{AC,clean}$), which is determined in a water-AC system, and the apparent adsorption capacity ($\log K_{AC,apparent}$), which is determined in a sediment-water-AC system. The models were trained using a dataset of 186 experimental data sets which are obtained from previous studies. Through feature selection and evaluation, we identified the most informative features for predicting AC's adsorption capacity. The models, trained separately for $\log K_{AC,clean}$ and $\log K_{AC,apparent}$, performed the best. The influencing factors for $\log K_{AC,clean}$ were the specific surface area of AC, molecular weight, and $\log K_{ow}$ of the adsorbate. On the other hand, $\log K_{AC,apparent}$ utilized five input variables, which are the sediment's total organic carbon content and black carbon content, the size of the activated carbon, the molecular weight of the compound, and the duration of the experiment. Our prediction models, employing support vector regression as the selected algorithm, demonstrated excellent performance, achieving a coefficient of determination of 0.91 and a root-mean-square error of 0.38 for $\log K_{AC,clean}$ and a coefficient of determination of

0.86 and a root-mean-square error of 0.40 for $\log K_{AC,apparent}$. This study shows the potential of machine learning in accurately predicting adsorption capacity and provides valuable insights into the impact of various factors on the adsorption process. The findings contribute to the design of effective in-situ remediation strategies and enhance our understanding of the adsorption process.

Keywords : sediment contamination, activated carbon, adsorption capacity, machine learning, hydrophobic organic compounds, prediction model.

Student Number : 2021-22740