TUNNEL & UNDERGROUND SPACE Vol.32, No.6, 2022, pp.586-597
https://doi.org/10.7474/TUS.2022.32.6.586 ISSN: 1225-1275(Print) ISSN: 2287-1748(Online)

ORIGINAL ARTICLE

낙동강 조간대 연약지반의 지역별 점성토층 두께 추정 모델 개발에 관한 연구

안성인[1], 류동우[2*]

[1]한국지질자원연구원 광물자원연구본부 자원탐사개발연구센터 연구원, [2]한국지질자원연구원 광물자원연구본부 자원탐사개발연구센터 책임연구원

A Study on the Development of Model for Estimating the Thickness of Clay Layer of Soft Ground in the Nakdong River Estuary

Seongin Ahn[1] and Dong-Woo Ryu[2*]

[1]Researcher, Mineral Exploration and Mining Research Center, Korea Institute of Geoscience & Mineral Resources (KIGAM)
[2]Principal Researcher, Mineral Exploration and Mining Research Center, Korea Institute of Geoscience & Mineral Resources (KIGAM)

***Corresponding author:** dwryu@kigam.re.kr

Received: December 19, 2022
Revised: December 26, 2022
Accepted: December 26, 2022

ABSTRACT

In this study, a model was developed for the estimating the locational thickness information of the upper clay layer to be used for the consolidation vulnerability evaluation in the Nakdong river estuary. To estimate ground layer thickness information, we developed four spatial estimation models using machine learning algorithms, which are RF (Random Forest), SVR (Support Vector Regression) and GPR (Gaussian Process Regression), and geostatistical technique such as Ordinary Kriging. Among the 4,712 borehole data in the study area collected for model development, 2,948 borehole data with an upper clay layer were used, and Pearson correlation coefficient and mean squared error were used to quantitatively evaluate the performance of the developed models. In addition, for qualitative evaluation, each model was used throughout the study area to estimate the information of the upper clay layer, and the thickness distribution characteristics of it were compared with each other.

Keywords: Nakdong river estuary, Ground layer modelling, Kriging, Machine learning

초록

본 연구에서는 국내 주요 연약지반으로 알려진 낙동강 조간대 지역의 압밀침하 취약성 평가에 활용할 상부 점성토층의 위치별 두께 정보를 추정할 수 있는 모델을 개발하였다. 두께정보 추정을 위하여 기계학습 알고리즘인 RF (Random Forest), SVR (Support Vector Regression), GPR (Gaussian Process Regression)과 지구통계기법인 정규크리깅(Ordinary Kriging)을 이용한 4가지 공간추정 모델을 개발하고 상호 비교하였다. 모델 개발을 위하여 수집한 연구지역의 시추공 자료 4,712개 중 상부점성토층이 존재하는 2,948개의 시추공 자료를 사용하였으며, 개발된 모델들의 성능을 정량적으로 평가하기 위하여 피어슨(Pearson) 상관계수와 오차제곱평균(mean squared error)을 사용하였다. 또한, 정성적 평가를 위하여 연구지역 전역에 상부점성토층의 두께를 추정하여 점성토층의 지역별 분포 특성을 상호 비교하였다.

핵심어: 낙동강 조간대, 지층 모델링, 크리깅, 기계학습

© The Korean Society for Rock Mechanics and Rock Engineering 2022. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/ licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

최근 지진, 싱크홀과 같은 지반 침하, 홍수 등 도심지에서의 지질재난에 대한 문제점이 크게 대두되고 있다. 이에 각각의 재난 종에 따른 취약성 평가와 관련된 연구가 국내에서도 다수 수행되었다(Kim et al., 2022, Suh et al., 2020). 도심지 재난의 취약성 평가를 위해서는 각 재난의 원인 규명이 연구의 주요한 목적이 되겠지만 이러한 연구의 기본적인 보조자료로서 해당 지역의 지하에 대한 정보를 수집하는 것도 필수적이다. 기본적인 정보에는 지하의 물성 정보와 함께 지층정보가 포함될 수 있으며, 도심지 내의 지층정보 모델은 사회적 및 경제적 필요해 의해 아래와 같이 다양한 분야에서 적용되고 있다.

- 지하수 용량(capacity) 및 취약성(vulnerability) 평가
- 지하 인프라 및 엔지니어링 측면에서의 지반조건(ground condition) 예측
- 광범위한 재해(hazards) 분야에서의 위험도 평가
- 가시화를 통한 지질자원 교육 분야에서의 활용

도심지의 지층정보 모델링은 일반적인 지질 모델링과 다르게 상대적으로 얕은 지중 심도를 대상으로 모델링하며, 단층의 유무에 따라 다르지만 일반적으로 층의 역전 현상이 상대적으로 적기 때문에 층서적 형태로 지층정보를 모델링 할 수 있다는 장점이 있다. 또한, 최근에는 전통적 층서 모델링 및 지구통계기반 모델링과 더불어 컴퓨터 연산 자원 및 알고리즘의 발전에 의해 기계학습 및 인공지능 기반의 모델링 기법도 연구되고 있다(Table 1).

Table 1. Summary of types and characteristics for ground layer modelling

Method	Algorithms	Modelling Cost	Software	References
Layer Modelling	Implicit Potential Calculation	High	GOCAD-SKUA, Geomodeller	Calcagno et al., 2008 Wellmann and Caumon, 2018
Geostatistical Modelling	Kriging (Ordinary, Indicator)	Low	ArcGIS, SGeMS	Journel and Kyriakidis 2004 Zhou et al., 2022
Machine Learning Modelling	RandomForest, SVR	Low	Python, Matlab, R	Smirnoff et al., 2008 Hengl et al., 2018

본 연구의 목적은 도심지 지질재난 취약성 평가를 위한 기본 자료인 지층정보를 모델링 하는 것으로 연구대상 지역으로 낙동강 조간대 지역을 선정하였다. 해당 지역은 연약 점토층이 두텁고 넓게 분포하는 지역으로 과거부터 대단위 지반개량 사업 등이 수행됐음에도 시료 교란 등의 이유로 압밀 침하량이 과소 평가되어 왔다(Chung et al., 2006). 이 지역의 연약지반 침하 취약성 평가에 활용될 상부점성토층의 두께 정보를 모델링하기 위하여 부산지역 시추공 자료 중 낙동강 조간대 지역 4,712개 시추공 자료를 사용하였다. 지층정보 모델링을 위하여 RF (Random Forest), SVR (Support Vector Regression), GPR (Gaussian Process Regression) 3종의 기계학습 모델을 개발하였으며, 동시에 전통적 지구통계기법인 정규크리깅(Ordinary Kriging) 모델을 사용하여 상부점성토층의 두께를 추정하였다. 최종적으로 개발한 모델들의 공간추정 성능 비교를 위해 성능지표로 평가하는 정량 평가 방법과 공간추정 분포 형태를 주관적으로 평가하는 정성평가 방법을 사용하여 비교하였다.

2. 연구지역

연구지역은 대한민국 부산광역시 강서구 낙동강 삼각주 일대 지역을 선정하였다(Fig. 1). 낙동강 삼각주는 육상에서 173 km^2의 면적을 가지며, 지형과 수리학적 특성에 따라 상부 삼각주 평원(양산 신도시, 예안리 등), 하부 삼각주평원(녹산, 가덕도 등), 삼각주 주변분지(장유)로 구분된다. 낙동강 삼각주는 지구조 요인에 의해 거대한 만입의 형상을 유지하다가 낙동강으로부터 공급된 퇴적물들이 하천과 바다 등의 상호작용으로 서서히 매립된 지형이다(Yang and Cho, 2011). 이러한 낙동강 삼각주 일대의 미고결 퇴적층은 해수면 상승 속도에 연계되어 상·하부의 점성토층으로 구분되며, 점성토층의 상·하부에는 사질토층이 발달 된 구조를 보인다. 또한, 해당 지역은 에코델타시티, 녹산공단 등이 위치하여 주거지와 공단 등 사람의 활동이 잦은 도심지 지역이다. 최근 도시 내 싱크홀과 같은 지반 침하는 잠재적인 대형사고를 유발한다는 우려와 함께 대국민적 불안감이 커지고 있다(MOLIT, 2015). 연약점토층이 두텁고 넓게 분포하고 있는 해당 지역에는 과거부터 대단위 지반개량 사업 등이 수행되었지만, 주로 시료교란 등의 이유로 압밀 침하량을 과소평가해 왔었다(Chung et al., 2006). 그러므로 본 연구지역의 연약지층에 대한 모델링은 추후 낙동강 삼각주 연약지반의 침하 취약성을 평가하는 데 기본 자료로 활용될 수 있을 것이다.

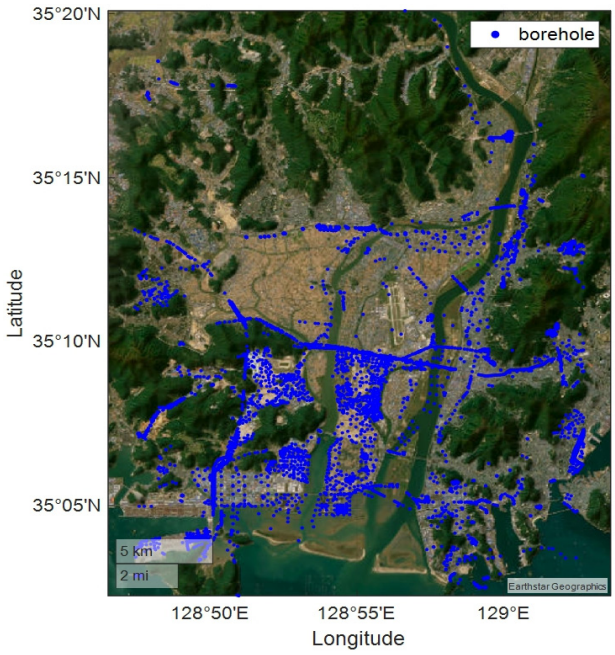


Fig. 1. Areal view of the study area with borehole data (image source: Esri)

3. 연구방법

3.1 시추공 데이터 현황 및 연약지층의 분류

본 연구에서는 연약지반 취약성 평가에 필요한 지층정보 자료로서 점성토 지층의 모델링을 위하여 낙동강 삼각주 일대의 시추공 자료를 활용하였다. 해당 자료는 국토지반정보 포털에서 제공하는 개방형 지반조사정보 데이터베이스에 포함된 부산광역시 시추

공 자료 중 중복된 자료를 제외하고 낙동강 삼각주 인근에 해당하는 4,712개의 시추공으로 구성되어 있으며, Fig. 1과 같이 분포하고 있다. 각 시추공 자료에는 시추지역, 시추공명, 위치정보(Transverse Mercator (TM), 경위도), 표고(Elevation), 지층별 심도 등의 자료가 포함되어 있다. 시추공의 지층정보는 시추주상도와 함께 낙동강 삼각주의 퇴적 특성을 고려하여 기반암 상부로부터 함력사질토, 하부점성토, 중간사질토, 상부점성토, 상부사질토, 표토(매립토)로 분류하였다. 본 연구에서 주요 모델링 대상 지층은 주된 압밀층인 점성토층이며 그 중에서도 표토와 가까운 상부점성토층을 모델링 대상으로 선정하였다. Fig. 2(a)는 상부점성토층이 낙동강 삼각주 일대에 분포하는 현황을 나타낸 것이다. 점성토층은 삼각주의 퇴적 기작에 의해 형성된 토층이기 때문에 고도가 높은 산간지역보다는 낙동강 하구를 따라 주로 분포하고 있는 것을 확인할 수 있으며, 연구지역의 시추공 자료 4,712개 중 2,948개의 시추공에 상부점성토층이 부존하였다. Fig. 2(b)는 연구지역 내 상부점성토층이 존재하는 시추공 자료에 대하여 해당 지층의 두께 정보를 히스토그램으로 나타낸 것이며 대체적으로 오른쪽 꼬리를 갖는 분포형태로 양의 왜도(Skewness)값을 나타내는 것을 확인할 수 있다.

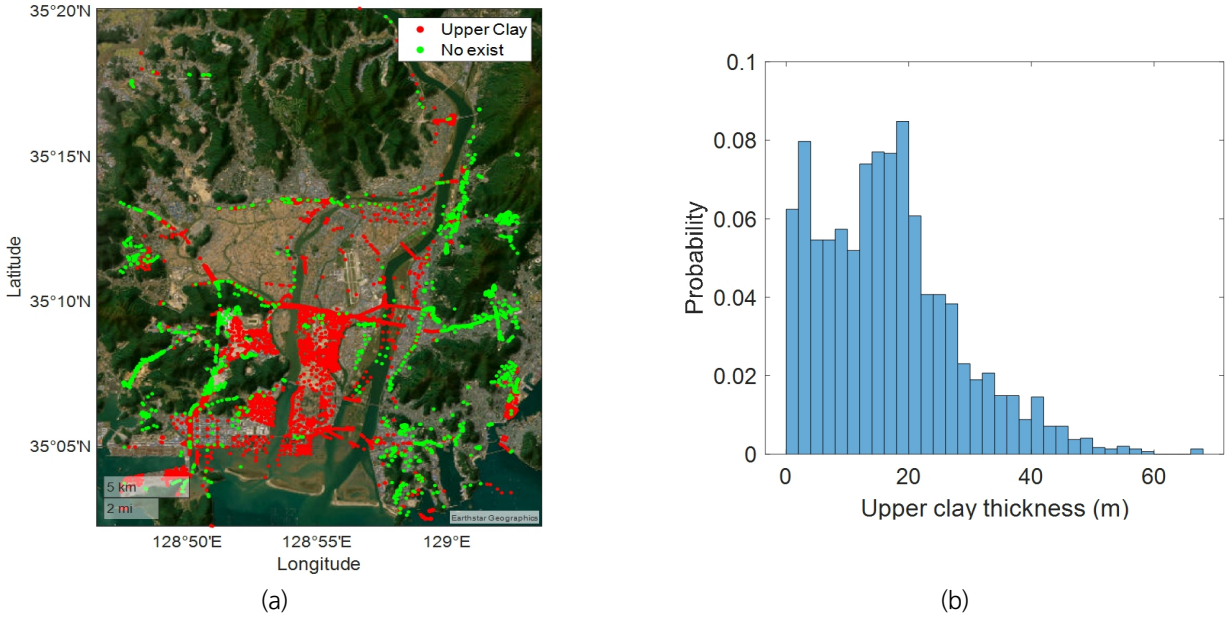


Fig. 2. Target data information. (a) Distribution of borehole with upper clay and (b) Histogram for upper clay thickness

3.2 모델 설계 및 지층정보 추론 기법

본 연구에서 개발하고자 하는 모델은 지층정보(점성토층 두께)를 추론하는 공간추정 문제이기 때문에 입력변수로는 공간정보가(TM X좌표, TM Y좌표, 표고) 구성될 수 있으며 출력변수로는 상부점성토층의 두께로 구성하였다. 이때 Fig. 2(a)와 같이 상부점성토층은 산간지역에 대부분 형성되어 있지 않은 것으로 확인되며, Fig. 3과 같이 상부점성토층의 두께와 표고의 관계를 도시해보았을 때 일부 산간지역에 분포하는 상부점성토층(해당 경우에도 매우 작은 두께의 점성토층만이 존재)을 제외했을 때 고도와 점성토층의 두께가 특정한 추세를 가지지 않는다는 것을 확인할 수 있다. 이러한 이유로 본 연구에서는 점성토층의 두께 추정을 위한 입력변수로 표고 자료의 사용 없이 TM좌표계만을 사용하여 점성토층의 모델링을 수행하였다.

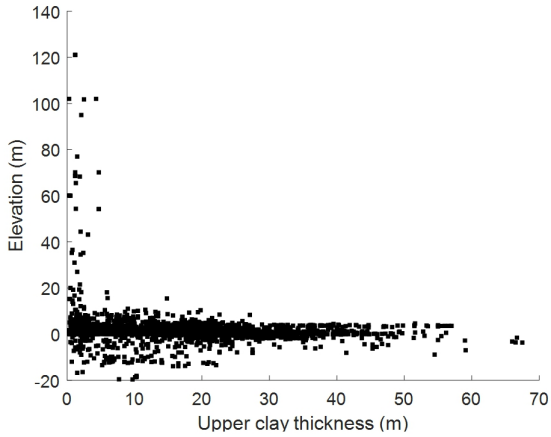


Fig. 3. Correlation analysis between target data and each elevation

또한, 개발한 모델의 성능을 시험하기 위하여 전체 시추공 데이터를 5개의 군집으로 나누어 전체의 80%(2,358개)에 해당하는 4개의 군집은 학습하고 나머지 20%(590개)인 1개의 군집을 시험하였다. 시험 군집을 5번 번갈아가며 5-fold 교차 시험을 수행하였다. 이러한 시험 방식을 이용하면 각 모델의 전체 데이터셋에 대한 추정 성능을 평가할 수 있다는 장점이 있다. 이 때, 지층정보에 대한 공간추론의 경우 지층의 연속적인 성질이 반영되어야 하므로 Fig. 4(a)와 같이 극단적인 형태이긴 하지만 지역 편향적으로 각각의 Fold를 구성하는 것이 아니라 Fig. 4(b)와 같이 각 Fold가 추정범위 전 범위를 커버할 수 있도록 구성되어야 한다.

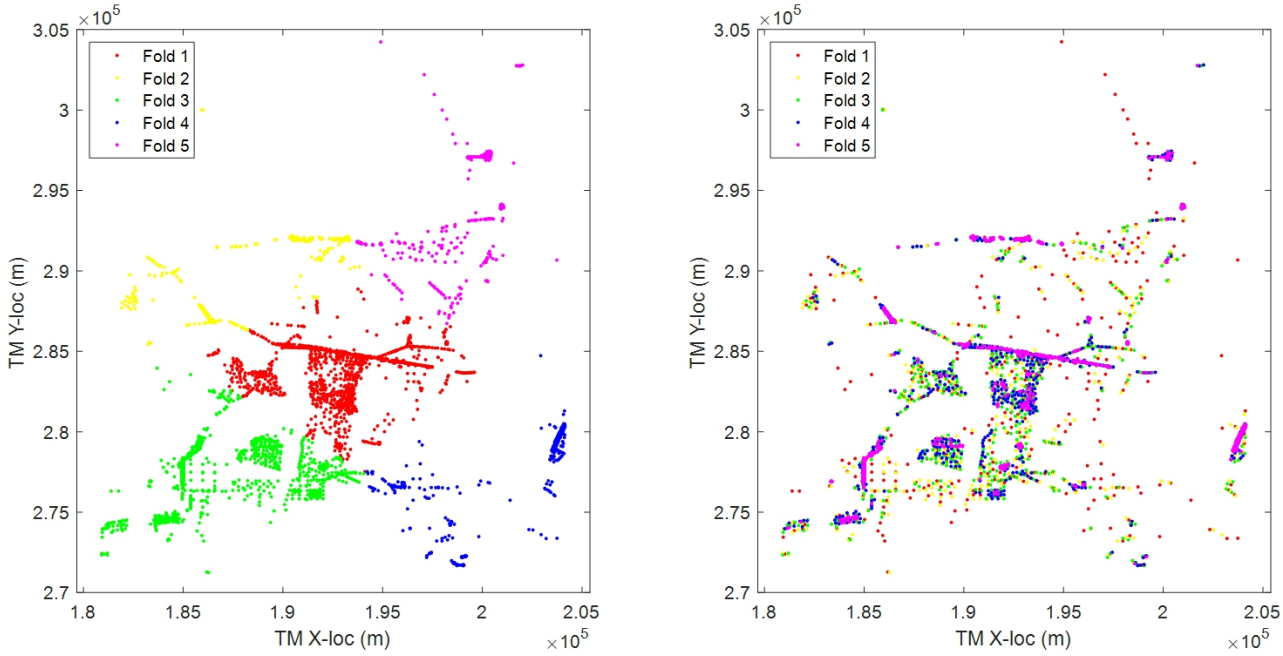


Fig. 4. Results of dividing sample data into five folds (a) without considering regional bias and (b) with considering it

3.2.1 지구통계기법: 정규크리깅(Ordinary Kriging)

크리깅(Kriging)은 미지의 관심 지점에서의 속성값을 예측하기 위하여 주변의 획득된 샘플 자료를 기반으로 정의된 확률변수의 결합을 통해 계산되는 방법론이다(Isaaks and Strivastava, 1989, Goovaerts, 1997). 이 때, 위치 기반의 공간적 상관성을 가지는 확률변수의 결합을 통해 생성되는 확률함수는 정상성(stationarity)을 만족하는 범위 내에서 일반적으로 식 (1)과 같은 변동도 모델로서 크리깅 계산에 반영된다.

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[z(x_i) - z(x_i + h)]^2 \tag{1}$$

여기서 $\gamma(h)$는 반-변동도(semi-variogram)이고, $z(x)$, $z(x+h)$는 각각, x, $x+h$ 위치에서의 속성값을 의미하며, h는 지연거리(lag distance)로서 샘플 속성 값들 사이의 거리를 의미한다.

정규크리깅(Ordinary Kriging, OK)은 편향된 추정식의 사용으로 인해 추정식의 평균이 모집단의 평균과 일치하지 않는 단순크리깅의 문제점을 극복하기 위해 추정식이 편향되지 않으면서 오차분산을 최소로 하는 크리깅 기법이다. 이를 위해 정규크리깅에서는 추정치가 편향되지 않도록 식 (2)와 같이 가중치의 합이 1이 되는 제약조건을 가진다. 주어진 제약조건에서 최대 및 최소를 구하는 문제를 푸는 방법으로 라그랑주 승수법(Lagrange multiplier method)이 있으며, 이를 이용하여 정규크리깅의 오차분산식을 최소화할 수 있도록 수학적으로 표현하면 식 (3)과 같이 나타낼 수 있다.

$$1 - \sum_{i=1}^{n}\lambda_i = 0 \tag{2}$$

$$L(\lambda_1, \lambda_2, \ldots, \lambda_n; \omega) = \sigma^2 - 2\sum_{i=1}^{n}\lambda_i\sigma_{0i}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\sigma_{ij}^2 + 2\omega(1 - \sum_{i=1}^{n}\lambda_i) \tag{3}$$

3.2.2 기계학습기반 회귀 알고리즘

본 연구에서는 시추공 자료를 활용하여 상부점성토층의 두께를 추정하기 위해 전통적 지구통계기법 이외에도 3종류의 기계학습 알고리즘을 최적화하여 학습시킨 후 모델들을 개발하였다. 3종류의 기계학습 알고리즘은 RF, SVR, GPR이며, 정규크리깅을 활용하여 개발한 4개 모델의 공간추정 성능을 비교하였다.

- Random Forest (RF)는 다수의 결정트리(decision tree)들을 학습하여 문제를 해결하는 기법(Breiman, 2001, Ho, 1998)으로 대표적인 앙상블 기법이다. 결정트리 방식은 학습 데이터에 따라 예측 성능이 크게 달라지며, 과적합하는 경향이 있다. 이러한 문제점을 극복하기 위해 다수의 결정트리를 고려하여 데이터를 학습시키는 앙상블 기법으로 배깅(Bagging)이 있다. 배깅은 각 결정트리를 독립적으로 학습하는 병렬 앙상블 모델로서 분산을 줄이고, 예측 모델의 과적합을 방지하는 특징을 가지고 있다. 배깅을 통해 예측된 값은 개별 트리 예측값의 평균값으로 표현될 수 있다. RF와 배깅은 거의 동일한 방식으로 계산되지만 한 가지 큰 차이점은 트리 내 노드의 분할 분기점에서 학습 데이터의 특징(feature)들을 무작위로 선택한다는 점이 있다.
- Support Vector Machine (SVM)은 대표적인 기계학습 분류 및 회귀 알고리즘의 하나로 회귀 모형에 사용될 때에는 SVR

(Support Vector Regression)으로 불린다(Cortes and Vapnik, 1995). SVM은 데이터를 분류하는 결정 경계(decision boundary)를 찾기 위해 가장 가까운 서포터 벡터(support vector)를 찾아 결정 경계에 평행한 초평면(hyperplane)을 계산하고 초평면과 결정 경계의 거리(margin)를 최대화하는 방향으로 최적해를 찾아간다. SVM은 이러한 거리를 계산하기 위해 전체 학습 자료가 아닌 서포트 벡터들만 필요하기 때문에 상대적으로 연산이 빠르다는 장점이 있다. 또한, 자료가 선형 분리기로 분류가 힘들 경우, 비모수적(nonparametric) 접근으로 커널 함수를 이용하여 고차원상(hibert space)에 자료를 투영하지 않고도 기존 차원의 자료를 구분하기 위한 초평면을 찾을 수 있다. 이때, 커널함수에 대한 매개변수로 C와 gamma값을 조정할 수 있다(Alpaydin, 2020).

- Gaussian Process (GP)는 비모수적 기계학습 모델로서 SVM과 마찬가지로 커널함수 기반으로 모수적(parametric) 모델이 복잡한 학습자료에서는 낮은 성능을 보인다는 단점을 개선하기 위해 등장하였다. GP는 식 (4)와 같이 결합 가우시안 분포(jointly gaussian distribution)를 따르는 임의 변수(random variable) 군을 의미한다. 일반적인 가우시안 분포와 가우시안 프로세스의 차이점은 평균과 분산이 식 (4)의 평균 함수인 $m(x)$와 공분산 함수인 $k(x, x')$로 표현된다는 것이다(Williams and Rasmussen, 2006).

$$f \sim GP(m(x), k(x, x')) \tag{4}$$

수학적으로 GP는 크리깅의 일반화로 간주할 수 있지만 크리깅은 일반적으로 공간추정 문제에 사용되는 반면 GP는 기계학습 분야에서 주로 사용된다. 또한 GP는 베이지안(Bayesian) 관점에서 파생되었지만 크리깅은 빈도주의(frequentist) 관점에서 파생되었다는 차이점이 있다(Erickson et al., 2018, Cui et al., 2021). 이러한 GP는 회귀 및 분류 문제 모두 활용될 수 있으며 회귀문제에 사용될 때 GPR (Gaussian Process Regression)로 불린다.

4. 연구결과 및 토의

4.1 상부점성토층 두께 추정 기법의 성능 비교

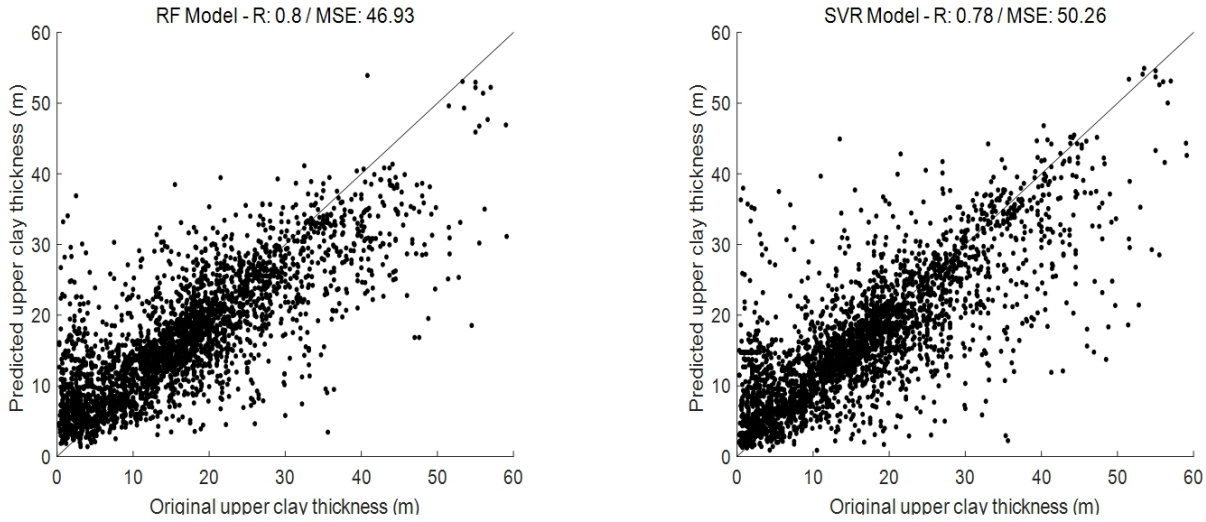
본 연구에서는 3종류의 기계학습 알고리즘(RF, SVR, GPR)과 지구통계기법인 정규크리깅을 사용하여 상부점성토층의 위치별 두께를 추정하였다. 공간추정을 수행하기 위한 기계학습 알고리즘은 MATLAB의 Machine learning and Statistics Toolbox (MATLAB, 2022)를 사용하였으며, 정규크리깅을 수행하기 위해서 Python의 GStools (Müller et al., 2022) 라이브러리를 사용하였다. 각 기계학습 알고리즘 모델의 하이퍼 파라미터(hyperparameter)를 탐색하여 최적의 모델을 개발하기 위해서 격자 탐색(grid search), 임의 탐색(random search) 등을 수행할 수 있지만 본 연구에서는 베이지안 최적화(Bayesian optimization) 탐색 알고리즘을 사용하여 모델을 최적화하였다. 베이지안 최적화(Malkomes et al., 2016)는 불필요한 하이퍼 파라미터의 반복 탐색을 줄여 보다 빠르게 파라미터를 찾을 수 있는 기법으로 알려지지 않은 목적 함수의 최대 또는 최소로 하는 해를 찾아가는 장점이 있다. 이에 RF모델은 Tree수와 최소 잎 크기, SVR모델은 커널함수와 C값과 gamma값, GPR 모델은 커널함수와 gamma값을 변화시켜가며 최적화하였다. 기계학습 모델의 최적화 결과 RF모델은 Tree수가 500, 최소 잎 크기 = 1일 때 가장 좋은 결과가 나왔으며, SVR은 커널함수로 가우스 함수를 이용하고 C = 21.7067, gamma = 0.0943일 때, GPR은 등방성 지수함수를 사용하고 gamma = 100일

때 가장 좋은 훈련 결과를 보였다.

최적화를 수행하여 결정한 하이퍼 파라미터로 설정된 각 기계학습 알고리즘과 정규크리깅 모델의 5-fold 교차 시험을 수행한 결과를 요약한 것이 Table 1과 Fig. 5에 나타나 있다. 각 모델의 성능을 비교하기 위하여 성능지표(피어슨 상관계수(Pearson correlation coefficient), 오차제곱평균(mean squared error))뿐만 아니라 추정한 상부점성토층의 기본 통계량(평균, 표준편차, 최댓값, 최솟값, 중앙값)을 참값과 비교하여 나타내었다. 성능지표를 비교해 보았을 때, 본 연구지역에서는 기계학습 알고리즘을 사용한 경우의 공간추정 성능이 정규크리깅을 사용한 공간추정 결과보다 대체적으로 높은 것을 확인할 수 있었다. 또한, Fig. 6과 같이 기계학습 알고리즘 중 상대적으로 성능이 가장 높았던 GPR과 정규크리깅을 사용하여 추정된 자료와 원본 자료의 분포를 비교해 보았을 때도 GPR 모델의 결과가 정규크리깅 모델에 대비하여 상대적으로 더 잘 모사한 것을 확인할 수 있다. 이러한 결과는 Table 2에 나타난 기본 통계량에서도 기계학습 알고리즘 모델들이 상대적으로 평균과 표준편차를 원본 자료와 근사하게 추정한 것으로 확인할 수 있다. 반면에 정규크리깅 모델을 통해 추정된 결과에서는 상대적으로 평균값과 표준편차가 더 작게 평가되었다. 이는 양의 왜도를 가지는 원본 자료의 분포 특성을 고려하여 로그정규화(log-normalization)를 수행하였음에도 불구하고 점성토층이 두꺼운 지역의 자료가 이상치(outlier)로 간주되어 계산된 결과로 추정할 수 있다. 한편 기계학습 알고리즘 중에서도 SVR의 경우 최솟값이 음의 값을 가지면서 원본자료의 자연현상적인 특성을 반영하지 못한 결과가 나온 것을 확인할 수 있다.

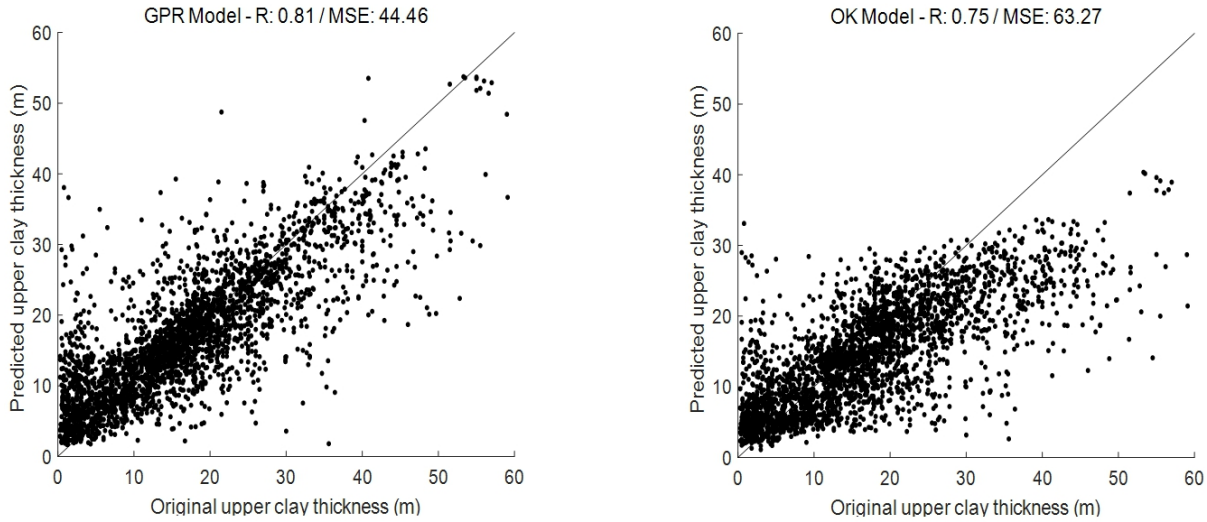
Table 2. Summary of the results for prediction performance of each methodology with basic statistics

Method	Metric		Basic Statistics					
	Corr. coeff.	MSE	Mean	Std.	Max	Median	Min	Skewness
RF	0.80	46.93	16.74	8.73	53.91	15.62	1.40	0.65
SVR	0.78	50.26	16.59	9.30	54.93	15.39	-1.20	0.79
GPR	0.81	44.46	16.78	9.03	53.74	15.51	1.69	0.77
OK	0.75	63.27	14.31	7.45	40.35	13.92	1.12	0.36
Ground Truth			16.68	11.30	67.50	15.60	0.30	0.86



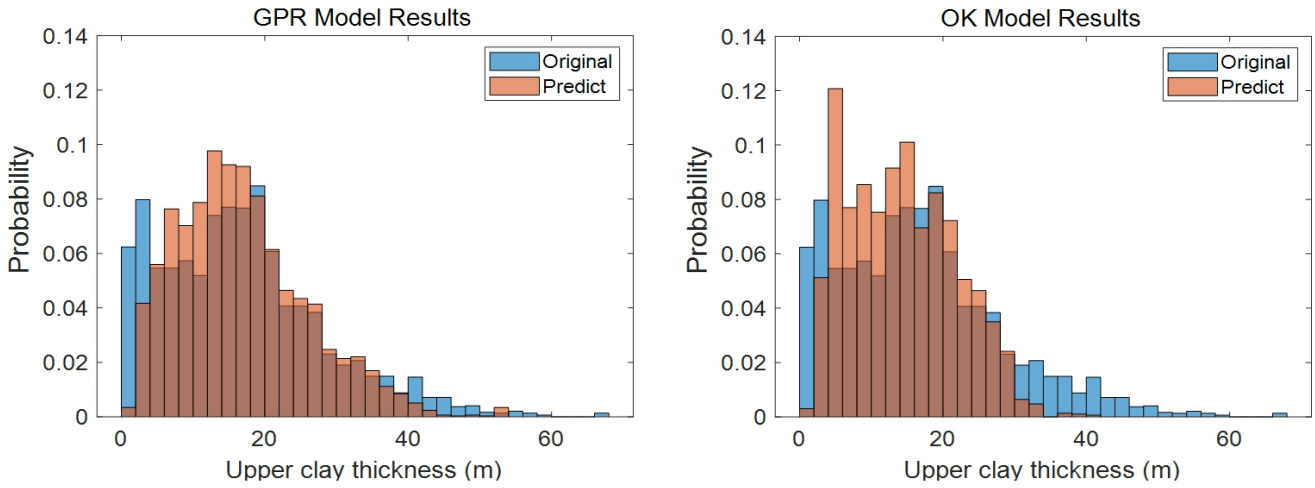
(a) (b)

Fig. 5. Results of correlation analysis between the predicted and ground truth. (a) RF, (b) SVR, (c) GPR, (d) OK



(c) (d)

Fig. 5. Results of correlation analysis between the predicted and ground truth. (a) RF, (b) SVR, (c) GPR, (d) OK (Continued)



(a) (b)

Fig. 6. Comparison results of histogram between predicted data of each methodology and ground truth. (a) GPR, (b) OK

4.2 지층정보 추정 모델별 가시화 결과 비교

각 공간추정 모델들의 정량적인 성능 평가와 함께 정성적인 평가를 위하여 낙동강 조간대 영역에 학습된 모델들을 사용하여 상부점성토층의 두께를 추정한 결과를 Fig. 7과 같이 가시화하여 비교하였다. 낙동강 조간대 영역 전체에 대한 추정에 앞서 연구지역 내에서도 상부점성토층이 존재할 것으로 추정되는 지역에 대해서만 상부점성토층의 두께를 추정하는 것이 결과의 신뢰성을 제공할 수 있을 것으로 판단하였다. 이에 Fig. 7(a)와 같이 연구지역 중 상부점성토층이 존재할 것으로 추정되는 지역을 선정하기 위하여 기계학습 알고리즘인 RF를 이용하였다. RF를 포함하여 다른 기계학습 분류 알고리즘(GP, SVM 등)을 사용한 결과 RF의 5-fold 시험 결과가 약 92 %의 정확도로 가장 높은 성능을 보였다. 상부점성토층이 존재할 것으로 추정되는 지역에 해당 지층의 정보를 각 모델들로 추정한 결과 4개의 모델 모두 상대적으로 지층이 두꺼운 부분을 포함하여 분포 형상은 유사하게 추정한 것으로 판단되었

다. 하지만 RF 모델의 경우 선형 분리기의 특성상 직선 형태의 인위적인 분포가 나타났으며, SVR의 경우 지역적 특잇값들 주위를 좁은 원형의 형태로 추정한 것이 확인되었다. 반면에 정규크리깅 모델의 경우 지역적 분포 특성을 정성적으로 확인할 수 있도록 추정되었으며, GPR은 SVR과 정규크리깅 모델의 공간추정 특징이 복합적으로 나타난 것으로 정성평가 될 수 있다.

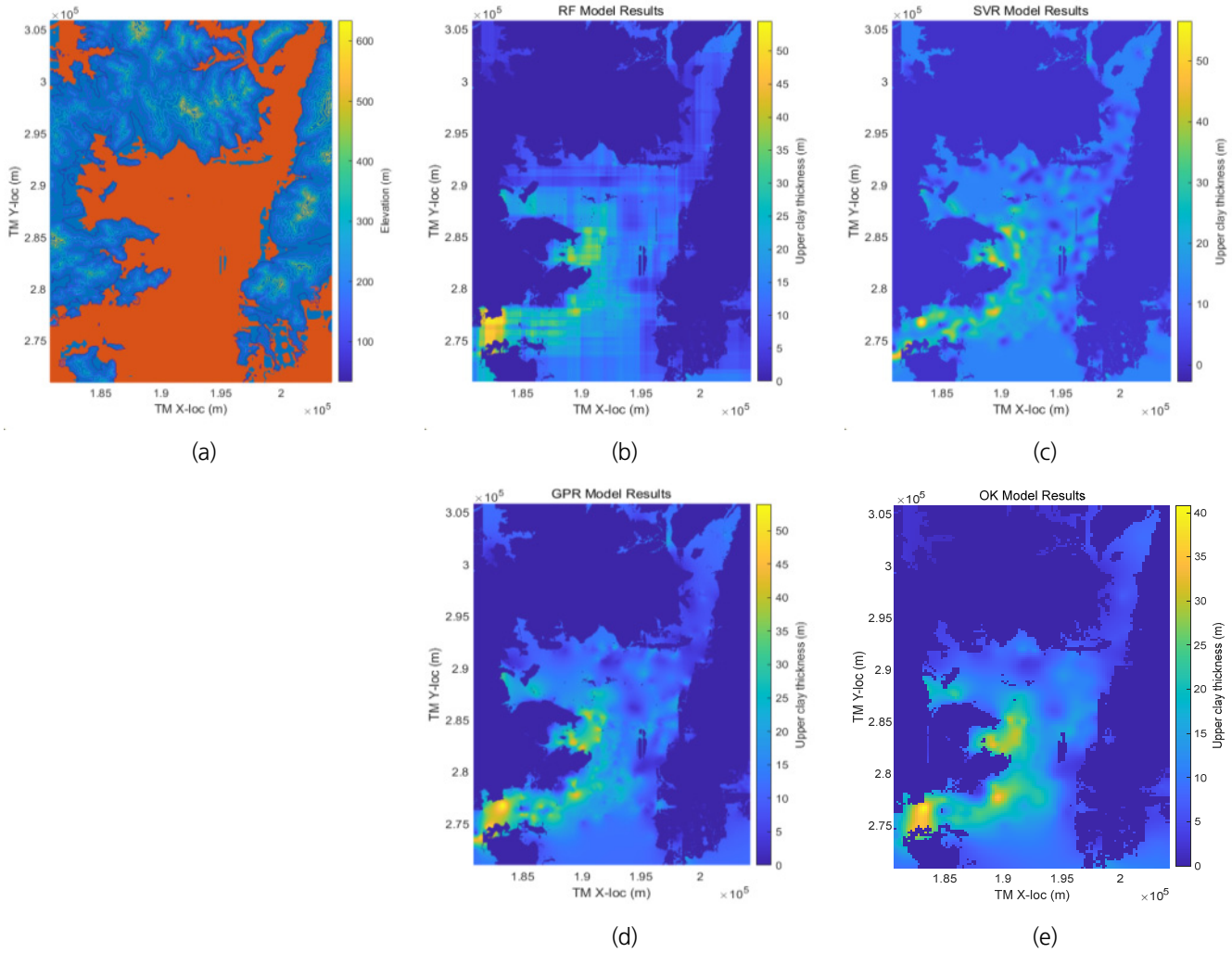


Fig. 7. Result of (a) masking the area where target data is expected to exist and comparison of estimation based on each methodology for predicting upper clay thickness of the study area. (b) RF, (c) SVR, (d) GPR, (e) OK

5. 결 론

본 연구에서는 국토지반정보포털에서 수집한 시추공 자료를 기반으로 낙동강 조간대 지역의 상부점성토층 두께 정보에 대한 공간추정을 수행하였다. 상부점성토층의 위치별 두께 정보를 예측하기 위해 3종의 기계학습 모델과 지구통계기반 공간추정 기법인 정규크리깅을 사용하여 총 4개의 모델을 개발하였으며, 그 성능을 비교하였다. 낙동강 조간대 지역 4,712개의 시추공 자료 중 상부점성토층이 존재하는 2,948개의 시추공 자료를 사용하여 지역적 편향을 가지지 않도록 5-fold 자료 군으로 구성하여 성능을 검증

한 결과, 기계학습 모델인 RF, SVR, GPR 모델과 정규크리깅 모델 중 GPR 모델의 성능이 R=0.81, MSE=44.46 (m)로 가장 좋은 성능을 보였다. 정량적인 공간추정 성능 평가와 함께 각 모델의 정성적 평가를 위하여 낙동강 조간대 전역에 상부점성토층의 정보를 추정해보았다.

그 결과, 지층의 공간적 분포 특성이 현실적으로 반영된 정규크리깅 모델과 달리 기계학습 모델들의 경우 인위적인 분포형태나 특잇값에 대해 국소적으로 분포하는 형태로 추정되었다. 이러한 결과는 점성토층의 두께를 비롯한 지층정보 추정 문제에 기계학습 알고리즘을 적용하는 것은 추정 성능 측면에서 정량적으로 우수할 수 있지만 자연현상적인 특성까지 잘 반영되는지에 대해 정성적으로 평가하는 것이 항상 동반되어야 한다는 것을 알려 주는 결과라고 판단된다. 그러나 지층정보 추정 문제에서 기계학습 알고리즘을 사용하는 것은 지구통계기법인 정규크리깅이 지역적 극댓값과 같은 특잇값의 추정에 취약하다는 단점을 보완할 수 있으며, 공간추정을 위한 보조자료의 추가와 같이 공간 모델링의 확장성 측면에 있어서도 장점이 있다. 그러므로 지구통계기법과 함께 기계학습 알고리즘을 통한 공간정보 모델링에 대한 연구는 서로 상호보완적으로 수행되어야 할 것이다.

마지막으로 본 연구에서 수행한 점성토층의 모델링 결과는 추후 낙동강 삼각주 연약지반의 침하 취약성을 평가하는데 기본 자료로 활용될 수 있을 것이다.

사 사

본 연구는 한국지질자원연구원 기본사업 '도시복합지질재난 능동 대응 스마트 통합 솔루션 기술 개발' 과제(GP2021-007)의 일환으로 수행되었습니다. 또한, 시추공 정보를 제공해 준 국토지반정보 통합DB센터에 감사드립니다.

REFERENCES

Alpaydin, E., 2020, Introduction to machine learning. MIT press.

Breiman, L., 2001, Random forests. Machine Learning. 45, 5-32.

Calcagno, P., Chilès, J. P., Courrioux, G., and Guillen, A., 2008. Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules. Physics of the Earth and Planetary Interiors, 171(1-4), 147-157.

Chung, S.G., Jang, W.Y., Ninjgarav, E., and Ryu, C.K., 2006, Compissibility Characteristics Associated with Depositional Environment of Pusan Clay in the Nakdong River Estuary. Journal of the Korean Geotechnical Society, 22(12), 57-65.

Cortes, C., and Vapnik, V., 1995, Support-vector networks. Machine learning, 20(3), 273-297.

Cui, T., Pagendam, D.E., and Gilfedder, M., 2021, Gaussian process machine learning and Kriging for groundwater salinity interpolation. Environmental Modelling & Software, 144, 105170.

Erickson, C.B., Ankenman, B.E., and Sanchez, S.M., 2018, Comparison of Gaussian process modeling software. European Journal of Operational Research, 266(1), 179-192.

Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation. Oxford University Press.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., and Graler, B., 2018, Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6.

Ho, T.K., 1998, The random subspace method for constructing decision forests. IEEE TPAMI, 20, 832-844.

Isaaks, E.H. and Srivastava, R.M., 1989, An Introduction to Applied Geostatistics. Oxford University Press.

Journel, A.G. and Kyriakidis, P.C., 2004, Evaluation of mineral reserves: a simulation approach. Oxford University Press.

Kim, H.R., Yu, S.H., Yun, S.T., Kim, K.H., Lee, G.T., Lee, J.H., Heo, C.H., and Ryu, D.W., 2022, Estimation of Spatial Distribution Using the Gaussian Mixture Model with Multivariate Geoscience Data. Economic and Environmental Geology, 55(4), 353-366.

Malkomes, G., Schaff, C., and Garnett, R., 2016, Bayesian optimization for automated model selection. Advances in Neural Information Processing Systems, 29.

MATLAB. 2022. 9.12.0.1884302(R2022a). Natick, Massachusetts: The MathWorks Inc.

MOLIT (Ministry of Land, Infrastructure and Transport), 2015, Guidance manual on subsidence control, 67p.

Müller, S., Schüler, L., Zech, A., and Heße, F., 2022, GSTools v1.3: a toolbox for geostatistical modelling in Python. Geoscientific Model Development, 15, 3161-3182.

Smirnoff, A., Boisvert, E., and Paradis, S. J., 2008, Support vector machine for 3D modelling from sparse geological information of various origins. Computers and Geosciences, 34(2), 127-143.

Suh, J.W., Ryu, D.W., and Yum, B.W., 2020, Logistic Regression and GIS based Urban Ground Sink Susceptibility Assessment Considering Soil Particle Loss. Tunnel and Underground Space, 30(2), 149-163.

Wellmann, F. and Caumon, G., 2018, 3-D Structural geological models: Concepts, methods, and uncertainties. Advances in Geophysics, 59, 1-121.

Williams, C.K. and Rasmussen, C.E., 2006, Gaussian processes for machine learning. MIT press.

Yang, J.H. and Cho, K.R. 2011., Geomorphological Development of Embayment Area at the estuary of Nakdong River. Journal of The Korean Association of Regional Geographers, 17(6), 649-665.

Zhou, F., Li, M., Huang, C., Liang, H., Liu, Y., Zhang, J., Wang, B., and Hao, M., 2022, Lithology-Based 3D Modeling of Urban Geological Attributes and Their Engineering Application: A Case Study of Guang'an City, SW China. Frontiers Earth Science, 10, 918285.