

국내 굴착 암반 부지에 대한 심부 전단파속도 주상도 예측 모델 개발

Predictive Models of Deep Shear Wave Velocity Profiles at Excavated Rock Sites in Korea

김지은¹⁾ · 김병민²⁾ · 조영규^{3)*}

Kim, Jieun¹⁾ · Kim, Byungmin²⁾ · Cho, Youngkyu^{3)*}

¹⁾울산과학기술원 지구환경도시건설공학과 대학원생, ²⁾울산과학기술원 지구환경도시건설공학과 부교수, ³⁾울산과학기술원 지구환경도시건설공학과 연구교수

¹⁾Graduate Research Assistant, Department of Civil Urban Earth and Environmental Engineering, Ulsan National Institute of Science and Technology,

²⁾Associate Professor, Department of Civil Urban Earth and Environmental Engineering, Ulsan National Institute of Science and Technology,

³⁾Research Professor, Department of Civil Urban Earth and Environmental Engineering, Ulsan National Institute of Science and Technology

/ A B S T R A C T /

In the context of site response analysis, the use of shear wave velocity (V_s) profiles that consider the seismological rock ($V_s \geq 3,000$ m/s) depth is recommended. This study proposes regression analysis and machine learning-based models to predict deep V_s profiles for a specialized excavated rock site in South Korea. The regression model was developed by modifying mathematical expressions from a previous study and analyzing the correlation between V_{S50} and model variables to predict deep V_s beyond 50 m. The machine learning models, designed using tree-based algorithms and a fully connected hierarchical structure, were developed to predict V_s from 51 m to 300 m at 1 m intervals. These models were validated by comparing them with measured deep V_s profiles and accurately estimating the trend of deep V_s variations. The proposed prediction models are expected to improve the accuracy of ground motion predictions for a specialized excavated rock site in Korea.

Key words: Deep shear wave velocity profiles, Excavated rock sites in Korea, Seismological rock depth, Regression analysis, Machine learning

1. 서론

국내에서는 2016년 경주 지진(리히터 규모(M_L) 5.8), 2017년 포항 지진(M_L 5.4), 그리고 2024년 부산 지진(M_L 4.8)과 같은 중규모 지진의 발생이 증가하는 추세다. 국내의 지진 발생 환경 변화에 따라 지진에 대한 구조물의 지진 안전성 확보 요구가 증대되고 있으며, 특히 주요 시설에 대한 안전성 확보가 중요 과제로 대두되고 있다.

실질적인 지진 안전성 확보를 위해서는 부지 특성이 반영된 부지 고유 응답 스펙트럼이 요구된다. 부지 고유 응답 스펙트럼은 대상 부지 지반의 고유한 물성이 반영된 응답 스펙트럼으로, 지진 발생 시 해당 부지에서 나타날 수

있는 지진동의 크기 및 주파수 특성을 산정하여 구조물에 예상되는 하중을 평가하는 기준이 된다. 진원에서 발생한 탄성파가 지층으로 전파될 때, 각 지층의 서로 다른 물성에 의한 임피던스 차이로 반사, 굴절 및 산란 등이 발생한다. 따라서 지진파는 각 지층을 통과할 때마다 감폭 또는 증폭되며, 이로 인해 임피던스(지반의 밀도(ρ)와 전단파속도(V_s)의 곱) 차이를 일으키는 지반의 물성(ρ , V_s)은 지표면에서의 지진동 특성을 결정짓는 주요 변수이다. V_s 를 지반의 깊이에 따라 표현한 V_s 주상도는 공간적으로 큰 변동성이 나타난다. 즉, 같은 부지에서도 평가하는 지점에 따라 상당한 차이를 보이며, 이러한 특징은 지표면 지진동 예측에 불확실성을 야기한다. 따라서 V_s 주상도의 변동성을 정량화한 부지 응답 해석이 필수적이다.

국내 내진설계 기준에서는 공학적 기반암(engineering bedrock)을 $V_s \geq 760$ m/s 인 지층으로 정의한다[1]. 그러나 주요 시설 건설을 위해 토층을 굴착한 암반 부지(굴착 암반 부지, Fig. 1)에 대해서는 이보다 높은 V_s 를 고려

*Corresponding author: Cho, Youngkyu

E-mail: youngkyu@unist.ac.kr

(Received November 16, 2024; Revised December 8, 2024; Accepted December 9, 2024)

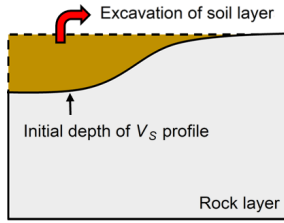


Fig. 1. The illustration of the excavated rock at the study site.

할 필요가 있다. 예를 들어, 국외의 연구 결과는 부지 증폭 계수 산정 시, 공학적 기반암이 아닌 지진학적 암반(seismological rock, $V_S \geq 3,000$ m/s) 깊이를 고려할 것을 권장한다[2-4]. 이는 진원으로부터의 파동이 부지 증폭 없이 도달하는 유일한 층이 지진학적 암반이기 때문이다[3]. 따라서 기존보다 정확한 부지 고유 응답 스펙트럼을 평가하기 위해서는 지진학적 암반을 고려한 심부 V_S 주상도(최종 평가 깊이 ≥ 100 m) 확보가 요구된다. 하지만 국내의 굴착 암반 부지에서 다량의 심부 V_S 주상도를 평가하는 것은 비경제적이고, 100 m 이상 깊이에서의 V_S 를 예측할 수 있는 모델 또한 부재하다. 따라서 기존에 평가된 천부 V_S 주상도(최종 평가 깊이 = 50 - 100 m)를 바탕으로 심부 V_S 를 예측할 수 있는 모델의 개발이 필요하다.

V_S 주상도 추정과 관련하여 회귀분석 기법을 활용한 연구는 국내 및 국외에서 활발히 진행되고 있다. Sun et al.[5]은 국내 부지에서 수집한 V_S 주상도의 평균을 기반으로 형상 곡선을 유도하여 얇은 깊이부터 30 m까지 V_S 주상도를 외삽하는 방법을 제안하였다. Shi and Asimaki[6]는 미국 캘리포니아 지역의 퇴적 분지에 대해 깊이 30 m까지의 평균 V_S (V_{S30})를 단일 입력변수로 취하여 V_S 주상도를 예측하는 수학적 모델을 제안했다. Passeri et al.[7]은 이탈리아 부지에서 공간적 변동성을 고려한 V_S 예측 모델을 개발하였다. 이러한 모델은 다수의 V_S 주상도를 평가하기 위하여 소요되는 비용 및 시간에 대한 보완책으로 평가받고 있다.

본 연구에서는 Shi and Asimaki[6]의 방법을 참고하여 국내 굴착 암반 부지의 심부 V_S 예측을 위한 회귀분석 모델을 제안한다. 굴착 암반 부지에서 기존에 평가되었던 179개의 천부 V_S 주상도를 수집하였으며, 현장 실험을 통하여 64개의 심부 V_S 주상도를 추가 획득하였다. 굴착 암반 부지에서 수집한 V_S 주상도의 특징을 고려해 50 m까지의 평균 V_S (V_{S50})을 산정하여 깊이 50 m 하부의 V_S 를 예측할 수 있는 수학적식을 작성하였다. V_{S50} 값에 따라 분류된 V_S 주상도의 평균을 계산하여 작성한 수학적식에 적합시켜 모델변수 값을 유도하였으며, 모델로 예측한 V_S 주상도를 실제 계측된 V_S 주상도와 비교하였다. 또한 본 연구에서는 머신러닝 기법을 활용한 V_S 예측 모델도 제안한다. 머신러닝 기법은 특정한 수학적식을 정의하지 않고도 입력변수와 출력변수 간의 비선형적 상관관계를 효과적으로 포착할 수 있다는 이점이 있으며, 이를 활용한 V_S 주상도 예측 모델 연구도 꾸준히 진행되고 있다[8-10]. 모델을 구축하기 위해 심부 깊이까지 평가된 64개의 V_S 주상도를 활용하였으며, 수목 기반 알고리즘 및 완전 연결 계층의 구조를 갖는 기법을 채택하였다. 개발된 머신러닝 모델은 깊이 50 m까지의 V_S 주상도를 입력값으로 받아 1 m 간격으로 깊이 300 m까지의 V_S 를 출력한다.

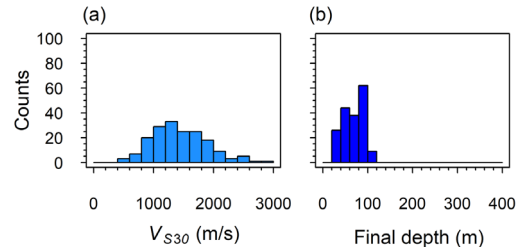


Fig. 2. Histograms of (a) V_{S30} and (b) final depths for the 179 shallow V_S profiles obtained from invasive method.

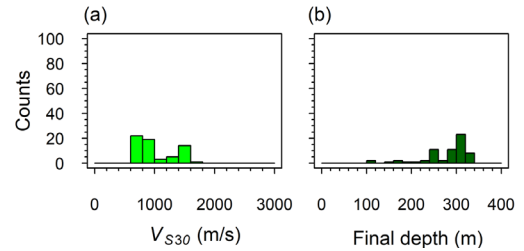


Fig. 3. Histograms of (a) V_{S30} and (b) final depths for the 64 deep V_S profiles obtained from non-invasive method

2. 자료 수집

기존에 굴착 암반 부지에서 평가된 179개의 V_S 주상도를 수집하였다. 해당 V_S 주상도는 공내 수평 탄성파 탐사(cross-hole test) 및 공내 하향 탄성파 탐사(down-hole test)로 평가되었다. Fig. 2는 수집한 V_S 주상도의 V_{S30} 및 최종 평가 깊이에 대한 히스토그램을 보여준다(보안상 위치 정보는 공개할 수 없다). V_{S30} 은 대부분 800 m/s에서 2,000 m/s 사이에 있으며(Fig. 2(a)), 이는 해당 부지가 보통암(760 m/s $\leq V_{S30} \leq 1,500$ m/s) 혹은 경암($V_{S30} \geq 1,500$ m/s) 지반임을 시사한다. 수집한 V_S 주상도 중 최종 평가 깊이가 100 m를 넘는 경우는 10개로 드물며(Fig. 2(b)), 이는 1절에 설명된 기존 부지 응답 해석의 기초에 의한 것으로 판단된다.

기존의 최종 탐사 깊이보다 더 깊은 깊이의 V_S 주상도를 획득하기 위하여 다중채널 표면파 탐사(Multichannel Analysis of Surface Wave, MASW)와 상시미동 배열 탐사(Microtremor Array Measurement, MAM)를 각각 수행하였다. MASW와 MAM은 각각 고주파수 영역 및 저주파수 영역의 분산곡선 획득에 적합하다. 두 탐사로 획득한 분산곡선을 취합하여 하나의 분산곡선을 형성한 후, 역산을 통해 64개의 심부 V_S 주상도를 평가하였다. Fig. 3은 표면파 탐사로 평가한 V_S 주상도의 V_{S30} 및 최종 평가 깊이를 보여준다. 대다수의 V_{S30} 은 700 m/s에서 1,600 m/s 범위에 있으며(Fig. 3(a)), 최종 평가 깊이의 최솟값과 최댓값은 각각 105 m와 330 m이다(Fig. 3(b)).

분산곡선을 통해 V_S 주상도를 평가할 때, 각 층의 두께는 층 비율(layering ratio)과 최소 해상 파장으로 결정한다[11]. 따라서 표면파 탐사로 평가한 V_S 주상도는 시추공 탐사로 평가한 V_S 주상도(1 m마다 V_S 평가 가능)보다 큰 층 두께 간격을 가지며, 이로 인해 두 V_S 주상도 세트의 V_{S30} 구간에 약간의 차이가 발생했을 가능성이 있다. 그러나, Fig. 4는 시추공 탐사로

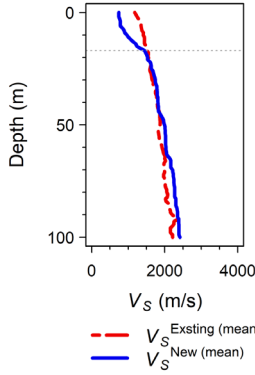


Fig. 4. The comparison between the average existing V_S profile evaluated through borehole investigations (i.e., cross-hole test and down-hole test) and the average newly obtained V_S profile evaluated through surface wave methods (i.e., MASW and MAM).

평가한 기존 V_S 주상도의 평균($V_S^{Existing(mean)}$)과 표면파 탐사로 추가 평가한 V_S 주상도의 평균($V_S^{New(mean)}$)이 심도 17 m 이후로 유사한 것을 보여준다. 따라서 두 V_S 주상도 세트는 병합되어 예측 모델 개발 데이터로 사용되었다. 예측 모델 개발에 사용된 총 V_S 주상도의 개수는 243 개이다.

3. 회귀분석 기반의 예측 모델 개발

3.1 개발 방법

Vrettos[12]는 지반의 깊이에 따라 변화하는 V_S 의 식을 아래와 같이 제안하였다.

$$V_S(z) = V_{S0}(1+kz)^{\frac{1}{n}}, \quad n > 1 \quad (1)$$

여기서, z 는 깊이(m)를 나타내는 변수이고, V_{S0} 은 지표면($z = 0$)에서의 V_S (m/s)를 의미하며, k 와 n 은 각각 V_S 주상도의 기울기와 곡률 계수이다. k 는 지표면 근처의 V_S 변화율을, n 은 심부 깊이의 V_S 변화율을 결정한다. 식(1)은 깊이가 깊어질수록 V_S 가 지속적으로 증가하는 것을 가정한다.

Shi and Asimaki[6]는 미국 캘리포니아 지역의 퇴적 분지에 대한 V_S 주상도 예측을 위해 식(1)을 아래와 같이 수정하였다.

$$V_S(z) = \begin{cases} V_{S0}, & 0 \leq z \leq 2.5 \text{ m} \\ V_{S0}(1+k(z-2.5))^{1/n}, & z > 2.5 \text{ m} \end{cases} \quad (2)$$

여기서, V_{S0} 은 지표면에서 깊이 2.5 m까지의 평균 V_S 를 의미하며, 이는 수집한 V_S 주상도 데이터의 평가 방법(시추공 탐사와 표면파 탐사)에 따른 지표 부근에서의 V_S 차이를 보정하기 위해 깊이 2.5 m까지의 V_S 를 상수로 가정한 것이다. Shi and Asimaki[6]는 수집한 V_S 주상도를 V_{S30} 기준으로 분류하고 평균 V_S 주상도를 산출한 후, 식(2)의 형태에 적합시켜 모델변수 V_{S30} , k , n 을 도출하였다. 이 모델 변수들은 V_{S30} 과 상관관계 분석을 통해 V_{S30} 의 함수식으로 표현되어 식(2)가 하나의 독립변수(V_{S30})로 V_S 주상도를 예측하도록 작성하였다.

본 연구는 Shi and Asimaki[6]에서 고려된 기법에 따라 수집한 V_S 주상도를 V_{S50} 값을 기준으로 분류하고 평균 V_S 주상도를 산출하여 깊이 50 m 하부의 V_S 주상도 예측을 위한 회귀분석 모델을 개발하였다. 식(2)를 본 연구에 적합하도록 수정한 식은 다음과 같다.

$$V_S(z) = V_{S50}^*(1+k(z-50))^{\frac{1}{n}}, \quad z > 50 \text{ m} \quad (3)$$

여기서, V_{S50}^* 은 깊이 50 m에서의 V_S 를 의미한다. 즉, 본 연구에서 개발한 회귀분석 모델은 깊이 50 m까지의 V_S 주상도의 정보를 활용하여 깊이 50 m 하부의 심부 V_S 주상도를 예측한다.

수집한 V_S 주상도를 6개의 V_{S50} 그룹으로 분류하였다. 통계적으로 편향되지 않는 평균 V_S 주상도의 추정을 위하여 각 V_{S50} 그룹은 20개 이상의 V_S 주상도를 포함하며, 인접한 그룹 간에 중복으로 포함된 V_S 주상도는 없다. V_S 가 로그 정규 분포를 따른다고 가정하여 깊이에 따른 V_S 의 평균(AVG_{V_S})을 다음과 같이 계산하였다[13].

$$AVG_{V_S} = \exp(\mu_{\ln(V_S)} + \frac{\sigma_{\ln(V_S)}^2}{2}) \quad (4)$$

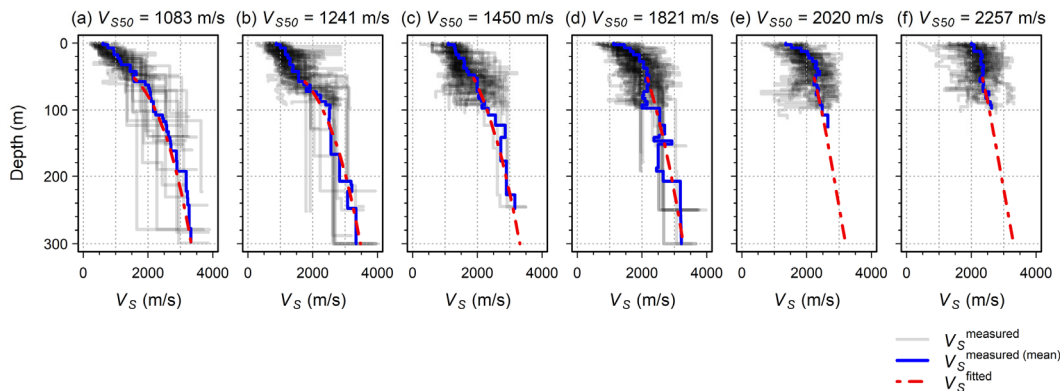


Fig. 5. Individual and average measured V_S profiles, along with the fitted V_S profiles, for six groups classified by V_{S50} .

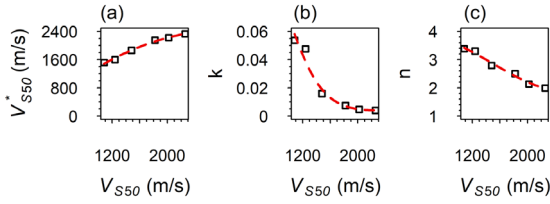


Fig. 6. Relationship between V_{S50}^* , k , n derived from the fitting and V_{S50} of averaged V_S profiles and their functions

여기서, $\mu_{\ln(V_S)}$ 와 $\sigma_{\ln(V_S)}$ 은 $\ln(V_S)$ 의 표본 평균과 표본 표준편차다. 계산된 평균 V_S 주상도는 모델변수 유도를 위하여 식 (3)에 적합(curve fitting)되었다. Fig. 5는 6개의 V_{S50} 그룹에 속한 개별 V_S 주상도(회색 실선)와 평균 V_S 주상도(파란색 실선), 식 (3)에 적합시킨 V_S 주상도(붉은색 점선)를 보여준다. 모든 그룹에서 평균 V_S 주상도는 전반적으로 깊이에 따라 증가하며, V_{S50} 값이 작은 그룹에서 V_S 변화율이 큰 경향을 보인다(Fig. 5(a)와 (b)). 식 (3)에 적합시킨 V_S 주상도를 깊이 300 m까지 나타냈으며, 모든 V_{S50} 그룹에서 깊이 300 m에서의 V_S 값이 3,000 m/s 이상으로 나타난다.

각 V_{S50} 그룹 내 평균 V_S 주상도를 식 (3)에 적합시켜 획득한 모델변수 V_{S50}^* , k , n 를 각 평균 V_S 주상도의 V_{S50} 값과 상관관계 분석하여 Fig. 6에 도시하였다. 분석한 경향은 다음과 같다: 1. V_{S50} 이 증가함에 따라 V_{S50}^* 이 증가한다. 이는 암반 부지의 평균 천부 강성이 클수록 50 m에서의 강성이 큰 것을 의미한다. 2. V_{S50} 이 증가함에 따라 k 와 n 이 감소한다. 이는 암반 부지의 천부 강성이 클수록 심부 깊이에서의 V_S 증가율이 작아진다는 것을 의미한다. 3. 보편적으로 천부 강성과 관계없이 심부 깊이가 갈수록 V_S 가 특정한 값에 접근함을 암시한다. 이를 기반으로 유도된 해석적 함수식은 다음과 같다.

$$V_{S50}^* = a_1 \times (V_{S50})^2 + a_2 \times (V_{S50}) + a_3 \tag{5a}$$

$$k = \exp(b_1 \times (V_{S50})^3 + b_2 \times (V_{S50})^2 + b_3 \times (V_{S50}) + b_4) \tag{5b}$$

$$n = c_1 \times (V_{S50})^3 + c_2 \times (V_{S50})^2 + c_3 \times (V_{S50}) + c_4 \tag{5c}$$

여기서, $a_1 = 118.02$, $a_2 = 1.74$, $a_3 = -44.13$, $b_1 = 1.92 \times 10^{-9}$, $b_2 = -8.52 \times 10^{-6}$, $b_3 = 9.48 \times 10^{-3}$, $b_4 = -5.56$, $c_1 = 4.72 \times 10^{-10}$, $c_2 = -2.13 \times 10^{-6}$, $c_3 = 1.78 \times 10^{-3}$, $c_4 = 3.39$ 이다. Fig. 6에 제시된 각 모델변수의 결정계수 R^2 은 모두 0.99로 우수한 적합성을 보인다.

3.2 모델 평가

회귀분석 모델의 평가를 위하여 세 곳의 지역(부지 a, b, c)에서 평가된 V_S 주상도를 검증 데이터로 사용하였다(보안상 위치 정보는 공개할 수 없다). 선택된 검증 데이터는 모델 개발에 사용되지 않은 굴착 암반 부지에서 획득한 V_S 주상도의 평균이다. 부지 a, b, c의 V_{S50} 을 산정한 후, 식 (5)를 통해 계산한 모델변수 (V_{S50}^* , k , n)를 식 (3)에 대입하여 예측 V_S 주상도

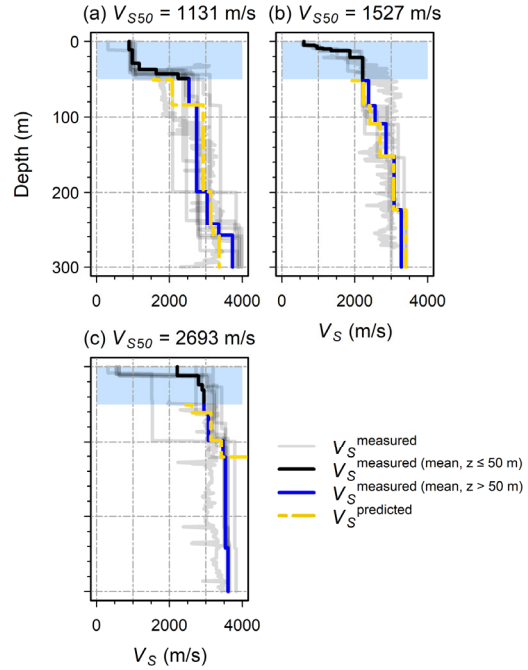


Fig. 7. Predicted deep V_S profiles from 50 m using the regression model and measured V_S profiles. The order of panel labels (a-c) represent three different sites

Table 1. Goodness-of-fit scores corresponding to the regression model (gof_{V_S}). The order of panel labels (a-c) represent three different sites

Site	gof_{V_S}	
	Before modification	After modification
(a)	8.0	9.2
(b)	9.3	
(c)	7.2	

($V_S^{predicted}$)를 획득하였으며, 각 부지의 V_{S50} 값과 함께 Fig. 7에 나타내 평균 예측 V_S 주상도 ($V_S^{measured(mean)}$)와 비교하였다. 이때, 정량적인 비교를 위하여 예측 V_S 주상도는 평균 V_S 주상도의 층두께를 따라 계단 형태의 주상도로 표현하였다. 부지 a와 b(Fig. 7(a)와 (b))의 V_{S50} 은 모델 개발에 사용된 평균 V_S 주상도의 V_{S50} 범위 $\{1,083 \leq V_{S50} \text{ (m/s)} \leq 2,257\}$ 내에 있으며, 부지 c의 V_{S50} 은 모델 개발에 사용된 평균 V_S 주상도의 V_{S50} 범위를 벗어난다. 모델 개발에 사용된 V_{S50} 범위 내의 V_{S50} 값을 갖는 경우, 예측 V_S 주상도는 전반적으로 평균 V_S 주상도와 유사한 경향을 보인다(Fig. 7(a)와 (b)). 그러나, 모델 개발에 사용된 V_{S50} 범위를 벗어난 V_{S50} 값을 갖는 경우, 예측 정확도가 저하된다(Fig. 7(c)).

모델의 성능을 정량적으로 평가하기 위해 Shi and Asimaki[6]가 제시한 적합도 지표(goodness of fit, gof_{V_S})를 활용하여 그 값들을 Table 1에 나열하였다. gof_{V_S} 는 다음과 같이 계산된다.

$$gof_{V_S} = 10 - 10 \left(\frac{10 \left| \ln V_S^{predicted} - \ln V_S^{measured} \right|}{\left| \ln V_S^{measured} \right|} \right) \tag{6}$$

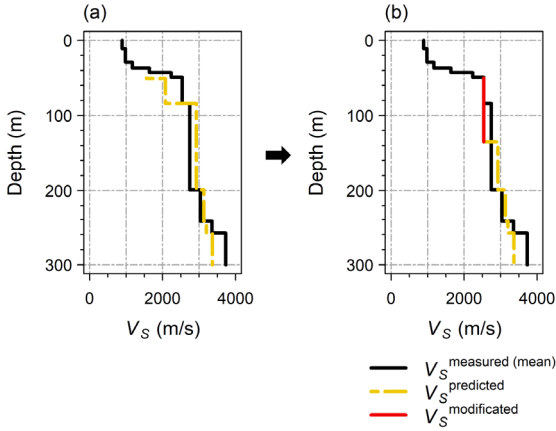


Fig. 8. Example of a modification method for improving velocity prediction when the predicted V_S at 50 m is significantly lower than the measured V_S : (a) original V_S profile prediction and (b) modified V_S profile prediction

분자에 곱해진 10은 모든 점수가 10에 가깝지 않도록 방지하며, 암반층의 큰 V_S 를 고려하여 적용한 값이다. 깊이 간격을 1 m로 하여 gof_{V_S} 를 계산한 후, 전체 깊이에 대해 평균화하였다. gof_{V_S} 가 10이면 완벽한 적합성을 의미하며, 0에 가까울수록 낮은 적합성을 의미한다. 부지 b에 대한 gof_{V_S} 은 9 이상으로 계산되어, 회귀분석 모델이 우수한 예측 성능을 제공함을 나타낸다. 반면, 부지 a와 c에서의 gof_{V_S} 는 각각 8과 7.2로 계산되었는데, 이는 다른 이유가 기인한다.

Fig. 7(a)에서 관찰되는 바와 같이, 부지 a에서는 예측이 시작되는 시점(깊이 50 m)에서 예측된 V_S (1,557 m/s)가 측정된 V_S (2,538 m/s)보다 현저히 낮은 값을 가진다. 이는 깊이 50 m 직전에서 V_S 가 급격히 증가하는 층이 존재하기 때문이며, Fig. 8과 같이 예측 V_S 주상도에서 깊이 50 m의 측정 V_S 값(2,538 m/s)에 상응하는 깊이(135 m)까지 일정한 V_S 값으로 가정하는 방법이 대안으로 제안된다. 이 방법으로 수정한 예측 V_S 주상도의 gof_{V_S} 는 9.2로 계산되었다. 부지 c에서의 오차는 입력변수인 V_{S50} 의 값이 커서 모델변수 V_{S50}^* , k , n 의 계산값에 영향을 주었기 때문이다. 따라서, 본 연구에서 개발한 회귀분석 모델을 사용하여 심부 깊이의 V_S 예측을 시도할 경우, V_{S50} 은 약 1,100 m/s에서 2,200 m/s 사이에 있어야 한다.

4. 머신러닝 기반의 예측 모델 개발

4.1 개발 방법

학습(Training) 데이터 세트는 깊이 300 m까지 평가된 64개의 심부 V_S 주상도를 사용하였다. 학습 데이터 세트는 깊이 50 m를 기준으로 이전 깊이의 V_S 주상도와 그 후 깊이의 V_S 주상도로 분리되었으며, 각각 입력변수(깊이와 V_S)와 목표값으로 활용되었다. 심부 깊이의 V_S 예측 모델 개발을 위하여 세 종류의 머신러닝 알고리즘{랜덤 포레스트(Random Forest, RF), 그라디언트 부스트(Gradient Boost, GB), 인공신경망(Artificial Neural

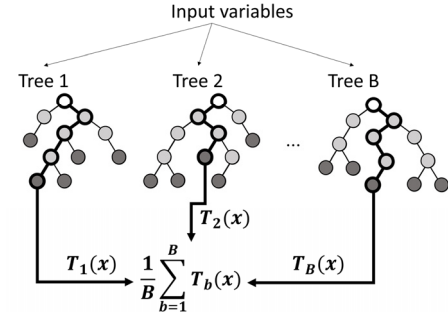


Fig. 9. Example of the random forest (RF) method architecture

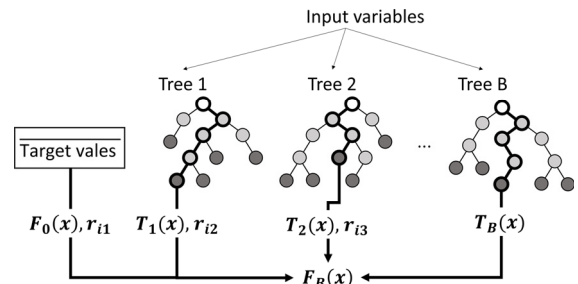


Fig. 10. Example of the gradient boosting (GB) method architecture

Network, ANN)}을 선택하였고, Figs. 9-11은 이들의 모식도를 보여준다.

RF와 GB는 결정 수목에 기반한 알고리즘이다. 결정 수목은 세 가지 유형의 노드(루트 노드, 내부 노드, 리프 노드)로 구성된다. 루트 노드는 트리의 시작점으로, 입력변수를 기준으로 데이터를 최초로 분리한다. 이후, 내부 노드에서 추가 조건에 따라 데이터를 더 세분화하고, 수목의 마지막에 위치한 리프 노드에 포함된 데이터로 예측값을 결정한다. 이러한 구조를 통해 결정 수목은 입력 공간을 여러 개의 영역으로 나누고, 각 노드에 해당하는 값을 할당한다. 식 (7)는 트리 모델 $T(x)$ 를 설명한다.

$$T(x) = \sum_{j=1}^J c_j I(x \in R_j) \quad (7)$$

여기서, x 는 입력변수, J 는 리프 노드 수, c_j 는 리프 노드 R_j 에 속하는 목표값의 평균을 의미한다. I 는 지표 함수로, x 가 R_j 에 속하면 1, 그렇지 않으면 0의 값을 가진다. $T(x)$ 는 평균 제곱 오차(mean square error, MSE)값을 최소로 갖는 R_j 에서의 예측값을 출력한다.

RF는 다수의 결정 수목(T_1, T_2, \dots, T_B)을 독립적으로 학습시키고, 각 수목에서 반환하는 예측값(F_1, F_2, \dots, F_B)의 평균을 최종 예측값($F(x)$)으로 출력한다(Fig. 9). 해당 알고리즘은 과적합을 방지하고 예측의 정확성을 높이는 데 효과적이다.

$$F(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (8)$$

여기서, B 는 수목의 개수, $T_b(x)$ 는 각 수목의 예측값이다. 본 연구에서는 50개의 수목을 고려하여 예측 모델을 개발하였다.

GB는 여러 개의 결정 수목(T_1, T_2, \dots, T_B)을 순차적으로 성장시키며, 각 수목은 이전 수목의 잔차(r_{ib})를 보완하도록 구축된다(Fig. 10). 이러한 방식은 오차를 점진적으로 줄이는 테이점을 갖는다. 이때, 첫 번째 수목(T_1)을 성장시킬 잔차(r_{i1})를 위한 $F_0(x)$ 을 목표값의 평균으로 설정하였다. r_{ib} 의 계산식은 아래와 같다.

$$r_{ib} = (V_S^{deep})_{mea,i} - F_{b-1}(x_i) \tag{9}$$

여기서, $(V_S^{deep})_{mea,i}$ 와 $F_{b-1}(x_i)$ 는 각각 측정된 i 번째 목표값과 이전 수목이 출력한 i 번째 예측값을 의미한다. 결정 수목(T_b)은 x_i 와 r_{ib} 의 쌍, $\{x_i, r_{ib}\}_{i=1}^{64}$ 를 사용하여 성장하며, 리프노드 R_{j_b} 를 결정한다. 각 R_{j_b} 에서 예측된 잔차의 평균값을 r_{j_b} 라고 할 때, r_{j_b} 를 생성하는 $T_b(x)$ 는 아래와 같이 설명할 수 있다.

$$T_b(x) = \nu \sum_{j=1}^{J_b} c_j I(x \in R_{j_b}) \tag{10}$$

여기서, J_b 는 T_b 의 리프노드 수를 의미한다. ν 는 학습률에 관한 가중치이며, 0.1로 적용하였다[14]. RF와 같이 50개의 수목을 고려하였으며, 최종 예측값($F_B(x)$)은 다음과 같이 계산된다.

$$F_B(x) = F_0(x) + \sum_{b=1}^B T_b(x) \tag{11a}$$

$$F_b(x) = F_{b-1}(x) + T_b(x) = F_0(x) + T_1(x) + \dots + T_b(x) \tag{11b}$$

ANN은 여러 개의 노드로 구성된 은닉층을 가지고 있으며, 각 노드는 입력 값과의 연산을 통해 예측값을 출력하는 구조이다(Fig. 11). 본 연구에는 완전 연결층을 사용한 두 개의 은닉층이 고려되었으며, 첫 번째와 두 번째 은닉층의 노드 개수는 각각 16개와 32개이다. 첫 번째 은닉층의 노드 값($h_j^{(1)}$) 및 두 번째 은닉층의 노드 값($h_k^{(2)}$)과 예측값($F(x)$)은 아래의 식과 같이 나타낼 수 있다.

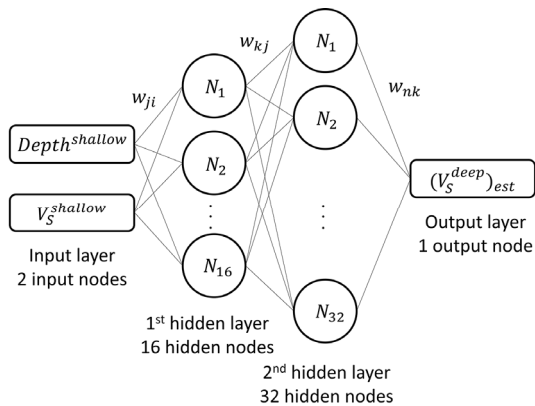


Fig. 11. Example of the artificial neural network (ANN) architecture that consists of two input layers, two hidden layers with 16 and 32 nodes (N), and an output layer. w represents the weights between nodes

$$h_j^{(1)} = f^{(1)}\left(\sum_{i=1}^2 (w_{ji}x_i + b_j)\right) \tag{12a}$$

$$h_k^{(2)} = f^{(2)}\left(\sum_{j=1}^{16} (w_{kj}h_j^{(1)} + b_k)\right) \tag{12b}$$

$$F(x) = \sum_{k=1}^{32} (w_{nk}h_k^{(2)} + b_n) \tag{12c}$$

여기서, x_i 는 i 번째 노드, w_{ji} 는 입력층 i 번째 노드와 첫 번째 은닉층의 j 번째 노드 사이의 가중치, w_{kj} 는 첫 번째 은닉층의 j 번째 노드와 두 번째 은닉층의 k 번째 노드 사이의 가중치, w_{nk} 는 두 번째 은닉층의 k 번째 노드와 출력층의 n 번째 노드(즉 $n=1$) 사이의 가중치를 의미한다. b_j, b_k, b_n 은 각각 첫 번째 은닉층의 j 번째 노드, 두 번째 은닉층의 k 번째 노드, 출력층의 n 번째 노드의 편향값으로 모델의 적합성을 높이는 역할을 한다. 정류 선형 함수(Rectified Linear Unit, ReLU)를 사용한 활성화 함수(f)는 모델에 비선형성을 부여하며, f 의 출력값($h_j^{(1)}, h_k^{(2)}$)은 다음 층 노드의 입력값이 된다. 이때, ANN으로 개발된 예측 모델은 데이터의 크기 및 범위에 영향을 받는 것이 확인되어, 정규화 방식(mix-max scaling)을 통해 입력변수와 목표값을 0부터 1 사이에 분포하도록 변환하였다.

4.2 모델 평가

머신러닝 모델의 평가를 위하여 회귀분석 모델 평가에 사용한 부지 세 곳의 V_S 주상도를 검증 데이터로 활용하였다. 머신러닝 모델은 데이터 학습 시 개별 V_S 주상도를 사용해 입력변수와 목표값을 1대 1로 대응시켰다. 따라

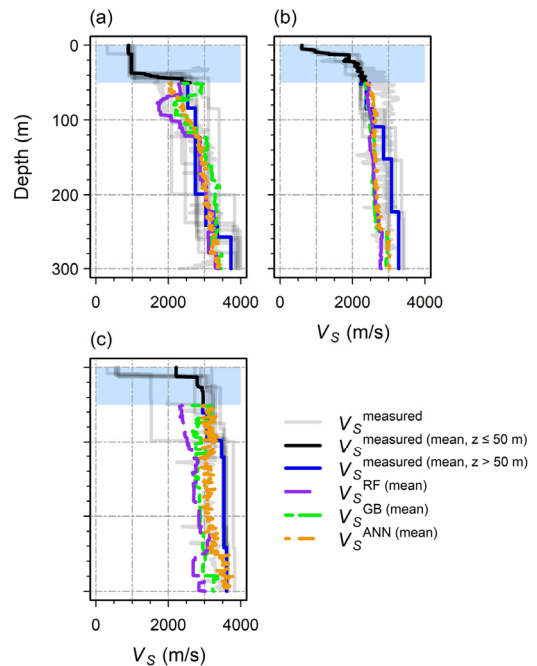


Fig. 12. Predicted deep V_S profiles from 50 m using the machine learning models (RF: random forest, GB: gradient boosting, ANN: artificial neural network) and measured V_S profiles. The order of panel labels (a-c) represent three different sites

Table 2. Goodness-of-fit scores corresponding to the machine learning models ($gof_{V_s}^{RF}$: for random forest, $gof_{V_s}^{GB}$: for gradient boosting, $gof_{V_s}^{ANN}$: for artificial neural network). The order of panel labels (a-c) represent three different sites

Site	$gof_{V_s}^{RF}$	$gof_{V_s}^{GB}$	$gof_{V_s}^{ANN}$
(a)	8.6	8.5	9.1
(b)	8.5	8.4	8.7
(c)	7.9	7.1	9.0

서, 검증 데이터로 부지별 평균 V_s 주상도가 아닌 23개의 개별 V_s 주상도를 활용하였다. 각 머신러닝 모델(RF, GB, ANN)은 23개의 예측 V_s 주상도를 출력하며, 각 머신러닝에서 부지별로 출력한 V_s 주상도를 평균하여 최종 예측 V_s 주상도를 획득하였다. Fig. 12는 세 종류의 머신러닝 모델로 예측한 평균 V_s 주상도를 평균 계측 V_s 주상도와 비교한 그림이다. $V_s^{measured (mean)}$ 는 계측 V_s 주상도의 평균이며, $V_s^{RF (mean)}$, $V_s^{GB (mean)}$, $V_s^{ANN (mean)}$ 은 각각 RF, GB, ANN 모델이 예측한 V_s 주상도의 평균을 의미한다. 예측 V_s 주상도는 계측 V_s 주상도와 비교적 유사함을 확인할 수 있다. Table 2에 예측 모델의 gof_{V_s} 결과를 수록하였다. ANN으로 개발된 예측 모델은 가장 높은 gof_{V_s} 점수를 가지며, 부지 c의 V_s 주상도에 대한 예측 성능도 우수한 것으로 확인되었다.

5. 결론

본 연구에서는 국내 굴착 암반 부지의 심부 V_s 주상도를 예측하기 위해 회귀분석 및 머신러닝 기반의 예측 모델을 개발하고, 그 성능을 검증하였다. 모델 개발에는 179개의 천부 V_s 주상도와 64개의 심부 V_s 주상도를 활용하였으며, 검증에는 23개의 심부 V_s 주상도를 활용하였다. 회귀분석 모델은 Shi and Asimaki[6]가 제안한 수학적식을 본 연구 목표에 적합하도록 수정하여 개발되었다. 이 모델은 V_{s50} 값을 기준으로 심부 V_s 의 추정치를 제공한다. 검증을 통해 회귀분석 모델은 입력변수와 모델변수 간 단순한 상관관계에 기반한 예측 기법임에도 불구하고 심부 V_s 의 V_s 변화 경향성을 잘 추정하는 것을 확인했다. 그러나 모델 개발에 사용되지 않은 V_{s50} 값에 대한 외삽은 불확실성이 있어 권장되지 않는다. 머신러닝 모델은 수목 기반 알고리즘과 완전 연결 계층 구조를 활용하여 개발되었다. 깊이 50 m까지의 V_s 를 입력으로 하여 300 m까지의 V_s 를 1 m 간격으로 출력하도록 설계하였으며, 완전 연결 계층 구조를 사용한 경우에 높은 정확성을 나타냈다. 본 연구는 회귀분석 모델과 머신러닝 모델을 통해 천부 V_s 로 심부 V_s 를 예측할 수 있는 가능성을 제시하였다. 예측 모델들은 시추공 탐사가 어려운 굴착 암반 부지의 심부 V_s 주상도 확보에 기여할 것으로 기대된다. 다만, 더욱 정교한 예측을 위하여 향후 연구에서 굴착 암반 부지의 특성을 반영한 깊이에 따른 층상 구조(지층 두께) 모델 개발이 필요할 것으로 사료된다.

/ 감사의 글 /

본 논문은 한국수력원자력(주)에서 재원을 부담하여 “제 2022-기술-03

호”에서 수행한 연구입니다. 머신러닝 기반의 예측 모델 개발과 관련하여 도움 주신 교토대학교 김지송 박사님께 감사를 표합니다.

/ REFERENCES /

1. Ministry of Land, Infrastructure and Transport. Korean Design Standard, KDS 17 10 00; c2024.
2. Hashash YM, Kottke AR, Stewart JP, Campbell KW, Kim B, Moss C, Nikolaou S, Rathje EM, Silva WJ. Reference rock site condition for central and eastern North America. *Bull Seismol Soc Am*. 2014; 104(2):684–701.
3. Poggi, V., Edwards, B., Fäh, D. Derivation of a reference shear-wave velocity model from empirical site amplification. *Bull Seismol Soc Am*. 2011;101(1):258–274.
4. Kawase H, Nagashima F, Nakano K, Mori Y. Direct evaluation of S-wave amplification factors from microtremor H/V ratios: Double empirical corrections to “Nakamura” method. *Soil Dyn Earthq Eng*. 2019;126:105067.
5. Sun CG, Chung CK, Kim DS. Determination of mean shear wave velocity to the depth of 30m based on shallow shear wave velocity profile. *J Earthq Eng Soc Korea*. 2007;11(1):45–57.
6. Shi J, Asimaki D. A generic velocity profile for basin sediments in California conditioned on V_{s30} . *Seismol Res Lett*. 2018;89(4): 1397–1409.
7. Passeri F, Foti S, Rodriguez-Marek A. A new geostatistical model for shear wave velocity profiles. *Soil Dyn Earthq Eng*. 2020;136: 106247.
8. Singh, S., Kanli, A. I. Estimating shear wave velocities in oil fields: a neural network approach. *Geosci J*. 2016;20:221–228.
9. Anemangely M, Ramezanzadeh A, Amiri H, Hoseinpour SA. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *J Pet Sci Eng*. 2019;174:306–327.
10. Rajabi M, Hazbeh O, Davoodi S, Wood DA, Tehrani PS, Ghorbani H, Mehrad M, Mohamadian N, Rukavishnikov VS, Radwan AE. Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. *J Pet Explor Prod Technol*. 2023;13(1):19–42.
11. Cox BR, Teague DP. Layering ratios: a systematic approach to the inversion of surface wave data in the absence of a priori information. *Geophys J Int*. 2016;207(1):422–438.
12. Vrettos C. Simple inversion procedure for shallow seismic refraction in continuously nonhomogeneous soils. *Soil Dyn Earthq Eng*. 1996; 15(6):381–386.
13. Toro G. Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies. Brookhaven National Library, Upton, New York; c1995.
14. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.