Machine learning-based estimation of soil’s true air-entry value from GSD curves

Mohammad Sadegh Es-haghi a,b, Mohammad Rezania c,d, Meghdad Bagheri d

a Escuela Técnica Superior de Ingenieros Navales (ETSin), Universidad Politécnica de Madrid (UPM), Av. de la Memoria, 4, 28040 Madrid, Spain
b Centre Internacional de Métodes Numèrics a l’Enginyeria (CIMNE), Edifici C1, Campus Norte, UPC, Gran Capitán s/n, 08034 Barcelona, Spain
c School of Engineering, University of Warwick, Coventry CV4 7AL, UK
d School of Energy, Construction and Environment, Coventry University, Coventry CV1 5FB, UK

A R T I C L E   I N   P R E S S

Abstract
The application of machine learning (ML) methods has proven to be promising in dealing with a wide range of geotechnical engineering problems in recent years. ML methods have already been used for the prediction of soil water retention curves (SWRC) and estimation of air-entry values (AEV). However, the reported works in the literature are generally based on limited data and conventional, less accurate approaches for AEV estimation. In this paper, a large database, known as UNSaturated SOil hydraulic DAtabase (UNSODA), is studied and the conventional and true AEVs of 790 soil samples are estimated based on determination methods reported in the literature. A ML approach is then employed for the development of a predictive model for the estimation of true AEV from water content-based SWRCs of a wide range of soil types taking into account the impact of bulk density and grain size distribution parameters. The obtained results reveal an enhanced accuracy in AEV determination, featuring R^2 values of 0.964, 0.901 and 0.851 for training, validation, and testing data, respectively, which confirm the high level performance of the developed ML model. Based on the results of a sensitivity analysis, the particle sizes of 50 and 250 µm are found to have the highest impact on the AEV estimation. © 2022 The Author(s). Published by Elsevier B.V. on behalf of International Association for Gondwana Research. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0).

1. Introduction

In unsaturated soil mechanics, the relationship between the amount of water held by the soil and the pore-water tension (suction) is generally presented in the form of water content (w) versus suction (s), or degree of saturation (Sr) versus suction in a semi-logarithmic graph known as soil water retention curve (SWRC). Generally, three distinct parts are recognized on a SWRC (Fig. 1). Initially, with the soil undergoing drying, the curve remains a relatively flat line until the air-entry value (AEV); during this stage, with an increase in suction, the pores remain saturated (Sr = 100%) with pore-water being under tension (saturated state or boundary effect zone). As suction exceeds the AEV, air breaks into the larger pores, and a continuous network of air-filled and water-filled pores is formed. During this stage, the degree of saturation (or water content) is progressively decreased as more water evaporates or leaves the pores (partially saturated state or transition effect zone). Finally, with an increase in suction, a state is reached where water only remains at the particle contacts and is no longer continuous with the pore space. At this stage, the SWRC almost flattens meaning that much less water will be expelled with an increase in suction, and the soil dries without significant volume changes (residual state). The two inflection points, namely AEV and residual suction (sres), are considered as fundamental parameters for the determination of soils’ water retention properties, with the former being a key input parameter in several unsaturated constitutive models (e.g., Alonso et al. 1990; Russell and Khalili 2006), hence, highlighting the importance of its accurate estimation from SWRCs.

On a degree of saturation-based SWRC, the AEV is graphically determined as the intersection point of a horizontal tangent line drawn to the saturated portion, with a tangent line drawn to the transition portion of the curve (see Fig. 1). Pasha et al. (2016) showed that using the same method to derive AEV from water content-based SWRCs can produce highly erroneous results. They proposed a simple method for estimation of true AEV from water content-based SWRC. This method was based on plotting the data on both semi–log and log–log scales on the same graph. In this way, the saturated part of the curve is identified by a linear or
Rezania (2022) later showed that direct measurement of soil suction is promising. Bagheri (2018), Bagheri et al. (2018), and Bagheri and Rezania (2016) reported in the literature on the estimation of the SWRC from water content-based methods (Schaap and Leij 1998) are also available. However, it is apparent that accurate estimation of AEV is profoundly subordinated to the accuracy of predicted SWRCs. Furthermore, the reported works on statistical and AI methods generally utilized databases that were not large enough to include a variety of soil types. It is therefore imperative to consider a relatively large database to directly estimate true AEV from common soil parameters.

Recently, machine learning (ML) approaches have proven to be promising in solving nonlinear and complicated problems using large databases (Rezania 2008; Javadi and Rezania 2009a; Zhang et al. 2020; Zhang et al. 2021; Wang et al., 2020b; Zhang et al. 2022; Wang et al., 2020c). In geotechnical engineering, the ML algorithms have been favorably employed for predicting various phenomena such as the settlement of shallow foundations on cohesionless soils (Rezania and Javadi 2007), thermo-hydraulic behavior of hydrate reservoirs (Zhou et al. 2020), non-stationary and non-Gaussian geotechnical properties (Shi and Wang 2021), soil constitutive modeling (Javadi and Rezania 2009b; Rezania and Ma 2019) and suction distribution in shallow soil layers (Cheng et al. 2020). In these studies, various ML algorithms including support vector machine (SVM), multi-expressions and AI methods (Schaap and Leij 1998) are also available. How-
grain size, etc.) on AEV is investigated. Finally, a neural network model is employed to estimate the true AEVs. Sensitivity analysis is also performed in order to determine the consideration of which parameters are more significant for the AEV prediction.

2. UNSODA

UNSODA is a collection of data for unsaturated hydraulic soil properties in Microsoft Access-97 format. The data for 790 soil samples have been categorized in 36 tables, arranging them in rational groups with relevant information. A 4-digit code has been allocated to each soil sample. Fig. 2 shows the frequency histogram of soil classification of the 790 samples according to USDA (United States Department of Agriculture) system and the summary of their geographical distribution represented in UNSODA. Most of the samples of the database are from Europe and North America. The majority of the soil samples are coarse-textured although there is an acceptable number of fine soil samples.

UNSODA provides a wide range of soil properties as shown in Table 1. It must be noted here that bulk density, particle density, porosity, saturated conductivity, and saturated water content are the only parameters required for the estimation of true AEV in the present study. In addition, UNSODA provides appropriate information about the grain size distribution of samples; however, there is no uniform set of particle sizes in the UNSODA, and the reported values of particle fraction are presented for different sizes in different samples due to the differences in experimental methods. For instance, in soil sample 1090, particle fractions for the particle sizes of 2, 20, 50, 250, 500, 1000, and 2000 μm have been determined; however, in the soil sample 1087 only the particle size of 2, 50, and 2000 μm have been considered.

In this study, the particle size fractions of 2, 20, 50, 250, 500, 1000, and 2000 μm, which have the most frequency in the database are used. Fig. 3 presents the frequency of particle sizes in the UNSODA database.

Table 2 presents the number of hydraulic curves and the number of data pairs partitioned into drying and wetting parts, as well as curves obtained from laboratory or field measurements. In this study, the water retention curves (a total of 902 curves) which are curves obtained from laboratory or field measurements. In this Figure, the range of changes and how suction and water content data are distributed for the lab and field data are well illustrated. As it turns out, in the lab-based works, the results involve a wider range of suction and water content values.

3. True air entry value

Although many researchers recognize the first break in the drying branch of the water content-based SWRC as evidence of air-entry point (Fredlund and Xing 1994; Zhai and Rahardjo 2012; Bagheri et al., 2019b, Rezania et al., 2020), this may not be true for some deformable soils, depending on factors such as the mechanical characteristics of the soil, pore-size distribution index, and the stress history of the soil. In the graphical method presented by Pasha et al. (2016), estimation of the true AEV was carried out based on the stress history of the soil, and consideration of three stress states, namely normally consolidated, overconsolidated with \( s_{aw} > p'_c \), and overconsolidated with \( s_{aw} < p'_c \), where \( s_{aw} \) is the suction at the air-entry and \( p'_c \) is the preconsolidation pressure of the soil.

From the 867 SWRCs in the UNSODA database, 221 curves could not be used for the analysis for various reasons, such as an insufficient number of data points or soil samples not entering the transition zone. Therefore, only SWRCs that would allow for appropriate graphical estimation of the AEV are selected and considered in his study. Of the remaining 644 curves, 212 curves are related to the normally consolidated state, 182 curves are related to the overconsolidated state with \( s_{aw} > p'_c \), and 250 curves are related to the overconsolidated state with \( s_{aw} < p'_c \). As shown in Fig. 5, for all these 644 curves, the AEVs are obtained following the graphical approach proposed by Pasha et al. (2016) (denoted as true AEV) and also according to the conventional method of finding the intersection of the tangent lines to the boundary effect zone and the transition zone (denoted as AEV). In this figure, the horizontal axis represents the codes allocated to the samples, which can be traced in the UNSODA database. With some examples, Fig. 6 illustrates the differences between the evaluated AEVs from the two methods. In addition, Fig. 7 presents the frequency of variations between the true AEVs and the conventional AEVs which for the majority of cases appear to be in the range of 30 to 300 kPa.

4. Preparing the database

The recorded data in the UNSODA database is presented in a way that is not convenient for training the ML model. On the other
In this database, some of the parameters are not reported for all samples. For instance, the data points of the GSD curves are presented in different particle sizes. Therefore, at first, to extract suitable information from the database and arrange it in the way that is useful to train the neural networks, a sorting code is developed in MATLAB software. The developed code also allows for extrapolation of the missing data and formation of the complete database required for the ML.

As shown in Fig. 3, the most common particle size values in UNSODA are $P_{2,610}$, $P_{2,000}$, $P_{250}$, $P_{500}$, $P_{1000}$, and $P_{2000}$. As a result, the fraction corresponding to these values are selected from the GSD curve as part of the ML inputs, and for samples that do not have these particle sizes, the corresponding values are estimated using the method proposed in Vaz et al. (2020). To overcome the common limitation of the GSD curves, which is the lack of standardization for granulometric fractions collected from various soil.

Table 1
Soil properties provided in UNSODA database.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Available Values</th>
<th>Missing Values</th>
<th>Considered in this study?</th>
</tr>
</thead>
<tbody>
<tr>
<td>bulk density</td>
<td>Bulk density as a mass of solids per bulk volume</td>
<td>762</td>
<td>28</td>
<td>Yes</td>
</tr>
<tr>
<td>particle density</td>
<td>Mass of solids per volume of solids</td>
<td>339</td>
<td>391</td>
<td>Yes</td>
</tr>
<tr>
<td>porosity</td>
<td>Volume of voids per bulk volume</td>
<td>270</td>
<td>420</td>
<td>Yes</td>
</tr>
<tr>
<td>OM content</td>
<td>Organic Matter Content. The mass of organic matter content as a percentage of the total solid mass.</td>
<td>388</td>
<td>402</td>
<td>No</td>
</tr>
<tr>
<td>$k_{sat}$</td>
<td>Saturated Conductivity. The measured saturated hydraulic conductivity</td>
<td>429</td>
<td>361</td>
<td>Yes</td>
</tr>
<tr>
<td>$\theta_{sat}$</td>
<td>Saturated Water Content. The experimental water content of a water-saturated sample</td>
<td>305</td>
<td>485</td>
<td>Yes</td>
</tr>
<tr>
<td>CEC</td>
<td>Cation Exchange Capacity. in cmol of charge per kg of dry soil (i.e., meq/100 g soil)</td>
<td>150</td>
<td>640</td>
<td>No</td>
</tr>
<tr>
<td>pH</td>
<td>Measured soil pH</td>
<td>300</td>
<td>490</td>
<td>No</td>
</tr>
<tr>
<td>electrolyte level</td>
<td>The approximate total solute concentration of the soil solution during the experiments</td>
<td>26</td>
<td>764</td>
<td>No</td>
</tr>
<tr>
<td>SAR</td>
<td>Sodium Adsorption Ratio</td>
<td>80</td>
<td>710</td>
<td>No</td>
</tr>
<tr>
<td>ESP</td>
<td>Exchangeable Sodium Percentage</td>
<td>19</td>
<td>771</td>
<td>No</td>
</tr>
<tr>
<td>free Fe Al oxide</td>
<td>The mass fraction of the Fe and Al oxides as a percentage of the total solid phase</td>
<td>14</td>
<td>776</td>
<td>No</td>
</tr>
</tbody>
</table>

Fig. 3. Frequency of the particle size in the UNSODA database.

Table 2
Summary of the number of hydraulic curves.

<table>
<thead>
<tr>
<th>Hydraulic Curves</th>
<th>Number of Curves</th>
<th>Total Number of Data Pairs</th>
<th>Average Number of Data Pairs in each Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Field</td>
<td>Lab</td>
<td>Field</td>
</tr>
<tr>
<td>Water Retention (h-h)</td>
<td>Drying</td>
<td>137</td>
<td>730</td>
</tr>
<tr>
<td></td>
<td>Wetting</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>902</td>
<td>11,223</td>
</tr>
<tr>
<td>Hydr. Conductivity (h-K)</td>
<td>Drying</td>
<td>133</td>
<td>730</td>
</tr>
<tr>
<td></td>
<td>Wetting</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>871</td>
<td>11,223</td>
</tr>
<tr>
<td>Hydr. Conductivity (h-K)</td>
<td>Drying</td>
<td>294</td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>Wetting</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>607</td>
<td>10,874</td>
</tr>
<tr>
<td>Soil Water Diffusivity (h-D)</td>
<td>Drying</td>
<td>56</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Wetting</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>150</td>
<td>2751</td>
</tr>
</tbody>
</table>
In the above equation, $a$, $b$, and $c$ are equation parameters, $d$ is the particle size and $F(d)$ is the fraction corresponding to the particle size of $d$. Thus, for each sample, the parameters of Eq. (1) are extracted and the fraction values in the required particle size (if not present) are calculated. For instance, in Fig. 8, the fitted curve for sample 3340 in the database is presented and the values of $P_{2000}$ and $P_{900}$ for this sample are estimated. This process is repeated for all samples in order to complete the database.

The detailed statistics of all variables in the database, with the values of minimum, maximum, median, mean, and standard deviation are summarized in Table 3 and their descriptive frequency histograms are shown in Fig. 9. Bulk density and porosity have the highest frequency at 1.4–1.6 and 0.4–0.5, respectively and follow an approximately normal distribution (Fig. 9a and b). Regarding mass fractions (Fig. 9c–j), it is observed that the highest frequency of $P_{25}$ is between 0.0 and 0.1 for 279 soils (Fig. 9c), while the highest frequency of $P_{2000}$ is between 0.9 and 1.0 for 641 soils (Fig. 9j). In terms of the true and conventional AEVs, most of the values are in the range between 0 and 50 kPa (Fig. 9k and l), while the true AEV distribution is slightly more uniform than the conventional AEV distribution. Since the values of $P_{2000}$ and $P_{1000}$ do not change across different samples, they are not considered in the subsequent analyses.

Fig. 10 presents a comparison of the AEVs and true AEVs obtained for different soil classifications. As can be seen, the range of the true AEVs has shifted to larger values compared to their corresponding conventional AEVs.

The variations of the AEV and the true AEV with each input soil property are shown in graphs of Fig. 11. Also shown in the figure are the linear trend lines fitted to the data along with the coefficient of determination (COD), $R^2$. The COD allows for the assessment of the fitting accuracy and it is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (Y_i - \overline{Y}_i)^2}{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}$$

and

$$\overline{Y}_i = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

where $Y_i$ and $\overline{Y}_i$ are real and predicted output values for the $i$th output, $N$ is the number of outputs, and $Y$ is the average value of the real outputs. Overall, a comparatively low fitting accuracy ($R^2 < 0.177$) is obtained for variations of both AEV and true AEV with input soil properties. The parameters $P_{50}$, $P_{200}$, $P_{500}$, $P_{250}$, $P_{1000}$, and bulk density, having the highest $R^2$ values are selected for developing the ML model and the parameters $P_{2000}$ and porosity, having the lowest $R^2$ values, are omitted. This is done to avoid the complexity associated with developing the ML model.

5. Results of analysis

According to Fig. 11, the database is described by seven input parameters namely, bulk density, $P_{25}$, $P_{200}$, $P_{500}$, $P_{250}$, $P_{1000}$, and one output parameter namely, true AEV. In order to accurately generate the predictive model, the ML method of multi-layer perceptron (MLP), which is a feedforward class of artificial neural networks (ANNs), is selected because of its simplicity and availability (comparing different ML methods is out of the scope of the current study).

ANNs are commonly used to develop practical AI frameworks to deal with complex pattern-oriented engineering problems. The nonparametric nature of ANNs permits models to be established without any prior knowledge of the distribution of the input values or presumed interactions between variables of a model as required.
Fig. 5. AEV and true AEV for each sample of UNSODA (Blue: true AEV, Green: miscalculated AEV, Triangle: normally consolidated, Square: overconsolidated with $s_{uv} > p_c$, and Circle: overconsolidated with $s_{uv} < p_c$).
Fig. 5 (continued)
ANN can be described as a multivariate and multi-dimensional function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. It consists of an input layer of $n$ neurons (input values), an output layer of $m$ neurons (output values) and an arbitrary number of interior layers with a variable number of neurons, called hidden layers. The neurons are storage cells for scalar values obtained by an activation function applied to the neuron values in the previous layer (Fig. 12). To each neuron in the output and the hidden layers, a vector of weights and a scalar bias are associated, and the value $u_k^l$ stored by the $k$th neuron in the $l$th layer can be written in the form (Krenker et al. 2011):

$$u_k^l = \sigma^l \left( \sum_j w_{kj}^l u_j^{l-1} + b_k^l \right)$$  \hspace{1cm} (4)

where $\sigma^l$ is the activation function for the $l^{th}$ layer, and $w_{kj}^l$ and $b_k^l$ are the weights and biases respectively. Typical activation functions

by frequently used statistical methods (Rezania, 2008; Walczak, 2019).

Fig. 6. Graphical determination of true and miscalculated AEVs for four soil samples in the UNSODA database (a: sample 4440, b: sample 4612, c: sample 2670, d: sample 1135).

Fig. 7. Histogram of the differences between the conventional and the true AEVs.
used can be linear (i.e., $\sigma(x) = x$), or non-linear (e.g., $\sigma(x) = \tanh(x)$). Under mild assumptions on the activation functions, it can be shown that a neural network with even a single hidden layer is a universal function approximator, as given enough neurons, any continuous function on a compact domain can be approximated with arbitrary precision (Csáji, 2001). In case where more than one hidden layer is used (a so-called deep neural network), very efficient approximations can be achieved with a relatively small number of network parameters, i.e., weights and biases (Huang, 2003).

To identify the best structure of the ML model, an initial analysis is carried out to determine the optimal number of layers and neurons that can suitably represent the relationship between the inputs and output of the database. To speed up the optimization procedure, only 25% of the database is considered for this step.

Fig. 13 presents a comparison of the performance of all the examined structures for one-hidden-layer and two-hidden-layer...
Fig. 10. Comparison of AEV and true AEV of soils with different classifications in the UNSODA database.

Fig. 11. Basic linear fittings between AEV and true AEV and each input soil property: (a) Bulk density, (b) Porosity, (c) $P_{20\mu}$, (d) $P_{50\mu}$, (e) $P_{500\mu}$, (f) $P_{2500\mu}$, (g) $P_{500\mu}$, (h) $P_{1000\mu}$, (i) $P_{2000\mu}$. 

$R^2_{AEV} = 0.0029$  $R^2_{True AEV} = 0.0293$

$R^2_{AEV} = 0.0153$  $R^2_{True AEV} = 0.0053$

$R^2_{AEV} = 0.0541$  $R^2_{True AEV} = 0.0972$

$R^2_{AEV} = 0.01265$  $R^2_{True AEV} = 0.1335$

$R^2_{AEV} = 0.0938$  $R^2_{True AEV} = 0.1772$

$R^2_{AEV} = 0.0642$  $R^2_{True AEV} = 0.1613$

$R^2_{AEV} = 0.0562$  $R^2_{True AEV} = 0.1489$

$R^2_{AEV} = 0.0280$  $R^2_{True AEV} = 0.0772$

$R^2_{AEV} = 0.0019$  $R^2_{True AEV} = 0.0149$
networks, at the end of the training process. The various architectures of ANNs are compared in terms of the RMSE given by:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}$$

(5)

In Fig. 13, the continuous line describes the result of the one-hidden-layer ANN, which is obviously uniquely determined by the number of neurons in the hidden layer itself. The dashed lines in this figure show the solutions for the two-hidden-layer ANNs. Each point of this graph has been calculated as the average of ten repetitions of the training process, to guarantee robustness against stochastic impacts. As shown in Fig. 13, the two-hidden-layer ANN with 5 and 23 neurons in the first and second layers provides the best performance. For cases with total number of neurons larger than the abovementioned values, there are no considerable changes in the RMSE. Therefore, the 7–5–23–1 ANN architecture is selected as the optimum ML model in this study.

In order to ensure the capabilities of the trained model to generalize its learning to unseen cases, and to prevent overfitting, only 70% of the database is used for the training process, 15% is kept for validation, and 15% is used for testing. Therefore, out of the total 644 data, 450 are used for training, 96 for validation, and 96 for testing.

Fig. 14 presents the comparison between the measured true AEVs with the ML predicted ones. It can be seen that not only the ML model exhibits a very high prediction accuracy for the training dataset, but also high-level prediction accuracies are
achieved for testing and validation datasets, with the $R^2$ values ranging from 0.851 to 0.964 across the three datasets.

Fig. 15 presents a comparison between the measured true AEVs for all cases in the UNSODA database and the predicted AEVs using the developed ML model. It is seen that the true AEVs estimated by the ML model match relatively well with their UNSODA counterparts; the insignificant scattering around the actual measured values are clearly in accordance with the results illustrated in Fig. 14.

6. Sensitivity analysis

To evaluate the effects of different soil parameters on soil’s AEV, using the developed model a sensitivity analysis is also performed according to the method proposed by Vu- Bac et al. (2016). A sensitivity, or uncertainty, analysis quantifies the impact of all input parameters with respect to the specific output of interest, which in this study are the AEV and the true AEV. Therefore, the graphically determined AEVs from cases in the UNSODA database have been assessed for this purpose. The sensitivities of the AEV and true AEV against different independent input parameters are shown in Fig. 16 confirming that the AEV is most sensitive to $P_{250}$, and $P_{500}$, respectively, and the true AEV is most sensitive to $P_{250}$ and $P_{50}$, respectively; therefore, these 2 parameters appear to have the most effect on soil’s air entry value. On the other hand, the bulk density and $P_{1000}$ appear to be the less influential parameters in terms of their effect on soil’s AEV.

7. Conclusions

In this paper, drawbacks in interpretation of water content-based SWRC and estimation of the AEV for a significant number of soil samples were investigated. The UNSODA database was used and thoroughly examined for this purpose. It was shown that there are 644 SWRCs in this database that possess the required information for the targeted analysis. For these SWRCs the AEVs were obtained using the conventional method and compared with the corresponding true AEVs obtained using a consistent graphical approach proposed by Pasha et al. (2016) which considers the stress history of the samples. For the cases in the database, the differences between the conventional AEVs and their corresponding true AEVs were found to be generally in the range of 30–300 kPa, although this could reach as high as 1500 kPa. A ML model was developed to estimate the true AEVs, with the bulk density and grain size distribution factors as input parameters. The obtained results revealed that the developed ML model operates with excellent accuracy and provides practical estimations of the true AEVs with $R^2$ values of 0.964, 0.901 and 0.851 for the training, validation and testing datasets, respectively. Furthermore, the sensitivity analysis showed that $P_{250}$ and $P_{50}$ are the most influential parameters on soil’s AEV. The study shows that the methodically trained ML approaches can be readily used for derivation of the true AEVs for a wide range of soils, provided appropriate information regarding the GSDs are available.

**CRediT authorship contribution statement**

Mohammad Sadegh Es-haghi: Conceptualization, Software, Validation, Investigation. Mohammad Rezania: Conceptualization, Methodology, Validation, Supervision, Writing – review & editing. Meghdad Bagheri: Conceptualization, Methodology, Validation, Writing – review & editing.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgements**

The second author acknowledges the financial support provided by the European Commission Research Fund for Coal and Steel (RFCS) for project MINRESCUE (Contract RFCS-RPJ-899518).

**References**


**ARTICLE IN PRESS**

M.S. Es-haghi, M. Rezania and M. Bagheri

**Gondwana Research xxx (xxxx) xxx**