# Predictive models for the residual saturation zone of the soil–water characteristic curve

Soner Uzundurukan[1]

## Abstract

**Purpose** Soil–water characteristic curve (SWCC) models have generally been developed based on measurement results in the low suction range. However, different mechanisms govern the SWCC in the low suction range (desaturation zone) and high suction range (residual zone). Accurate estimation of the residual zone is critical for analysing the behaviour of soils in arid regions. This study aimed to develop prediction models for the residual zone based on selected properties of soils.

**Materials and methods** To achieve the abovementioned purpose, 162 total suction (in pF) and gravimetric water content (%) measurement pairs in the high suction range were made on 40 cohesive soil samples with known physical, chemical, and spectral properties. A semi-logarithmic linear model was used to define the residual saturation zone of the SWCC. The model parameters were the slope of the SWCC ($s_r$) and total suction at zero water content ($\psi_{dry}$). Correlation and stepwise regression analyses were carried out between the model parameters and selected soil properties. The regression equations were validated using the four-fold cross-validation procedure. Water content (%) estimation models were developed using combinations of different regression equations for $s_r$ and $\psi_{dry}$, and their estimates were evaluated using performance metrics.

**Results and discussion** The $s_r$ values for the soils studied ranged from 1.589 to 13.035, with an average of 6.007. Although some studies have shown strong correlations between clay content and $s_r$, no significant relationship was found between clay content and $s_r$ for the soils in this study. However, significant correlations were found between the consistency limits, some spectral parameters, and $s_r$. The $R^2$ was 0.88 when the liquid limit (LL) and depth of the 1900 nm wavelength band ($D_{1900}$) values were used as descriptive variables. The $\psi_{dry}$ values for the soils studied ranged from 6.483 to 7.370 pF, with an average of 6.855 pF. The relationships between the selected soil properties and $\psi_{dry}$ were weak, consistent with previous research.

**Conclusion** The spectral absorption characteristics of the soils in this study had a high potential for estimating the SWCC. Prediction models based on various $s_r$ and $\psi_{dry}$ equation combinations could predict measured water content values with varying degrees of accuracy. The SWCC's residual saturation zone was accurately estimated using soil properties such as liquid limit, electrical conductivity, and spectral characteristics that can be determined quickly and inexpensively.

**Keywords** Multivariate linear stepwise regression · Residual saturation zone · Total suction · SWCC

## 1 Introduction

Swelling and collapsing upon wetting, water retention, and infiltration are vital behaviours of soils, especially in arid regions. Soil suction, which reflects the ability of the soil to attract and retain water, is a critical stress state variable that influences the strength, deformation, and hydraulic behaviour of unsaturated soils. The term total suction describes the combined effect of all the mechanisms that influence the water retention behaviour of soil. Total suction or total soil–water potential comprises several components, including matric, osmotic, gravitational, pneumatic, and piezometric potential. However, in the absence of externally applied gradients and under isothermal conditions, the matric and osmotic components are sufficient to describe the soil–water potential of unsaturated soils (Bulut and Wray 2005; Yong and Warkentin 1975; Yong 1999). Matric suction arises due

✉ Soner Uzundurukan
soneruzundurukan@sdu.edu.tr

1 Department of Civil Engineering, Suleyman Demirel
University, 32260 Isparta, Turkey

to adsorptive and capillary forces in the soil matrix, whereas osmotic suction is caused by salts or contaminants in the soil pore water. Since matric suction is the most significant component of total suction, changes in the total suction of soil are usually caused by variations in matric suction (Agus et al. 2001; Arifin and Schanz 2009; Nam et al. 2010; Ng and Menzies 2007; Malaya and Sreedeep 2012).

Soil suction can be measured directly or indirectly in the laboratory or the field using various techniques and instruments. In direct measurements, the water and air phases are separated by a ceramic cup or disk, and the negative pore water pressure is measured directly. The suction plate, pressure plate, and tensiometer methods are examples of the direct measurement of matric suction (Agus and Schanz 2005). Indirect techniques are generally based on measuring relative humidity in a climate with a moisture balance between the soil and the environment and estimating suction using the Kelvin equation. Indirect suction measurement techniques include the thermocouple, transistor, or chilled-mirror psychrometer methods, the filter paper method, the thermal conductivity sensor technique, the evaporation method (HYPROP), and the electrical conductivity sensor technique (Woodburn et al. 1993; Leong et al. 2003; Fredlund and Wong 1989; Likos and Lu 2002; Erzin 2007; Satyanaga et al. 2019). The suction of a given soil varies with its moisture conditions. Suction is zero in a fully saturated state but can exceed 1 GPa in a completely dry state. The soil–water characteristic curve (SWCC) describes the relationship between water content and suction in the region between the saturated and dry states for a given soil (Zhai et al. 2019). Each instrument or technique used to measure suction has some limitations in terms of reliability, practicality, or cost, and none of them can individually provide satisfactory measures for the entire suction or moisture range (Guan 1996; Vanapalli et al. 1999; Rahardjo and Leong 2006; Bulut and Leong 2008). The WP4-T dew point potentiometer provides satisfactory results in the high suction range region of the SWCC while reducing the time and costs associated with suction measurements (Ebrahimi-Birang and Fredlund 2016).

Several mathematical models (Brooks and Corey 1964; van Genuchten 1980; Williams et al. 1983; McKee and Bumb 1987; Fredlund and Xing 1994) have been developed to simulate the SWCC of a particular soil. On the other hand, these models have only been tested in certain soils and are only valid for specific suction ranges; for example, the Brooks and Corey, van Genuchten, McKee, and Bumb models are unsuitable for high suction ranges (Zapata 1999). Nam et al. (2010) observed that the Fredlund and Xing (1994) and van Genuchten (1980) models provide nearly identical SWCCs, except at high suction values. According to Lu and Khorshidi (2015), while many empirical SWCC models have been developed to cover the entire suction

range, matric suction data greater than 10 MPa is limited; thus, the validity of these models in the high suction range is uncertain.

On a semi-logarithmic scale, the SWCC is typically S-shaped and consists of three zones separated by specific limit values. These limits can be expressed as saturated state water content, air entry value (AEV), and residual water content. The region between the saturated water content and the AEV is referred to as the capillary saturation zone; the region between the AEV and the residual water content is referred to as the desaturation zone or transition zone; and the region between the residual water content and the completely dry state is referred to as the residual saturation zone. The fundamental mechanisms influencing the water retention behaviour in these three regions are distinct. In the transition zone (the region dominated by capillary water), pore water is retained primarily as capillary menisci located between soil particles; consequently, capillary forces are effective, and suction is mainly a function of pore size distribution (Lu and Likos 2006; Zhai et al. 2018, 2021). In the residual zone (the adsorbed water-dominant region), the majority of pore water is retained as hydration on the particle surfaces; adsorptive forces are more critical in this region, and the effect of soil structure on the SWCC is negligible (Lu and Likos 2006; Malaya and Sreedeep 2012; Fang et al. 2022). In other words, different mechanisms control the moisture–suction relationship of soils over different suction ranges. It is difficult to explain the entire suction range using a model that only focuses on capillary mechanisms and ignores soil physicochemical interactions, especially in cohesive soils (Guo et al. 2021). Salager et al. (2013) indicated that the retention curves corresponding to different densities of the same soil converge and become unique after a specific suction value is exceeded in the suction versus water content plane. Chen et al. (2022) also demonstrated that at high suction levels, the SWCCs in terms of water content versus initial void ratio are independent of the initial void ratio. It can be deduced that the region where the void ratio or density changes do not affect suction is located within the SWCC's residual zone.

The validity of most current prediction models for soils under low moisture conditions — typical for arid and semi-arid regions — is unclear. Models are needed to reliably estimate SWCC under these moisture conditions, particularly in cohesive soils. The present study attempted to develop prediction models for the residual saturation zone of the SWCC in cohesive soils based on the soils' physical, chemical, and spectral properties. For this purpose, 162 gravimetric water content and total suction (> 10 MPa) measurements for 40 different soils were investigated. The SWCC at large suctions was simulated using a semi-logarithmic linear model with two parameters: the slope of the residual saturation zone of the SWCC ($s_r$) and suction at zero water content

($\psi_{dry}$). Correlation and regression analyses were conducted between these parameters and several soil properties. As a result of these analyses, empirical estimation models for the SWCC residual saturation zone in cohesive soils were developed. The estimation capabilities of the proposed models were evaluated using different performance measures and then discussed.

## 2 Material and methods

### 2.1 Some physical, chemical, and spectral properties of soils

The current study utilised experimental data obtained from a comprehensive survey by Demirci (2010). Forty soil samples for which sufficient suction and water content measurement data existed to establish a reliable relationship at high suction levels (> 10 MPa) were selected for evaluation. These soils had plasticity index values ranging from 9.5 to 70.1 and fine contents ranging from 46.8 to 97.2%. The soil sampling sites were located in the Turkish cities of Tokat (S1 to S6), Sivas (S7), Mersin (S8), Adana (S9), Burdur (S10 and S18), Antalya (S15 and S16), Mugla (S11 to S14), Isparta (S17), Afyon (S19, S20, and S22), Kutahya (S21), Ankara (S23 to S31), Istanbul (S32 to S38), and Kirklareli (S39 and S40), as shown in Fig. 1.

The WP4-T dew point potentiometer was used to measure the total suction of the soils under specific moisture conditions following the ASTM D6836-02 D protocol. The ASTM D4318-10 standard was followed to determine the consistency limits of the soils. Mechanical sieving (following ASTM D422-63) and laser diffraction (Malvern Instruments Mastersizer 2000) were used to determine the grain size distributions of the soil samples. The specific surface area ($SSA_{LD}$) values were calculated from the grain size distributions using Malvern Mastersizer software. The surface area activity ($A_{SSA}$) values were calculated by dividing the specific surface area by the clay content. The free swell index (FSI) values of the soil samples were determined using the procedure described by Sridharan and Rao (1988). The carbonate content of the soil samples was determined using the method described by Loring and Rantala (1992). The electrical conductivity (EC) values were measured using a conductivity meter (Delta OHM, model HD 2106.1), and the pH values were measured with a pH/mV meter (Delta OHM HD 2105.1). The soils' cation exchange capacity (CEC) values were determined using the sodium acetate method (Bower et al. 1952). Following this method, the soil was washed with sodium acetate, soluble salts were extracted with ethyl alcohol, and the sodium produced by washing with ammonium acetate was quantified (Azadi and Baninemeh 2022). The specific surface charge (SSC) was

calculated as the ratio of CEC/$SSA_{LD}$ (Pesch et al. 2022). The reflectance properties of the soil samples were measured in the laboratory using an ASD field spectrometer. The spectral absorption feature characteristics (asymmetry and depth of an absorption band) were determined according to procedures outlined by van der Meer (2004). The above-mentioned textural, physical, and chemical soil properties of cohesive soils were used as explanatory variables to develop prediction equations that define the residual saturation zone of the SWCC. Table 1 presents the descriptive statistics, such as the standard deviation, minimum and maximum values, and quartiles for some of the textural, physical, and chemical properties of the 40 cohesive soils examined in this study.

### 2.2 Statistical analysis

Multiple linear regression is an analysis method that uses observed data to develop a prediction model for a dependent variable by describing the relationship between the dependent variable and explanatory variables using a linear equation. This study used stepwise multiple linear regression to develop prediction models for the residual saturation zone. Stepwise regression is an iterative procedure in which the possible explanatory variables to be used in a prediction model are added to or removed from the model in each iteration based on the test statistics and statistical significance of the coefficients (Hosseinpour et al. 2018). Prediction models were developed based on the results of stepwise regression analysis, as shown in Eq. 1:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \cdots + \beta_n \cdot X_n \qquad (1)$$

where $Y$ is the dependent variable; $\beta_0, \beta_1,\ldots, \beta_n$ are regression coefficients; and $X_1, X_2, \ldots,X_n$ are explanatory variables.

One of the most essential aspects of the predictive stability of a regression model is the presence of multicollinearity. A high correlation between two or more explanatory variables is defined as multicollinearity. Multicollinearity between explanatory variables in a regression model can lead to unstable model estimates. The variance inflation factor (VIF) is a common indicator used to assess the multicollinearity of explanatory variables in a model. A value of 1 indicates that the predictive variables are independent of one another, while a value greater than 5 indicates that the variables are highly correlated (Daoud 2017; Shrestha 2020). In this study, the upper limit for the VIF value was set at 3.

Validation of a regression model is another crucial aspect of predictive stability. The two main methods for validating regression models are external validation and internal validation (cross-validation). External validation is achieved by collecting, analysing, and comparing new and previous data (Lucko et al. 2006). Cross-validation is a variable selection
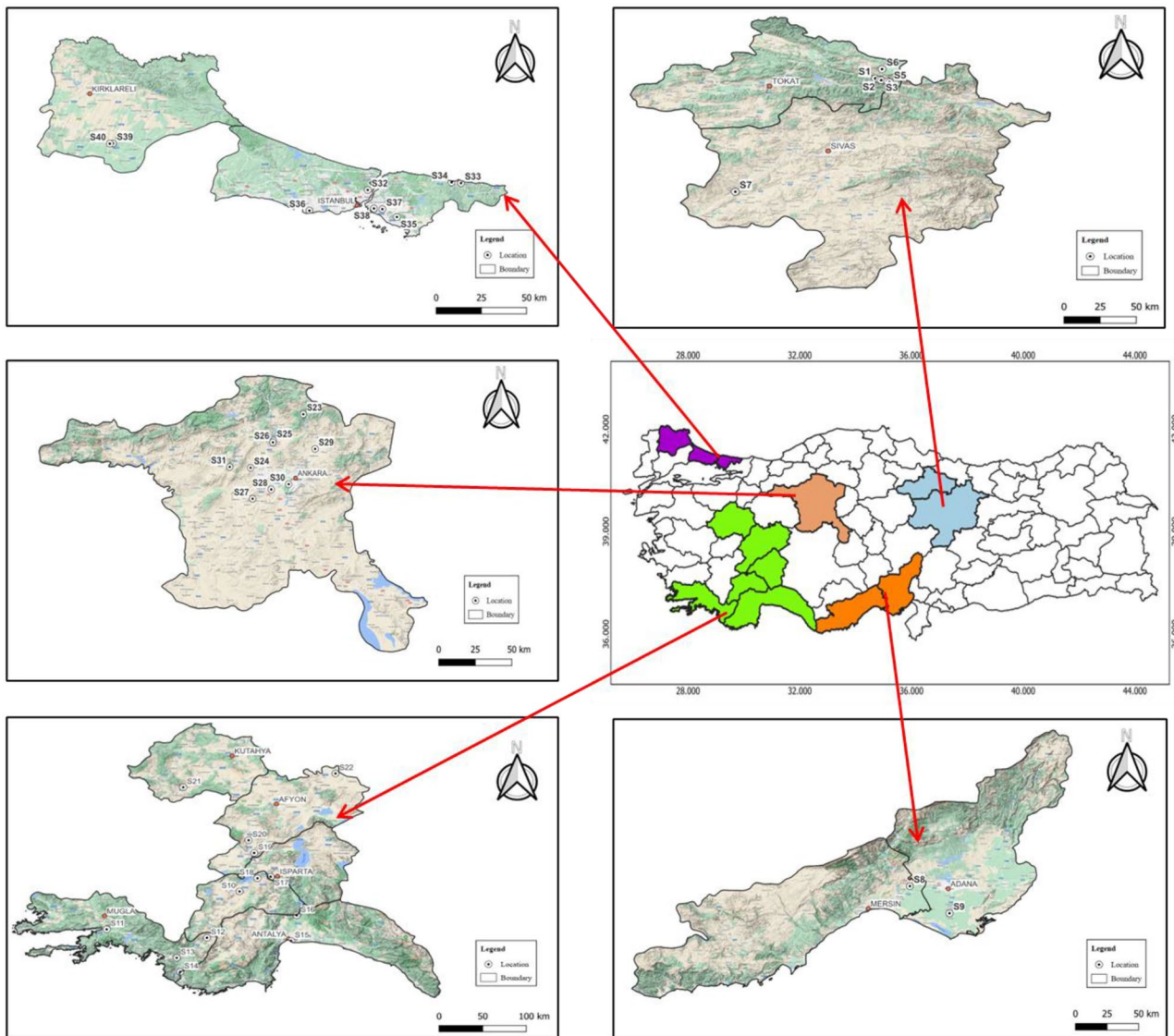
**Fig. 1** Sampling sites of soils

technique used to evaluate prediction models with a small data sample, and it is an effective method of model validation when collecting new data is impractical (Snee 1977). In this study, cross-validation of the regression models was performed using the K-fold procedure. The dataset was divided into four random and equal parts. A candidate model was trained on three subsets of the data, and its prediction performance was evaluated on a test set containing the remaining data. The model training and testing procedure was repeated to provide each of the four parts as a test set. The results from all four processes were combined to create a single validation set, and the cross-validated performance of the predictive model was evaluated by the performance measurements on the validation set (Jung and Hu 2015; Li et al. 2020; Liu et al. 2021). SPSS (version 25.0) software was used to perform the statistical analysis.

## 2.3 Performance metrics

In this study, the predictive performance of models was evaluated using six distinct metrics: the coefficient of determination ($R^2$), the mean absolute error (MAE), the root mean square error (RMSE), the mean absolute percentage error (MAPE), the ratio of performance to deviation (RPD), and the ratio of performance to interquartile range (RPIQ). These metrics were calculated using the following equations:

**Table 1** Descriptive statistics of selected textural, physical, and chemical properties of soil samples

| Soil property (unit) | Min | Q1 | Q2 | Q3 | Max | SD |
|---|---|---|---|---|---|---|
| Sand (%) | 2.8 | 13.4 | 20.4 | 37.3 | 52.0 | 14.8 |
| Silt (%) | 26.1 | 40.1 | 50.9 | 54.4 | 69.5 | 10.6 |
| Clay (%) | 9.8 | 17.3 | 24.8 | 33.7 | 50.0 | 10.8 |
| d10 (mm) | 0.00072 | 0.00140 | 0.00165 | 0.00235 | 0.00530 | 0.00094 |
| d30 (mm) | 0.00170 | 0.00338 | 0.00465 | 0.00805 | 0.01700 | 0.00364 |
| d60 (mm) | 0.00600 | 0.00943 | 0.02000 | 0.05575 | 0.30000 | 0.06736 |
| LL (%) | 27.3 | 44.2 | 56.0 | 65.2 | 106.0 | 19.2 |
| PL (%) | 13.9 | 21.3 | 25.6 | 32.3 | 44.7 | 7.6 |
| SL (%) | 11.3 | 13.4 | 15.5 | 17.4 | 31.2 | 3.8 |
| R | 1.292 | 1.508 | 1.598 | 1.667 | 1.900 | 0.128 |
| $G_S$ | 2.593 | 2.705 | 2.730 | 2.753 | 2.822 | 0.047 |
| A | 0.53 | 0.80 | 1.05 | 1.41 | 4.61 | 0.77 |
| SSA (gr/ m$^2$) | 75.54 | 129.89 | 166.01 | 198.72 | 310.44 | 52.91 |
| $A_{SSA}$ | 4.78 | 5.57 | 6.63 | 8.26 | 13.30 | 1.89 |
| FSI (cm$^3$/gr) | 1.00 | 1.20 | 1.28 | 1.50 | 4.30 | 0.56 |
| CaCO$_3$ (%) | 0.00 | 1.48 | 8.66 | 21.65 | 71.12 | 16.78 |
| pH | 6.50 | 7.88 | 8.14 | 8.47 | 9.60 | 0.62 |
| EC (mS/m) | 34.00 | 115.03 | 183.05 | 311.50 | 1807.00 | 319.22 |
| CEC (meq/100gr) | 0.68 | 7.34 | 14.41 | 18.15 | 56.42 | 10.50 |
| CEA | 0.04 | 0.31 | 0.51 | 0.74 | 5.39 | 0.85 |
| SSC (meq/100m$^2$) | 0.00483 | 0.05119 | 0.07229 | 0.12383 | 0.58500 | 0.09232 |
| $S_{1400}$ | 0.37666 | 0.44934 | 0.5294 | 0.61546 | 0.96441 | 0.14014 |
| $D_{1400}$ | 0.05099 | 0.07756 | 0.11152 | 0.13861 | 0.23717 | 0.04593 |
| $S_{1900}$ | 0.28085 | 0.30233 | 0.31693 | 0.33631 | 0.67494 | 0.06087 |
| $D_{1900}$ | 0.01381 | 0.15181 | 0.25091 | 0.29497 | 0.44744 | 0.09659 |
| $S_{2200}$ | 0.42234 | 0.74133 | 0.803 | 0.85736 | 1.61396 | 0.17274 |
| $D_{2200}$ | 0.0087 | 0.10091 | 0.13336 | 0.15397 | 0.27058 | 0.04488 |
| $\Psi$ (pF) ($N=162$) | 5.003 | 6.629 | 5.421 | 5.745 | 6.131 | 0.439 |
| w (%) (N=162) | 0.820 | 17.353 | 3.019 | 6.017 | 8.835 | 3.686 |

$d_{10}$, $d_{30}$, and $d_{60}$ are diameters through which 10%, 30%, and 60%, respectively, of the total soil mass passes. LL, PL, and SL are liquid, plastic, and shrinkage limits, respectively. R is shrinkage ratio. $G_S$ is specific gravity. A is activity. $S_{1400}$, $S_{1900}$, and $S_{2200}$ are SWIR spectral parameters that correspond to asymmetry factors at the 1400 nm, 1900 nm, and 2200 nm wavelength bands, respectively. $D_{1400}$, $D_{1900}$, and $D_{2200}$ are SWIR spectral parameters that correspond to depths of 1400 nm, 1900 nm, and 2200 nm wavelength bands, respectively. SD is the standard deviation

$$R^2 = \left( \frac{\sum \left( Y_i - \overline{Y_i} \right)\left( Y_i^* - \overline{Y_i^*} \right)}{\sqrt{\sum \left( Y_i - \overline{Y_i} \right)^2 \sum \left( Y_i^* - \overline{Y_i^*} \right)^2}} \right)^2 \quad (2)$$

where $Y_i$ is the actual value of the dependent variable, $\overline{Y_i}$ is the average of the actual values of the dependent variable, $Y_i^*$ is the predicted value of the dependent variable, and $\overline{Y_i^*}$ is the average of the predicted values of the dependent variable.

$$MAE = \frac{\sum \left| Y_i - Y_i^* \right|}{n} \quad (3)$$

where $n$ is the number of observations.

$$RMSE = \sqrt{\frac{\sum \left( Y_i - Y_i^* \right)^2}{n}} \quad (4)$$

$$MAPE = \frac{\sum \left| \frac{Y_i - Y_i^*}{Y_i} \right|}{n} \quad (5)$$

$$RPD = \frac{SD}{RMSE} \quad (6)$$

$$RPIQ = \frac{IQ}{RMSE} = \frac{Q_3 - Q_1}{RMSE} \quad (7)$$

The MAE determines the magnitude of the mean errors between the predicted and actual values, regardless of the direction of the errors. In the case of large error values, the model's prediction performance may not be accurately reflected. Unlike the MAE, which provides equal weight to all errors, the RMSE accounts for variance by giving greater weight to larger error values. The MAPE helps evaluate the performance of predictive models since it is independent of the sizes of the observed and anticipated values. Lewis (1982) categorises a model's predictive capacity based on its MAPE values as highly accurate forecasting (10%), good forecasting (10 to 20%), reasonable forecasting (20 to 50%), or inaccurate forecasting (> 50%). The RPD is another metric used to assess prediction performance; higher values indicate the model's robustness. If the RPD value is less than 1.5, the model lacks the ability to predict; if it is between 1.5 and 2, the model can distinguish between high and low values; if it is between 2 and 2.5, the model can be used for quantitative predictions; if it is between 2.5 and 3, the model has good estimation ability; and if it is greater than 3, the model has excellent predictive ability (Saeys et al. 2005). Ludwig et al. (2017) adapted the classification developed by Saeys et al. (2005) for RPD values to categorise the prediction performance of a model based on RPIQ values. According to Ludwig et al. (2017), if the RPIQ value is between 2.02 and 2.70, the model can distinguish between high and low values; if it is between 2.70 and 3.37, the model provides approximate quantitative predictions; if it is between 3.37 and 4.05, the model has good estimation ability; and if it is greater than 4.05, the model has excellent predictive power.

The Wilcoxon signed-rank sum test was employed in addition to the performance indicators mentioned above to evaluate the statistical significance of the differences between the actual and predicted values. The Wilcoxon signed-rank test is a non-parametric test used to determine whether two populations differ statistically based on rank. The following equation can be used to calculate the Wilcoxon $z$-score:

$$z = \frac{\max\left(R^-, R^+\right) - \left(\frac{nn(+1)}{4}\right)}{\sqrt{\frac{nn(+1)(2n+1)}{24}}} \tag{8}$$

where $R^+$ and $R^-$ are the sums of the ranks of the positive paired difference and negative paired difference $(S_i - S_i^*)$, respectively.

A two-tailed test is applied to decide if the difference between the actual and predicted values is statistically significant. Since the critical $z$ value for the 95% confidence interval (or a 5% significance level) in a two-tailed test is 1.96, the null hypothesis of no difference is rejected if the absolute value of the calculated $z$ value is greater than

1.96 or the $p$-value is less than the significance level. On the other hand, it is assumed that there is no significant difference between the actual and predicted values if the $z$-score is less than 1.96 or the $p$-value is greater than 0.05 (Uzundurukan 2023).

## 2.4 Model used to describe the residual saturation zone of SWCC

In the current study, the residual saturation zone of SWCC was defined using the linear relationship expressed in Eq. 9 between total suction ($\psi$) values (in pF) and gravimetric water contents ($w$).

$$w = -a \cdot \psi + b \tag{9}$$

where $a$ and $b$ are the equation coefficients of the relationship.

The equation coefficients of the relationships between total suction (in pF) and gravimetric water content (%) have specific values for each soil included in this study. The assessment of equation coefficients for all soils demonstrates a linear proportional relationship between coefficients $a$ and $b$. Thus, Eq. 10 can be written as follows:

$$w = a \cdot \left(^b/_a - \psi\right) \tag{10}$$

where $a$ represents the slope of SWCC in the residual saturation zone ($s_r$) and $^b/_a$ represents total suction at zero water content ($\psi_{dry}$) of a particular soil. As a result, the expression in Eq. 9 can be transformed into the following form, which gives the gravimetric water content ($w_i$) corresponding to a specific suction value ($\psi_i$) within the boundaries of the residual saturation zone:

$$w_i = s_r \cdot \left(\psi_{dry} - \psi_i\right) \tag{11}$$

# 3 Results and discussions

## 3.1 Obtaining the model parameters for the residual saturation zone of the SWCC

Total suction and gravimetric water content datasets with suction values larger than 10 MPa (> 5 pF) were used to develop a model defining the residual saturation zone of SWCC. The linear relationship expressed by Eq. 9 was obtained between the total suction (in pF) and the gravimetric water content (%) for each of the soils examined. It was observed that the equation coefficients of this relationship had specific values for each soil. Figure 2 provides an example of the relationship obtained for the S24 sample.
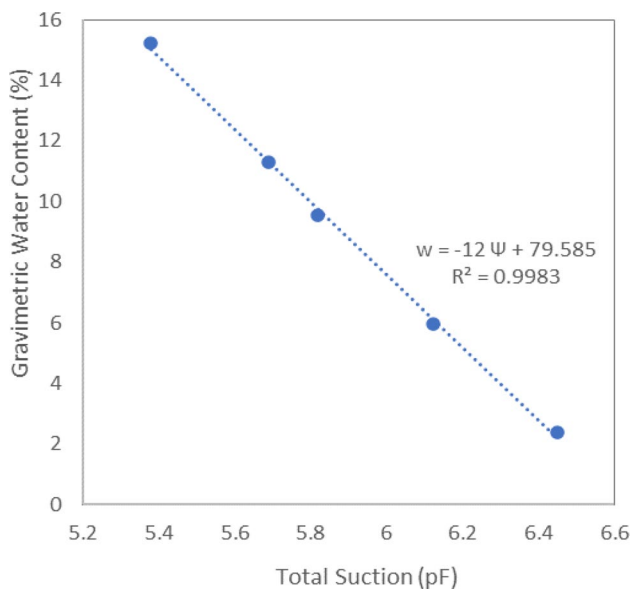
**Fig. 2** Gravimetric water content (%)–total suction (in pF) relationship (for S24 soil sample)

The coefficients a and b, both of which have a specific value for each soil, were observed to have a linear proportional relationship (Fig. 3).

Figure 3 demonstrates that the ratio b/a produces an approximate value applicable to all soils evaluated in the study. This ratio represents the suction at zero water content ($\psi_{dry}$), while the coefficient a indicates the slope of the soil–water characteristic curve ($s_r$) in the residual saturation zone. Consequently, Eq. 11 was derived by making appropriate modifications to Eq. 9, resulting in a model that properly
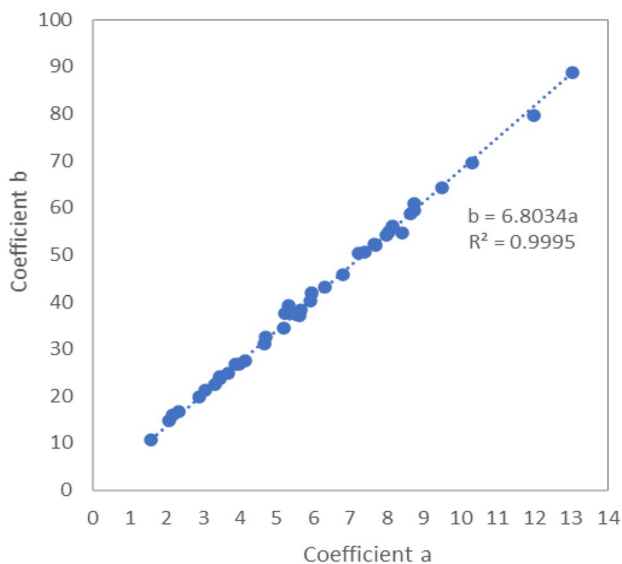


**Fig. 3** Relationship between the a and b coefficients for all soil samples investigated in the study

reflects the behaviour of the SWCC within the residual saturation zone. It must be noted that Eq. 11 presents another representation of the semi-log-linear function initially developed by Campbell and Shiozawa (1992) to describe the relationship between water content and suction in soil.

Table 2 provides descriptive statistics for the $s_r$ and $\psi_{dry}$ values obtained from the 40 soils examined in this study.

The $s_r$ values of the investigated soil samples were between 1.589 and 13.035. The values corresponding to the dimensionless slope $SL$ ($^1/_{s_r} \cdot 100$) of the Campbell–Shiozawa model can be calculated as 62.93 and 7.67, respectively. Resurreccion et al. (2011) reported $SL$ values between 19.2 and 276.92 for 41 illite-predominant clay samples. Chen et al. (2014) determined that the $SL$ values of 24 clayey soils ranged from 7.6 to 135.6. Pittaki-Chrysodonta et al. (2019) investigated 144 different soils with $SL$ values ranging between 16.71 and 165.54.

The $\psi_{dry}$ values obtained for the soils examined in this study ranged from 6.483 to 7.370 pF, with an average of 6.855 pF. Although previous research has assumed that the suction at zero water content can reach up to 1 GPa (or 7 pF), the exact limits have not been defined. Several studies have considered a value ranging from 6.7 to 7.1 for $\psi_{dry}$ (Ross et al. 1991; Rossi and Nimmo 1994; Fayer and Simmons 1995; Webb 2000; Groenevelt and Grant 2004; Schneider and Goss 2012; Lu and Khorshidi 2015). Karup et al. (2017) revealed that the matric potential at zero water content varies between 6.65 and 7.1, even in soils with similar mineralogy, in their study of 171 undisturbed soil samples.

Consequently, the $s_r$ and $\psi_{dry}$ values obtained from the semi-logarithmic linear function for the soils examined in this study are consistent with data previously reported in the literature.

### 3.2 Correlations between the model parameters and the soil properties

Spearman's rho analyses were performed to determine whether the $s_r$ and $\psi_{dry}$ values are related to the soil

**Table 2** Descriptive statistics for $s_r$ and $\psi_{dry}$ values for the investigated soils

| Statistics | $s_r$ | $\psi_{dry}$(in pF) |
|---|---|---|
| Min | 1.589 | 6.483 |
| Q1 | 3.835 | 6.763 |
| Q2 (median) | 5.633 | 6.807 |
| Q3 | 7.999 | 6.922 |
| Max | 13.035 | 7.370 |
| Average | 6.007 | 6.855 |
| Std. Dev | 2.735 | 0.187 |
| Kurtosis | 0.517 | 1.012 |
| Skewness | −0.084 | 1.313 |

properties selected for the study, and the statistical significance of the correlations was determined. The significance levels of the correlations and their corresponding correlation coefficients are shown in Table 3.

Based on WP4T measurements on 41 Danish soils, Resurreccion et al. (2011) found strong correlations between the slope $SL$ of the Campbel and Shiozawa (1992) model and the clay content and specific surface area. Arthur et al. (2013) and Schneider and Goss (2012) also observed significant relationships between $1/SL$ and clay content. However, this study observed no significant relationship between the slope values and the clay content. According to Spearman's rho analyses, there are significant correlations ($p < 0.01$) between $S_r$ and the liquid limit, the

**Table 3** Results of Spearman's Rho correlation analysis

| Soil properties | Spearman's Rho correlation analysis | | | |
| --- | --- | --- | --- | --- |
| | $s_r$ | | $\psi_{dry}$ | |
| | r | p | r | p |
| Sand % | −0.305 | 0.056 | 0.431** | 0.005 |
| Silt % | 0.209 | 0.196 | −0.410** | 0.009 |
| Clay % | 0.162 | 0.316 | −0.266 | 0.098 |
| d10 | −0.097 | 0.551 | 0.141 | 0.384 |
| d30 | −0.198 | 0.220 | 0.269 | 0.093 |
| d60 | −0.367* | 0.020 | 0.313* | 0.050 |
| LL | 0.888** | 0.000 | −0.401* | 0.010 |
| PL | 0.820** | 0.000 | −0.165 | 0.310 |
| SL | −0.118 | 0.468 | −0.015 | 0.927 |
| R | −.878** | 0.000 | 0.353* | 0.025 |
| $G_S$ | −0.073 | 0.656 | 0.079 | 0.629 |
| A | 0.572** | 0.000 | −0.231 | 0.152 |
| SSA | −0.052 | 0.751 | −0.129 | 0.427 |
| SAA | −0.311 | 0.051 | 0.409** | 0.009 |
| FSI | 0.757** | 0.000 | −0.544** | 0.000 |
| $CaCO_3$ | −0.096 | 0.555 | 0.061 | 0.710 |
| pH | 0.247 | 0.124 | 0.008 | 0.961 |
| Ec | 0.393* | 0.012 | −0.381* | 0.015 |
| CEC | 0.294 | 0.066 | 0.133 | 0.413 |
| CEA | 0.160 | 0.323 | 0.221 | 0.170 |
| SSC | 0.307 | 0.054 | 0.170 | 0.293 |
| S1400 | −0.258 | 0.108 | −0.376* | 0.017 |
| D1400 | 0.625** | 0.000 | −0.004 | 0.980 |
| S1900 | −0.034 | 0.833 | −0.123 | 0.450 |
| D1900 | 0.680** | 0.000 | 0.040 | 0.807 |
| S2200 | −0.271 | 0.090 | −0.254 | 0.114 |
| D2200 | 0.256 | 0.111 | −0.232 | 0.151 |

*$p < 0.05$; **$p < 0.01$

plastic limit, the shrinkage ratio, the activity, the free swell index, and the SWIR spectral parameters $D_{1400}$ and $D_{1900}$. According to Pittaki-Chrysodonta et al. (2019), higher absolute values of $SL$ indicate sandier soils, while lower absolute values indicate clayey soils. However, examining Table 3 demonstrates that defining this judgement based on consistency limits rather than clay content is more significant because consistency limits are controlled by the mineral composition, particle size distribution, and pore fluid chemistry, which govern the physicochemical forces that exist between soil grains and soil water (Zhou and Lu 2021). Therefore, even if it contains fewer clay particles, soil with higher plasticity and activity may have larger suction under the same humidity conditions than one with lower plasticity (Kocaman et al. 2022).

Several studies have been conducted to determine the relationship between spectral adsorption features and the physical and chemical properties of soils (Escadafal 1993; Kariuki et al. 2003; Kariuki et al. 2004; Viscarra Rossel et al. 2006; Ben-Dor et al. 2009; Mulder et al. 2013; Dufréchou et al. 2015; Zhou et al. 2022; Taghdis et al. 2022). In most of these studies, geometric parameters, such as band position, band depth, band width, band area, and asymmetry, have been used to express spectral absorption properties (van der Meer 1999; Byun et al. 2023). Some studies have explored the relationships between water retention characteristics and spectral soil properties (Janik et al. 2007; Santra et al. 2009; Babaeian et al. 2015; Pittaki-Chrysodonta et al. 2021), but studies on the residual saturation zone remain limited (Pittaki-Chrysodonta et al. 2019; Norouzi et al. 2022). The spectral parameters of soil are strongly affected by clay mineralogy, and the presence of kaolinite and smectite significantly affects the intensity of the 1900 nm and 2200 nm features (Kariuki et al. 2003, 2004). Santra et al. (2009) discovered a strong correlation between basic soil and hydraulic properties and spectral reflectance in the 2000–2300 nm wavelength bands. Babaeian et al. (2015) stated that wavelength bands in the NIR-SWIR (700–2500 nm) region had predictive capacity in their study, which aimed to estimate Mualem–van Genuchten model parameters using spectral reflectance data. Pittaki-Chrysodonta et al. (2021) indicated that visible near-infrared spectroscopy can effectively estimate the soil water retention curve. Consequently, the strong relationship between $s_r$ and the SWIR spectral parameters $D_{1400}$ and $D_{1900}$ observed in this study is consistent with the literature.

Karup et al. (2017) stated that they could not find a distinct relationship between the matric potential at zero water content and clay mineralogy, organic matter, or clay content. In this study, the relationships between $\psi_{dry}$ and the

**Table 4** The regression equations for $s_r$

| Eq. ID | Regression equation | $R^2$ | Adjusted $R^2$ | Std. error of the estimate |
|---|---|---|---|---|
| E1 | $s_r = -0.9455 + 0.1228LL$ | 0.744 | 0.737 | 1.402 |
| E2 | $s_r = -2.1803 + 0.0945LL + 11.9523D_{1900}$ | 0.883 | 0.876 | 0.962 |
| E3 | $s_r = -2.3963 + 0.0971LL + 10.459D_{1900} + 4.5676SSC$ | 0.904 | 0.896 | 0.881 |
| E4 | $s_r = -1.9599 + 0.0976LL + 9.8193D_{1900} + 4.3509SSC - 0.0202CaCO_3$ | 0.919 | 0.910 | 0.822 |

selected soil properties are not as strong as those observed for $S_r$. The total suction at zero water content was observed to have weak correlations with the free swelling index, particle fractions, surface area activity, liquid limit, and electrical conductivity.

### 3.3 Regression analysis between model parameters and soil properties

Following the correlation analysis between the selected soil properties and the $S_r$ and $\psi_{dry}$ values, the goal was to develop regression models for the $s_r$ and $\psi_{dry}$ values. Multiple linear regression analyses were performed using IBM SPSS Statistic Data Editor v25 and the stepwise regression method. Table 4 reveals the linear regression equations obtained between $s_r$ and the selected soil properties. Table 5 presents the regression equations for $\psi_{dry}$.

Assessing Tables 4 and 5, it is evident that the coefficients of determination of the regression equations for $s_r$ values are acceptable, whereas those for $\psi_{dry}$ are relatively low. A four-fold cross-validation procedure was used to validate the regression models. The dataset was randomly divided into four equal parts, and the model training and testing procedure was repeated four times, with each of the four parts serving as a test set. The results of all four processes were combined to form a single validation set, and the predictive model's performance was assessed using validation set performance measures. Figure 4 compares the performance metrics obtained for each fold, the combined validation set, and the regression set for $s_r$.

The predictive performance comparison for $\psi_{dry}$ is given in Fig. 5.

Figure 4 shows that, except for E1, all other regression equations can be used to accurately predict $s_r$ for the soils examined. The calculated performance criteria, on the other hand, show that E1's predictive ability is acceptable.

The values of the performance criteria of the regression equations used for the $\psi_{dry}$ predictions are, as expected, lower than those obtained for $s_r$. Based on the calculated performance criteria, it can be concluded that the regression equations obtained for $\psi_{dry}$ cannot provide reliable predictions. The Wilcoxon signed-rank sum test was performed to determine whether the regression equations produced statistically significant results by comparing the predicted values of $s_r$ and $\psi_{dry}$ with their corresponding actual values. Table 6 summarises the results.

The Wilcoxon signed-rank sum test results revealed that the E5 equation used for $\psi_{dry}$ estimation did not produce statistically significant results. Consequently, this equation was not used as a part of the prediction models in the next part of the study.

Combinations of $s_r$ and $\psi_{dry}$ values predicted by different regression equations were substituted into Eq. 11, and the actual water content values of the 162 measured $w_i - \psi_i$ data pairs were compared to the predicted values. Table 7 shows the prediction performance for all predictive model combinations. Table 7 also shows the prediction performances for cases where the average value obtained for the soils studied (6.855 pF) was used instead of the suction value at zero water content. The most accurate predictions

**Table 5** The regression equations for $\psi_{dry}$

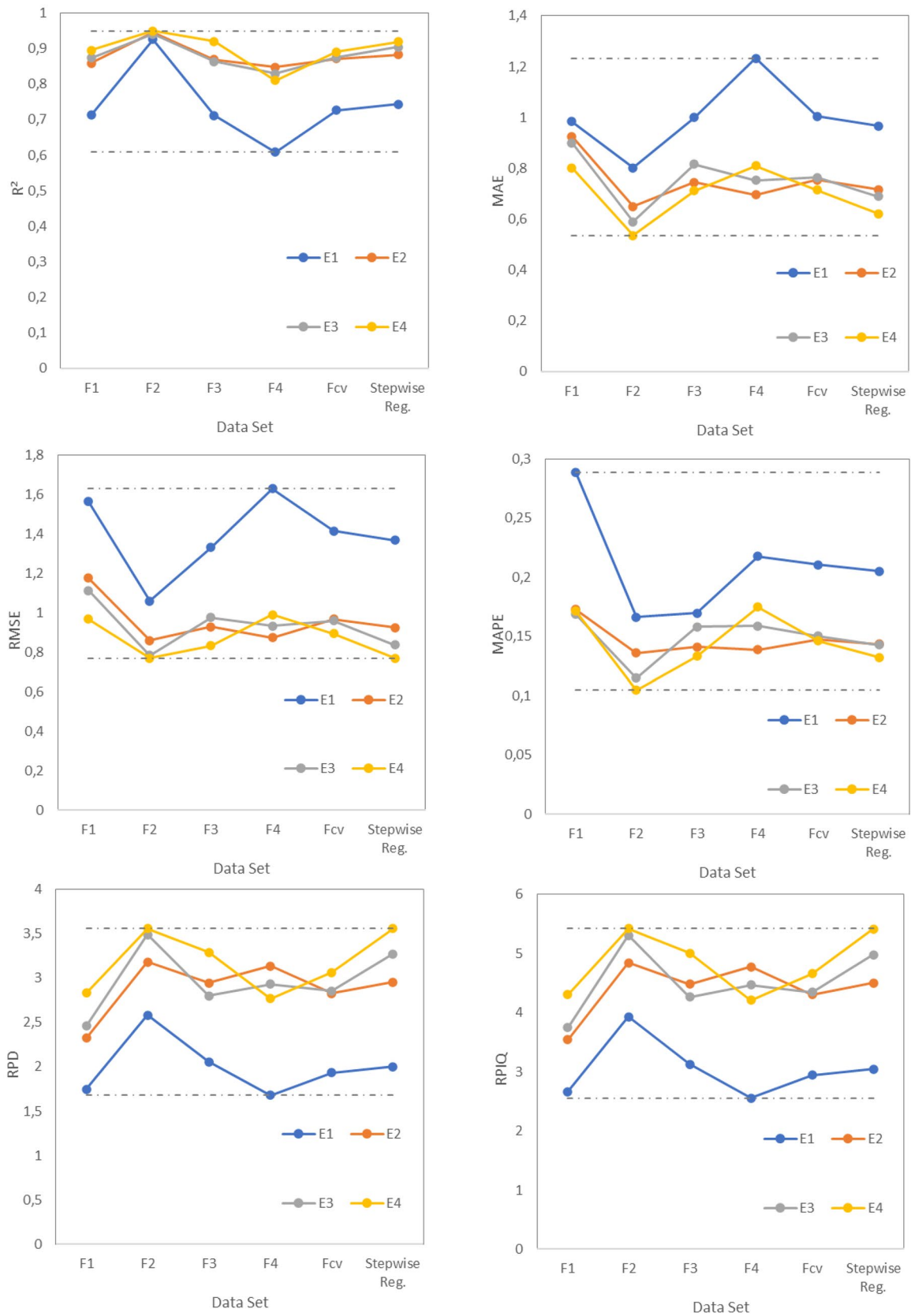| Eq ID | Regression equation | $R^2$ | Adjusted $R^2$ | Std. error of the estimate |
|---|---|---|---|---|
| E5 | $\psi_{dry} = 6.9308 - 0.0003EC$ | 0.219 | 0.198 | 0.167 |
| E6 | $\psi_{dry} = 7.1177 - 0.0003EC - 1.4047D_{2200}$ | 0.332 | 0.295 | 0.157 |
| E7 | $\psi_{dry} = 6.401 - 0.0002EC - 1.3076D_{2200} + 0.4287R$ | 0.403 | 0.353 | 0.150 |
| E8 | $\psi_{dry} = 6.3462 - 0.0002EC - 0.9566D_{2200} + 0.5882R - 0.4664S_{1400}$ | 0.507 | 0.451 | 0.138 |

**Fig. 4** Variations of the performance metrics according to datasets and regression equations for $s_r$
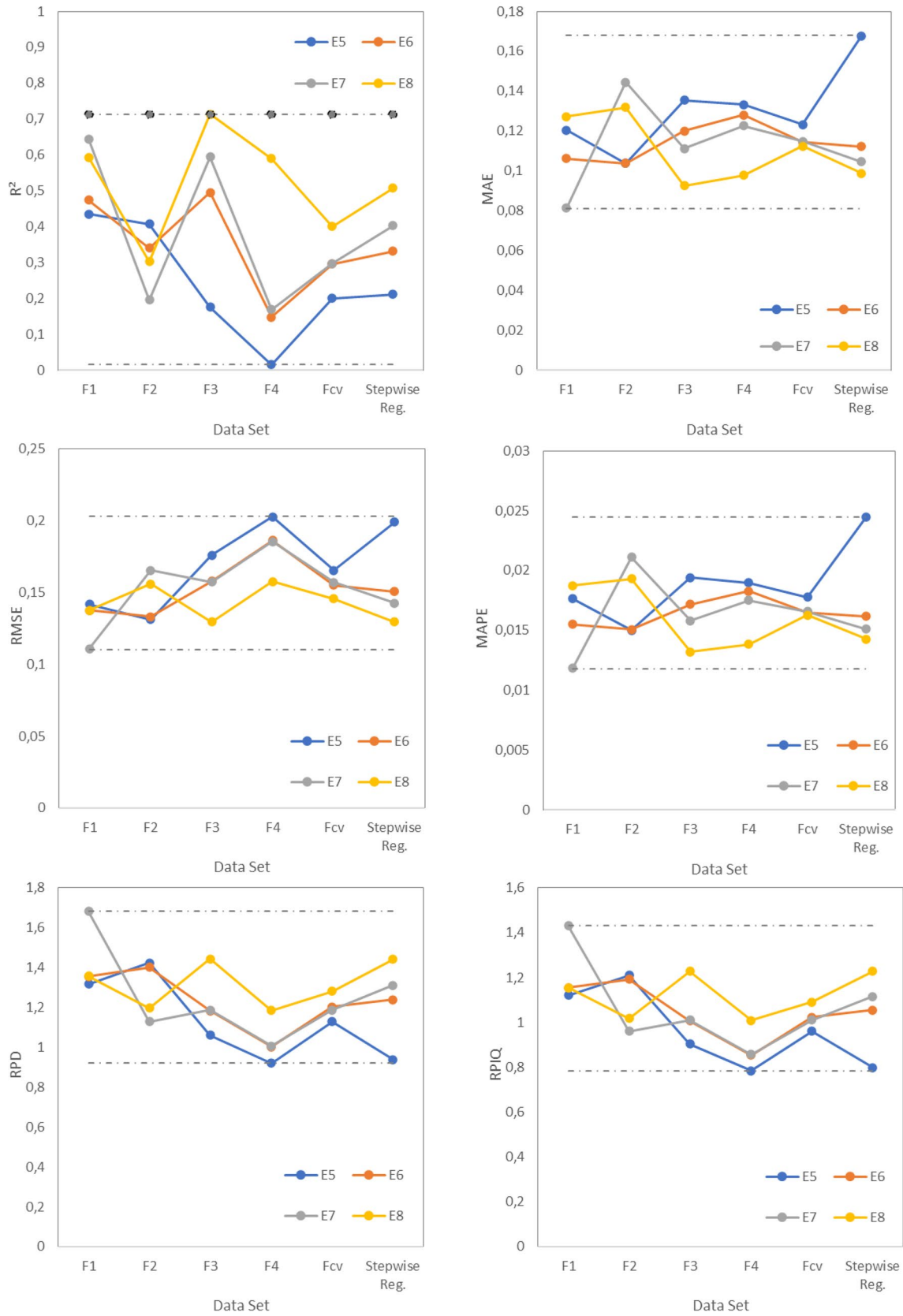
**Fig. 5** Variations of the performance metrics according to datasets and regression equations $\psi_{dry}$

**Table 6** Wilcoxon signed rank sum test result for actual values vs. predicted values

| Dependent variable | Equation no | Wilcoxon signed rank sum test (%95 confidence interval) | |
|---|---|---|---|
| | | Z | p |
| $s_r$ | E1 | −0.175 | 0.861 |
| | E2 | −0.161 | 0.872 |
| | E3 | −0.148 | 0.882 |
| | E4 | −0.175 | 0.861 |
| $\psi_{dry}$ | E5 | **−2.796*** | **0.005**** |
| | E6 | −0.780 | 0.436 |
| | E7 | −0.726 | 0.468 |
| | E8 | −0.390 | 0.697 |

*$z > 1.96$; **$p < 0.05$

were provided by the P13 model, which employs the E4 equation for $s_r$ estimations and the E6 equation for $\psi_{dry}$ estimations. The P4 model, which uses the E1 equation for $s_r$ estimates and the average value of 6.855 pF for $\psi_{dry}$, had the worst predictive accuracy. Figure 6 compares the water content values calculated from the models with the best and worst predictive performances with the measured water content values.

Figure 7 depicts a visual comparison of the performances of the prediction models for the threshold values given for the performance criteria in the literature.

As shown in Table 7 and Fig. 7, the regression equation selected to predict the slope of the residual saturation

zone of the SWCC has the greatest influence on the prediction performance of different models. Models employing regression equations other than the E1 equation, which employs only the liquid limit as an explanatory variable to predict $s_r$, provide acceptable estimates. Models that use the E3 and E4 regression equations to predict $s_r$ have the best predictive performance. However, using the E2 regression equations to predict $s_r$ yields good predictions. The regression equation used to predict suction at zero water content has a limited effect on the prediction performance of the models. Even in models employing the average value of 6.855 pF obtained for the soils examined in this study without any estimation equations for $\psi_{dry}$, acceptable estimations are achieved. For the P8 model, for example, knowing the liquid limit and the $D_{1900}$ values, which are inexpensive and easily identifiable, is sufficient. The performance metrics demonstrate that the P8 model makes good predictions for the soils under consideration in the study.

## 3.4 Limitations of the study

The study's results were obtained from the analysis of experimental data from 40 cohesive soils with plasticity index values between 9.5 and 70.1 and fine content between 46.8 and 97.2%. The suction–water content measurements in the high suction range (> 10 MPa) were considered in line with the study's objectives. Consequently, the models proposed in this study can only be used to predict the residual saturation zone of the SWCC in cohesive soils.

**Table 7** Prediction performances of different prediction models

| Model no | Prediction equation | | Performance criteria | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $s_r$ | $\psi_{dry}$ | $R^2$ | MAE | RMSE | MAPE | RPD | RPIQ |
| P1 | E1 | E6 | 0.78 | 1.29 | 1.72 | 0.26 | 2.15 | 3.32 |
| P2 | E1 | E7 | 0.76 | 1.33 | 1.79 | 0.26 | 2.06 | 3.18 |
| P3 | E1 | E8 | 0.79 | 1.22 | 1.69 | 0.24 | 2.19 | 3.37 |
| P4 | E1 | Average | 0.69 | 1.45 | 2.12 | 0.29 | 1.74 | 2.68 |
| P5 | E2 | E6 | 0.89 | 0.93 | 1.22 | 0.19 | 3.04 | 4.67 |
| P6 | E2 | E7 | 0.88 | 0.92 | 1.26 | 0.18 | 2.93 | 4.51 |
| P7 | E2 | E8 | 0.88 | 0.93 | 1.30 | 0.18 | 2.85 | 4.39 |
| P8 | E2 | Average | 0.85 | 1.02 | 1.48 | 0.21 | 2.51 | 3.86 |
| P9 | E3 | E6 | 0.92 | 0.86 | 1.08 | 0.06 | 3.43 | 5.28 |
| P10 | E3 | E7 | 0.91 | 0.86 | 1.10 | 0.06 | 3.36 | 5.18 |
| P11 | E3 | E8 | 0.91 | 0.86 | 1.12 | 0.18 | 3.31 | 5.10 |
| P12 | E3 | Average | 0.88 | 0.96 | 1.32 | 0.20 | 2.79 | 4.30 |
| P13 | E4 | E6 | 0.93 | 0.78 | 0.98 | 0.17 | 3.77 | 5.81 |
| P14 | E4 | E7 | 0.93 | 0.76 | 1.00 | 0.16 | 3.69 | 5.68 |
| P15 | E4 | E8 | 0.93 | 0.78 | 1.00 | 0.16 | 3.70 | 5.69 |
| P16 | E4 | Average | 0.89 | 0.91 | 1.28 | 0.20 | 2.90 | 4.47 |
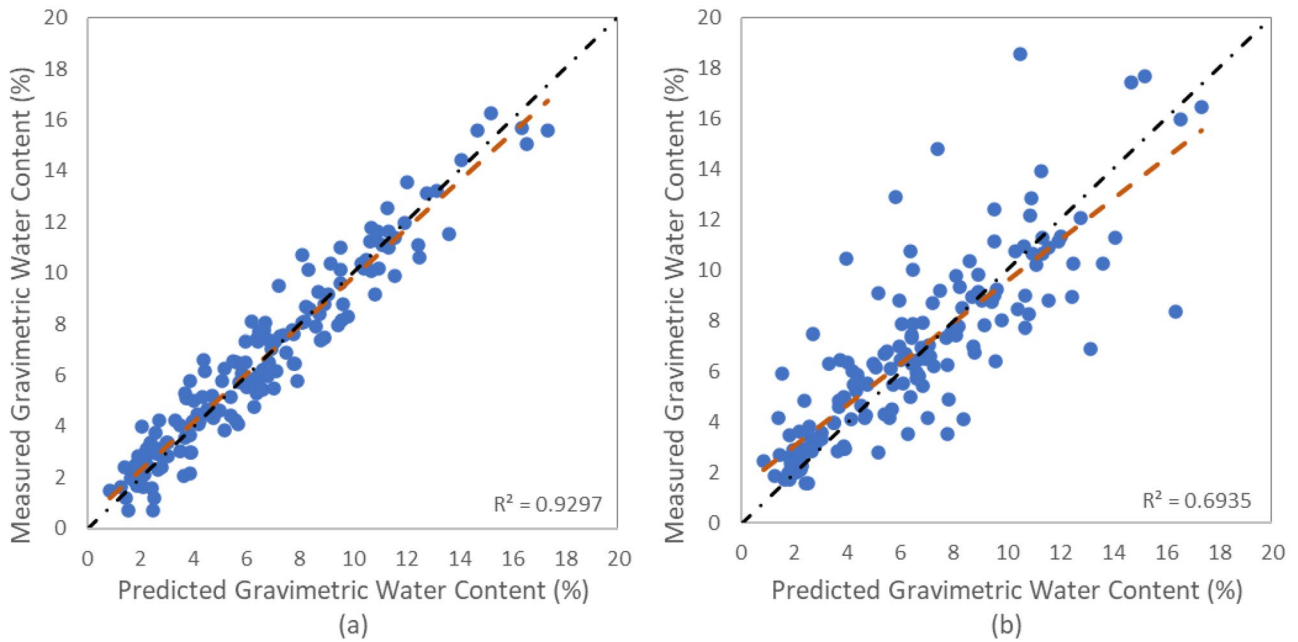
**Fig. 6** Predicted and measured gravimetric water contents for models with the best and worst prediction performance: **a** $s_r$ was estimated from E4, and $\psi_{dry}$ was estimated from E6; **b** $s_r$ was estimated from E1, and average value was used for $\psi_{dry}$
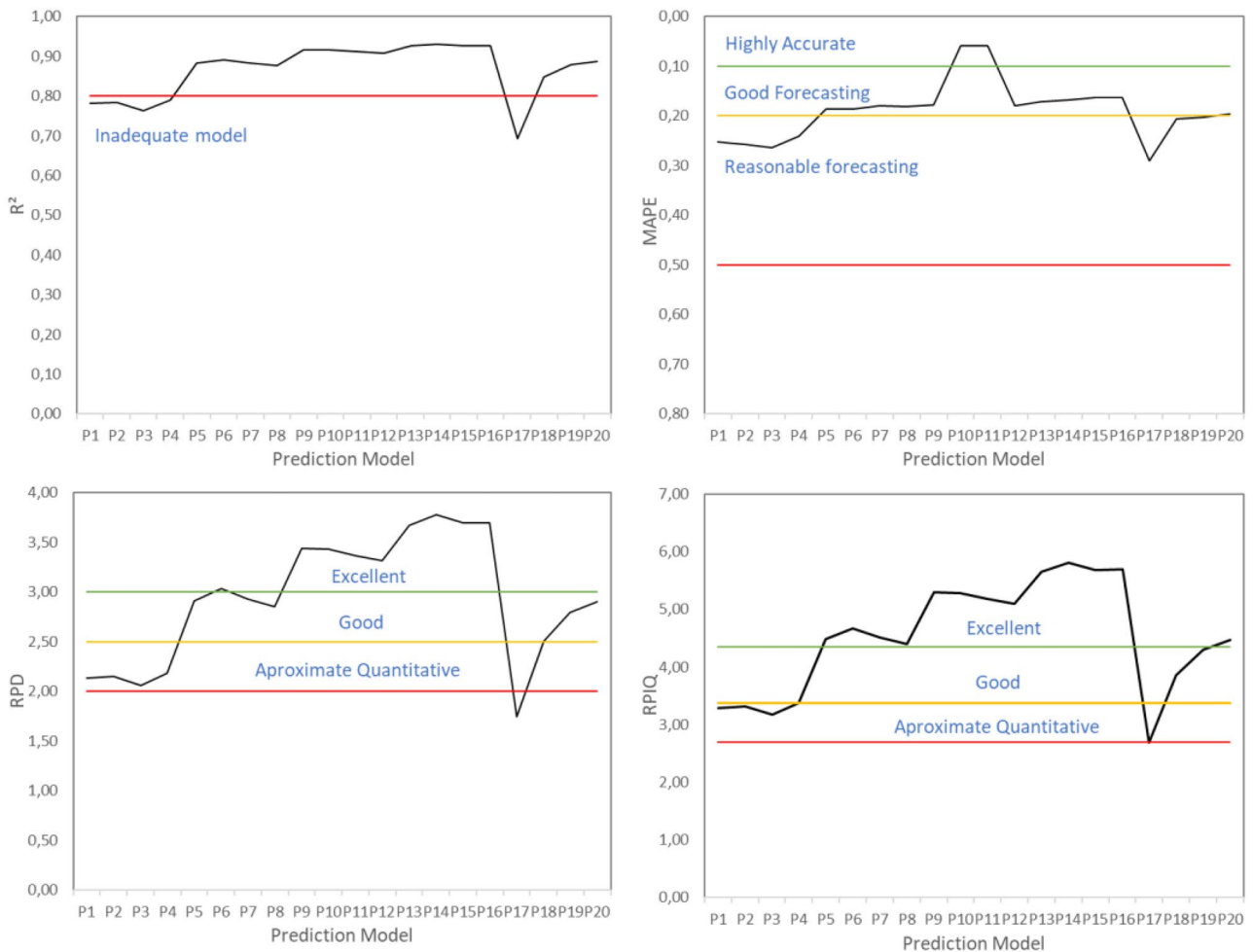


**Fig. 7** Comparison of prediction performances of the models

# 4 Conclusion

Models used to predict the SWCC generally depend on suction measurements within the transition zone; however, the validity of these models for soils under low-moisture conditions, typical of arid and semiarid regions, is uncertain. Thus, models are needed that can accurately estimate the residual saturation zone of the SWCC, especially in cohesive soils.

In this study, a semi-logarithmic linear model was used to simulate the residual saturation zone of the SWCC. The model has two parameters: the suction at zero water content ($\psi_{dry}$) and the slope of the SWCC in the residual saturation zone ($s_r$). Using an experimental dataset for 40 cohesive soil samples, correlations between model parameters and the soils' physical, chemical, and mineralogical properties were examined. The $s_r$ values were observed to have strong correlations with the consistency limits, activity, and some spectral properties that reflect the quantity and mineralogical composition of the clay minerals in the soils. However, the correlations between $\psi_{dry}$ and soil properties in this study's dataset are not very strong, consistent with previous studies. Using the physical, chemical, and spectral properties of soils as explanatory variables, stepwise regression analyses were used to obtain empirical equations that could be used to predict the model parameters. Sixteen estimation models for predicting the residual saturation zone of the SWCC in cohesive soils were developed by inserting combinations of empirical equations based on regression analyses for $s_r$ and $\psi_{dry}$ into the two-parameter semi-logarithmic linear model. The results show that using the semi-logarithmic linear model to describe the SWCC's residual saturation zone provides satisfactory estimations. The fact that the prediction equations include physicochemical soil properties, such as the liquid limit and specific surface charge, indicates that parameters related to the amount and type of clay minerals should be included for good predictions of the models that define the dry region of the SWCC in cohesive soils. Furthermore, the results demonstrated that spectral absorption properties in the SWIR region have the potential to predict the residual saturation zone of the SWCC in cohesive soils.

**Data availability** The data underlying this article will be shared on reasonable request to the corresponding author.

## Declarations

**Conflict of interest** The author declares no competing interests.

# References

Agus SS, Leong EC, Rahardjo H (2001) Soil—water characteristic curves of Singapore residual soils. In: Toll, D.G (eds) Unsaturated soil concepts and their application in geotechnical practice. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-9775-3_4

Agus SS, Schanz T (2005) Comparison of four methods for measuring total suction. Vadose Zone J 4(4):1087–1095. https://doi.org/10.2136/vzj2004.0133

Arifin YF, Schanz T (2009) Osmotic suction of highly plastic clays. Acta Geotech 4:177–191. https://doi.org/10.1007/s11440-009-0097-0

Arthur E, Tuller M, Moldrup P, Resurreccion AC, Meding MS, Kawamoto K, Komatsu T, de Jonge LW (2013) Soil specific surface area and non-singularity of soil-water retention at low saturations. Soil Sci Soc Am J 77(1):43–53. https://doi.org/10.2136/sssaj2012.0262

Azadi A, Baninemeh J (2022) Performance analysis of geostatistical approach and PCA techniques for mapping cation exchange capacity in the South West of Iran. Eurasian Soil Sci 55:1749–1760. https://doi.org/10.1134/S1064229322601494

Babaeian E, Homaee M, Montzka C, Vereecken H, Norouzi AA (2015) Towards retrieving soil hydraulic properties by hyperspectral remote sensing. Vadose Zone J 14(3):1–17. https://doi.org/10.2136/vzj2014.07.0080

Ben-Dor E, Chabrillat S, Demattê JAM, Taylor GR, Hill J, Whiting ML, Sommer S (2009) Using imaging spectroscopy to study soil properties. Remote Sens Environ 113(1):S38–S55. https://doi.org/10.1016/j.rse.2008.09.019

Bower CA, Reitemeier RF, Fireman M (1952) Exchangeable cation analysis of saline and alkali soils. Soil Sci 73(4):251–261. https://doi.org/10.1097/00010694-195204000-00001

Brooks RH, Corey AT (1964) Hydraulic properties of porous media. hydrology papers 3, Colorado State University, Fort Collins

Bulut R, Wray WK (2005) Free energy of water-suction-in filter papers. Geotech Test J 28(4):355–264. https://doi.org/10.1520/GTJ12307

Bulut R, Leong EC (2008) Indirect measurement of suction. In: Tarantino, A, Romero, E, Cui, YJ (eds) Laboratory and field testing of unsaturated soils. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-8819-3_3

Byun Y, Seo C, Yun T, Joo Y, Jo HY (2023) Prediction of Na-and Ca-montmorillonite contents and swelling properties of clay mixtures using Vis-NIR spectroscopy. Geoderma 430:116294. https://doi.org/10.1016/j.geoderma.2022.116294

Campbell GS, Shiozawa S (1992) Prediction of hydraulic properties of soils using particle size distribution and bulk density data. International Workshop on Indirect Methods for Estimating the Hydraulic properties of Unsaturated Soils. University of California Press, Berkeley, CA, pp. 317–328

Chen C, Hu K, Arthur E, Ren T (2014) Modeling soil water retention curves in the dry range using the hygroscopic water content. Vadose Zone J 13(11):vzj2014.06.0062. https://doi.org/10.2136/vzj2014.06.0062

Chen J, Li Z, Han Z, Gao Y, Xiong Y, Ding R (2022) Effect of initial void ratio on the soil water characteristics of unsaturated soil at high suctions. Hydrogeol Eng Geol 49(4):47–54. https://doi.org/10.16030/j.cnki.issn.1000-3665.202106043

Daoud JI (2017) Multicollinearity and regression analysis. In Journal of Physics: Conference Series (Vol. 949, No. 1, p. 012009). IOP Publishing. https://doi.org/10.1088/1742-6596/949/1/012009

Demirci Ö (2010) Determination of the relationship between the physical chemical and mineralogical properties of fine grained soils (in Turkish). Thesis, Department of Civil Engineering, Suleyman Demirel University, Isparta, Türkiye, MSc

Dufréchou G, Grandjean G, Bourguignon A (2015) Geometrical analysis of laboratory soil spectra in the short-wave infrared domain: clay composition and estimation of the swelling potential. Geoderma 243:92–107. https://doi.org/10.1016/j.geoderma.2014.12.014

Ebrahimi-Birang N, Fredlund DG (2016) Assessment of the WP4-T device for measuring total suction. Geotech Test J 39(3):500–506. https://doi.org/10.1520/GTJ20150118

Erzin Y (2007) Artificial neural networks approach for swell pressure versus soil suction behaviour. Can Geotech J 44(10):1215–1223. https://doi.org/10.1139/T07-052

Escadafal R (1993) Remote sensing of soil color: principles and applications. Remote Sens Rev 7(3–4):261–279. https://doi.org/10.1080/02757259309532181

Fang Q, Ren X, Zhang B, Chen X, Guo Z (2022) A flexible soil-water characteristic curve model considering physical constraints of parameters. Eng Geol 305:106717. https://doi.org/10.1016/j.enggeo.2022.106717

Fayer MJ, Simmons CS (1995) Modified soil water retention functions for all matric suctions. Water Resour Res 31(5):1233–1238. https://doi.org/10.1029/95WR00173

Fredlund DG, Wong DK (1989) Calibration of thermal conductivity sensors for measuring soil suction. Geotech Test J 12(3):188–194. https://doi.org/10.1520/GTJ10967J

Fredlund DG, Xing A (1994) Equations for the soil-water characteristic curve. Can Geotech J 31(4):521–532. https://doi.org/10.1139/t94-061

Groenevelt PH, Grant CD (2004) A new model for the soil-water retention curve that solves the problem of residual water contents. Eur J Soil Sci 55(3):479–485. https://doi.org/10.1111/j.1365-2389.2004.00617.x

Guan Y (1996) The measurement of soil suction, Ph (Doctoral dissertation, Thesis, Department of Civil Engineering, University of Saskatchewan, Saskatoon, Canada)

Guo L, Chen G, Li C, Xia M, Gong S, Zheng L (2021) A bound water model for numerical simulation of SWCC in the wide suction range based on DDA. Comput Geotech 139:104378. https://doi.org/10.1016/j.compgeo.2021.104378

Hosseinpour M, Sharifi H, Sharifi Y (2018) Stepwise regression modeling for compressive strength assessment of mortar containing metakaolin. Int J Modell Simul 38(4):207–215. https://doi.org/10.1080/02286203.2017.1422096

Janik LJ, Merry RH, Forrester ST, Lanyon DM, Rawson A (2007) Rapid prediction of soil water retention using mid infrared spectroscopy. Soil Sci Soc Am J 71(2):507–514. https://doi.org/10.2136/sssaj2005.0391

Jung Y, Hu J (2015) A K-fold averaging cross-validation procedure. J Nonparam Stat 27(2):167–179. https://doi.org/10.1080/10485252.2015.1010532

Kariuki PC, van der Meer F, Siderius W (2003) Classification of soils based on engineering indices and spectral data. Int J Remote Sens 24(12):2567–2574. https://doi.org/10.1080/0143116031000075927

Kariuki PC, Woldai T, van der Meer F (2004) Effectiveness of spectroscopy in identification of swelling indicator clay minerals. Int J Remote Sens 25(2):455–469. https://doi.org/10.1080/0143116031000084314

Karup D, Moldrup P, Tuller M, Arthur E, de Jonge LW (2017) Prediction of the soil water retention curve for structured soil from saturation to oven-dryness. Eur J Soil Sci 68(1):57–65. https://doi.org/10.1111/ejss.12401

Kocaman K, Ozocak A, Edil TB, Bol E, Sert S, Onturk K, Ozsagir M (2022) Evaluation of soil-water characteristic curve and pore-size distribution of fine-grained soils. Water 14(21):3445. https://doi.org/10.3390/w14213445

Leong EC, Tripathy S, Rahardjo H (2003) Total suction measurement of unsaturated soils with a device using the chilled-mirror dew-point technique. Géotechnique 53(2):173–182. https://doi.org/10.1680/geot.2003.53.2.173

Lewis CD (1982) Industrial and business forecasting method. Butterworth Scientific Publishers, London

Li N, He F, Ma W, Wang R, Zhang X (2020) Wind power prediction of kernel extreme learning machine based on differential evolution algorithm and cross validation algorithm. IEEE Access 8:68874–68882. https://doi.org/10.1109/ACCESS.2020.2985381

Likos WJ, Lu N (2002) Filter paper technique for measuring total soil suction. Transp Res Rec 1786(1):120–128. https://doi.org/10.3141/1786-14

Liu Y, Heuvelink GB, Bai Z, He P, Xu X, Ding W, Huang S (2021) Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. Field Crops Res 264:108098. https://doi.org/10.1016/j.fcr.2021.108098

Loring DH, Rantala RT (1992) Manual for the geochemical analyses of marine sediments and suspended particulate matter. Earth Sci Rev 32(4):235–283. https://doi.org/10.1016/0012-8252(92)90001-A

Lu N, Khorshidi M (2015) Mechanisms for soil-water retention and hysteresis at high suction range. J Geotech Geoenviron Eng 141(8):04015032. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001325

Lu N, Likos WJ (2006) Suction stress characteristic curve for unsaturated soil. J Geotech Geoenviron Eng 132(2):131–142. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:2(131)

Lucko G, Anderson-Cook CM, Vorster MC (2006) Statistical considerations for predicting residual value of heavy equipment. J Constr Eng Manage 132(7):723–732. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:7(723)

Ludwig B, Vormstein S, Niebuhr J, Heinze S, Marschner B, Vohland M (2017) Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects. Geoderma 288:37–46. https://doi.org/10.1016/j.geoderma.2016.10.022

Malaya C, Sreedeep S (2012) Critical review on the parameters influencing soil-water characteristic curve. J Irrig Drain Eng 138(1):55–62. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000371

McKee CR, Bumb AC (1987) Flow-testing coalbed methane production wells in the presence of water and gas. Society of Petroleum Engineers (SPE) Formation Evaluation, Richardson, Tex. pp. 599–608.

Mulder VL, Plötze M, de Bruin S, Schaepman ME, Mavris C, Kokaly RF, Egli M (2013) Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1–2.4 µm) and regression tree analysis. Geoderma 207:279–290. https://doi.org/10.1016/j.geoderma.2013.05.011

Nam S, Gutierrez M, Diplas P, Petrie J, Wayllace A, Lu N, Muñoz JJ (2010) Comparison of testing techniques and models for establishing the SWCC of riverbank soils. Eng Geol 110(1–2):1–10. https://doi.org/10.1016/j.enggeo.2009.09.003

Ng CWW, Menzies B (2007) Advanced Unsaturated Soil Mechanics and Engineering Taylor Francis. New York

Norouzi S, Sadeghi M, Tuller M, Liaghat A, Jones SB, Ebrahimian H (2022) A novel physical-empirical model linking shortwave infrared reflectance and soil water retention. J Hydrol 614:128653. https://doi.org/10.1016/j.jhydrol.2022.128653

Pesch C, Weber PL, Moldrup P, de Jonge LW, Arthur E, Greve MH (2022) Physical characterization of glacial rock flours from fjord deposits in South Greenland-Toward soil amendment. Soil Sci Soc Am J 86(2):407–422. https://doi.org/10.1002/saj2.20352

Pittaki-Chrysodonta Z, Arthur E, Moldrup P, Knadel M, Norgaard T, Iversen BV, de Jonge LW (2019) Comparing visible–near-infrared spectroscopy and a pedotransfer function for predicting the dry region of the soil-water retention curve. Vadose Zone J 18(1):1–13. https://doi.org/10.2136/vzj2018.09.0180

Pittaki-Chrysodonta Z, Hartemink AE, Huang J (2021) Rapid estimation of a soil–water retention curve using visible–near infrared spectroscopy. J Hydrol 603:127195. https://doi.org/10.1016/j.jhydrol.2021.127195

Rahardjo H, Leong EC (2006) Suction measurements. In: Unsaturated Soils 2006 (pp. 81–104). https://doi.org/10.1061/40802(189)3

Resurreccion AC, Moldrup P, Tuller M, Ferré TPA, Kawamoto K, Komatsu T, de Jonge LW (2011) Relationship between specific surface area and the dry end of the water retention curve for soils

with varying clay and organic carbon contents. Water Resour Res 47(6):W06522. https://doi.org/10.1029/2010WR010229

Ross PJ, Williams J, Bristow KL (1991) Equation for extending water-retention curves to dryness. Soil Sci Soc Am J 55(4):923–927. https://doi.org/10.2136/sssaj1991.03615995005500040004x

Rossi C, Nimmo JR (1994) Modeling of soil water retention from saturation to oven dryness. Water Resour Res 30(3):701–708. https://doi.org/10.1029/93WR03238

Saeys W, Mouazen AM, Ramon H (2005) Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosyst Eng 91(4):393–402. https://doi.org/10.1016/j.biosystemseng.2005.05.001

Salager S, Nuth M, Ferrari A, Laloui L (2013) Investigation into water retention behaviour of deformable soils. Can Geotech J 50(2):200–208. https://doi.org/10.1139/cgj-2011-0409

Santra P, Sahoo RN, Das BS, Samal RN, Pattanaik AK, Gupta VK (2009) Estimation of soil hydraulic properties using proximal spectral reflectance in visible, near-infrared, and shortwave-infrared (VIS–NIR–SWIR) region. Geoderma 152(3–4):338–349. https://doi.org/10.1016/j.geoderma.2009.07.001

Satyanaga A, Rahardjo H, Koh ZH, Mohamed H (2019) Measurement of a soil-water characteristic curve and unsaturated permeability using the evaporation method and the chilled-mirror method. J Zhejiang Univ -Sci A 20:368–374. https://doi.org/10.1631/jzus.A1800593

Schneider M, Goss KU (2012) Prediction of the water sorption isotherm in air dry soils. Geoderma 170:64–69. https://doi.org/10.1016/j.geoderma.2011.10.008

Shrestha N (2020) Detecting multicollinearity in regression analysis. Am J Appl Math Stat 8(2):39–42. https://doi.org/10.12691/ajams-8-2-1

Snee RD (1977) Validation of regression models: methods and examples. Technometrics 19(4):415–428. https://doi.org/10.1080/00401706.1977.10489581

Sridharan A, Rao SM (1988) A scientific basis for the use of index tests in identification of expansive soils. Geotech Test J 11(3):208–212. https://doi.org/10.1520/GTJ10008J

Taghdis S, Farpoor MH, Mahmoodabadi M (2022) Pedological assessments along an arid and semi-arid transect using soil spectral behavior analysis. Catena 214:106288. https://doi.org/10.1016/j.catena.2022.106288

Uzundurukan S (2023) Prediction of soil–water characteristic curve for plastic soils using PSO algorithm. Environ Earth Sci 82:37. https://doi.org/10.1007/s12665-022-10687-0

Vanapalli SK, Pufahl DE, Fredlund DG (1999) The influence of soil structure and stress history on the soil–water characteristics of a compacted till. Géotechnique 49(2):143–159. https://doi.org/10.1680/geot.1999.49.2.143

van der Meer F (1999) Can we map swelling clays with remote sensing? Int J Appl Earth Obs Geoinf 1(1):27–35. https://doi.org/10.1016/S0303-2434(99)85025-9

van der Meer F (2004) Analysis of spectral absorption features in hyperspectral imagery. Int J Appl Earth Obs Geoinf 5(1):55–68. https://doi.org/10.1016/j.jag.2003.09.001

van Genuchten M, Th, (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 44(5):892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131(1–2):59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

Webb SW (2000) A simple extension of two-phase characteristic curves to include the dry region. Water Resour Res 36(6):1425–1430. https://doi.org/10.1029/2000WR900057

Williams J, Prebble R, Williams W, Hignett C (1983) The influence of texture, structure and clay mineralogy on the soil moisture characteristic. Aust J Soil Res 21:15–32

Woodburn JA, Holden JC, Peter P (1993) The transistor psychrometer: a new instrument for measuring soil suction. In: Houston SL, Wray WK (eds), Unsaturated Soils. Geotechnical Special Publication No. 39. ASCE, New York

Yong RN, Warkentin BP (1975) Soil properties and soil behavior. Elsevier, Amsterdam

Yong RN (1999) Soil suction and soil-water potentials in swelling clays in engineered clay barriers. Eng Geol 54(1–2):3–13. https://doi.org/10.1016/S0013-7952(99)00056-3

Zapata CE (1999) Uncertainty in soil-water characteristic curve and impacts on unsaturated shear strength predictions. Ph.D. Dissertation, Arizona State University, Tempe, United States

Zhai Q, Rahardjo H, Satyanaga A (2018) Estimation of air permeability function from soil-water characteristic curve. Can Geotech J 56(4):505–513. https://doi.org/10.1139/cgj-2017-0579

Zhai Q, Rahardjo H, Satyanaga A (2019) Uncertainty in the estimation of hysteresis of soil–water characteristic curve. Environ Geotech 6(4):204–213. https://doi.org/10.1680/jenge.17.00008

Zhai Q, Rahardjo H, Satyanaga A, Zhu Y, Dai G, Zhao X (2021) Estimation of wetting hydraulic conductivity function for unsaturated sandy soil. Eng Geol 285:106034. https://doi.org/10.1016/j.enggeo.2021.106034

Zhou B, Lu N (2021) Correlation between Atterberg limits and soil adsorptive water. J Geotech Geoenviron Eng 147(2):04020162. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002463

Zhou P, Sudduth KA, Veum KS, Li M (2022) Extraction of reflectance spectra features for estimation of surface, subsurface, and profile soil properties. Comput Electron Agric 196:106845. https://doi.org/10.1016/j.compag.2022.106845