

Article

Estimation of Cation Exchange Capacity for Low-Activity Clay Soil Fractions Using Experimental Data from South China

Jun Zhu ¹ and Zhong-Xiu Sun ^{2,*}

¹ School of Computer and Software, Nanjing Vocational University of Industry Technology, Nanjing 210023, China; junzhu@niit.edu.cn

² College of Land and Environment, Shenyang Agricultural University, Shenyang 110866, China

* Correspondence: zhongxiusun@syau.edu.cn; Tel.: +86-15734005989

Abstract: The cation exchange capacity (CEC) of the clay fraction (<2 μm), denoted as CEC_{clay}, serves as a crucial indicator for identifying low-activity clay (LAC) soils and is an essential criterion in soil classification. Traditional methods of estimating CEC_{clay}, such as dividing the whole-soil CEC (CEC_{soil}) by the clay content, can be problematic due to biases introduced by soil organic matter and different types of clay minerals. To address this issue, we introduced a soil pedotransfer functions (PTFs) approach to predict CEC_{clay} from CEC_{soil} using experimental soil data. We conducted a study on 122 pedons in South China, focusing on highly weathered and strongly leached soils. Samples from the B horizon were used, and eight models and PTFs (four machine learning methods, multiple linear regression (MLR) and three PTFs from publication) were evaluated for their predictive performance. Four covariate datasets were combined based on available soil data and environmental variables and various parameters for machine learning techniques including an artificial neural network, a deep belief network, support vector regression and random forest were optimized. The results, based on 10-fold cross-validation, showed that the simple division of CEC_{soil} by clay content led to significant overestimation of CEC_{clay}, with a mean error of 14.42 cmol(+) kg⁻¹. MLR produced the most accurate predictions, with an R² of 0.63–0.71 and root mean squared errors (RMSE) of 3.21–3.64 cmol(+) kg⁻¹. The incorporation of environmental variables improved the accuracy by 2–10%. A linear model was fitted to enhance the current calculation method, resulting in the equation: $CEC_{clay} = 15.31 + 15.90 \times (CEC_{soil} / Clay)$, with an R² of 0.41 and RMSE of 4.48 cmol(+) kg⁻¹. Therefore, given limited soil data, the MLR PTFs with explicit equations were recommended for predicting the CEC_{clay} of B horizons in humid subtropical regions.



Citation: Zhu, J.; Sun, Z.-X. Estimation of Cation Exchange Capacity for Low-Activity Clay Soil Fractions Using Experimental Data from South China. *Agronomy* **2024**, *14*, 2671. <https://doi.org/10.3390/agronomy14112671>

Academic Editor: Carlo Leifert

Received: 7 October 2024

Revised: 3 November 2024

Accepted: 11 November 2024

Published: 13 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: pedotransfer function; modeling techniques; machine learning; multiple linear regression; deep learning

1. Introduction

The cation exchange capacity (CEC) of the clay fraction (<2 μm) (CEC_{clay}) has been employed to indicate the activity of the clay fraction [1,2]. Soil with a CEC_{clay} value greater or less than 24 cmol(+) kg⁻¹ is referred to as high-activity clay (HAC) or low-activity clay (LAC) [1], respectively, of which the threshold value was set according to substantial soil survey and laboratory analysis for purpose of soil classification [3–6] and can be used to indicate the ability of soil to adsorb and retain cations. From the perspective of pedogenesis, LAC reflects the weathering degree of the mineral portion of the pedons [1]. LAC soils may be characterized as having high contents of kaolinite in the clay fraction and can be regarded as an advanced stage of soil formation in subtropical and tropical areas [7]. Due to extensive leaching over time, the soil nutrient contents in LAC soils are generally lower than those in HAC soils, which can result in a rapid acidification and depletion. Identifying LAC soils is beneficial for precision agriculture, especially for secondary food crops [8].

The value of CEC_{clay} not only can be used to indicate the soil quality and productivity but also provides important input for soil classification purposes, such as the USDA Soil

Taxonomy (ST) [3] (Table S1), World Reference Base for Soil Resources (WRB) [4] (Table S2) and Chinese Soil Taxonomy (CST) [5,6] (Table S3). Taking the ST as an example, the CEC_{clay} is a required characteristic for determining the kandic and oxic horizons and other subgroups (Table S1). In the CST, the LAC ferric horizons and ferralic horizons are employed to indicate medium and high ferrallitization degrees (Table S3), respectively. LAC ferric and ferralic horizons are defined as the diagnostic subsurface horizons for the orders of Ferrosols and Ferraliosols, respectively, which are widely distributed across subtropical and tropical regions [9,10].

Clay extraction measurement is very time-consuming and costly. Thus, the CEC_{clay} is usually calculated as the ratio of the CEC of the fine earth fraction (<2 mm) (CEC_{soil}) to the clay percentage for simplicity [4,5,11]. This method assumes that the CEC_{soil} for genetic B horizons is mainly contributed by clays [12,13]. The clay content does not always contribute significantly to the CEC_{soil} in soils in subtropical and tropical regions [14,15], as well as in other regions [16–18]. Other factors, such as silts and the soil organic matter (SOM) also contribute to the CEC_{soil} [19,20]. For example, SOM may account for 20–60% of the CEC_{soil} in Spodosols (Podzols in WRB) according to laboratory tests in Russia and Canada [16]. It is therefore understandable that between the CEC_{soil} and CEC_{clay} , there is no simple transformation equation that applies to different soils. The ratios of the CEC_{soil} and clay percentages may introduce biases when allocating soils to a classification system (Tables S1–S3), even if the soils have clear morphological evidence. However, little attention has been paid to this overestimation or to updating the CEC_{clay} calculation method so far.

The pedotransfer function (PTF) is an effective alternative for estimating the CEC_{clay} based on the relationships between the CEC_{clay} and easily-measured soil properties and other possible soil-forming factors. The predictive techniques range from multiple linear regression (MLR) [21], multiple nonlinear regression [22], and principal component regression [18] to machine learning [23] and ensemble models [24]. The most important covariates can be identified [19]. In recent years, machine learning techniques have been widely adopted for the soil quality assessment [25], digital soil mapping [26–29], prediction of soil properties that are difficult to measure (e.g., hydraulic conductivity) [30,31], and the estimate of changes in soil properties [32,33]. However, much attention has been paid to CEC_{soil} PTF fitting [23,24,34], soil mapping [35] and CEC_{soil} determination using proximal soil sensing techniques, such as visible and near-infrared spectroscopy [36] and electromagnetic instrumentation [37]. Nevertheless, few CEC_{clay} PTFs have been presented in the literature.

In this study, two hypotheses were examined: (1) nonlinear relationships between CEC_{clay} and predictors that were captured by machine learning techniques may generate more accurate predictions than those obtained by MLR, and (2) the performance of PTFs that only use measured soil properties as covariates can be enhanced by incorporating environmental variables as auxiliary predictors (e.g., terrain attributes and climatic variables). Eight models and PTFs were evaluated to establish an optimal predictive model for CEC_{clay} , including an artificial neural network (ANN), a deep belief network (DBN), support vector regression (SVR), random forest (RF), MLR and three published PTFs [3,38,39]. Thus, this study could be considered as one attempt to investigate the prediction of CEC_{clay} based on advanced machine learning and environmental variables in which soil samples were collected from soil profiles in South China. The comparison of PTFs and machine learning techniques contributes to the prediction of soil properties that are laborious, difficult or time-consuming to measure.

2. Methods and Materials

2.1. Study Areas

This study was conducted in South China, mainly south of the Yangtze River (Figure 1) (17°30' N to 34°29' N and 95°24' E to 124°11' E), covering approximately 1.7 million km². The climate is predominantly subtropical, characterized by hot and humid summers and mild winters. The mean annual precipitation (MAP) is 1415 mm, with a mean annual

air temperature (MAT) of 17.9 °C [40]. The soil temperature and moisture regimes are mainly thermic and udic and perudic, respectively. The most widely distributed land cover in this area is forest (51.2%), followed by grassland (26.3%) and cropland (18.9%) [41]. This area is rarely affected by dust deposits from northwestern regions [42], and the main soil types are Ferralosols (48.7%), Argosols (22.9%) and Cambosols (20.6%) according to CST [6]. The statistics of main soil types based on soil maps [41] are shown in Table 1. Ferralosols are mainly referred to as Ultisols, Alfisols and Inceptisols in ST and as Acrisols, Lixisols, Plinthosols and Nitisols in WRB. Argosols can be mainly referred to as Alfisols or Ultisols in ST and as Luvisols, Lixisols, Acrisols or Alisols in WRB, and Cambosols can be mainly referred to as Inceptisols in ST and as Cambisols in WRB. The main parent materials in this area are clastic (42.0%) and calcareous sedimentary rocks (29.4%). The values of the aforementioned environmental variables were obtained from a national dataset in [32,41]. The soil-forming environments may lead to notable changes in soil physicochemical properties in acidic soils, which makes this area ideal for investigating the CEC_{soil} and CEC_{clay} relationship.

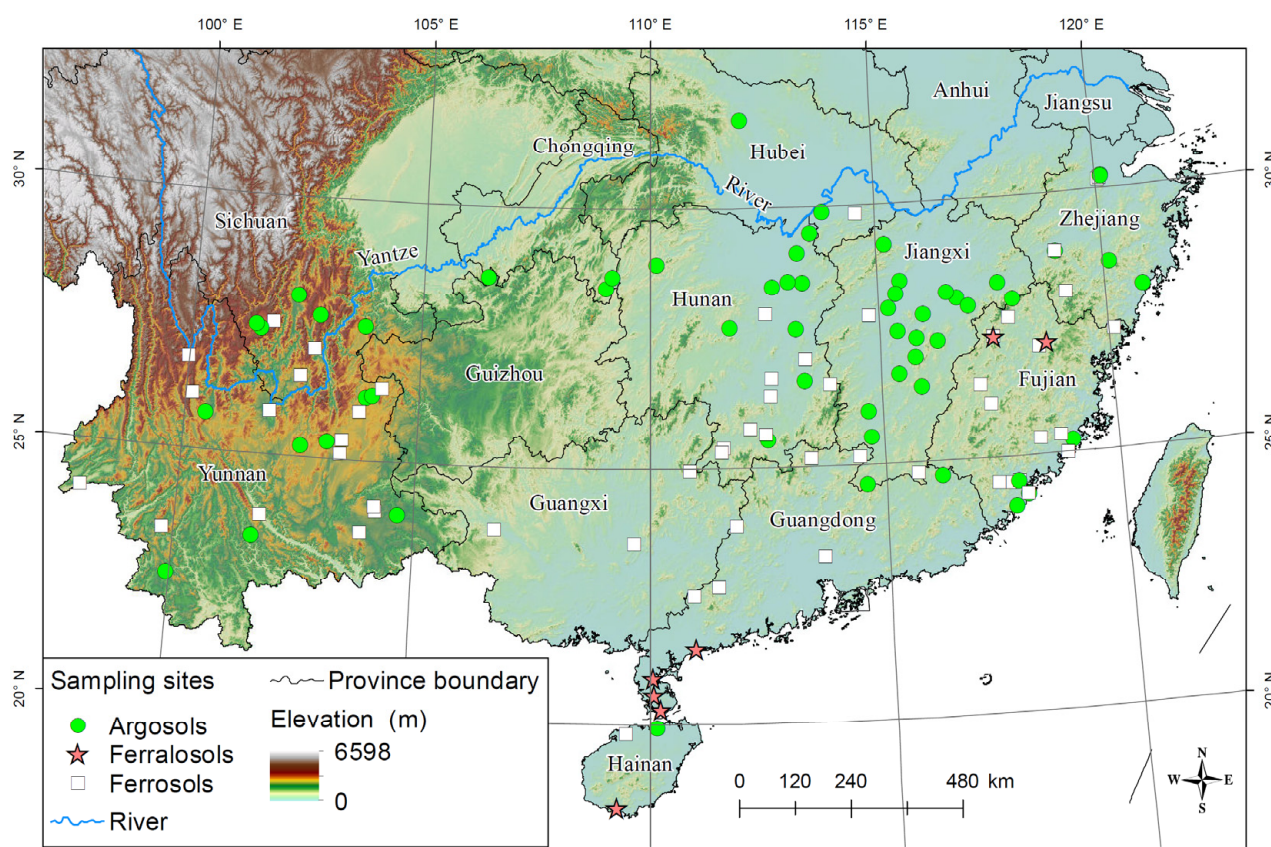


Figure 1. Spatial distribution of the sampled pedons in this study ($n = 122$). Argosols are soils with an argillic horizon and a CEC_{clay} of more than $24 \text{ cmol}(+) \text{ kg}^{-1}$, while Ferralosols and Ferralosols are soils with CEC_{clay} values of less than 16 and $24 \text{ cmol}(+) \text{ kg}^{-1}$, respectively [6]. Ferralosols are mainly referred to as Ultisols, Alfisols and Inceptisols in ST and as Acrisols, Lixisols, Plinthosols and Nitisols in WRB. Ferralosols are mainly referred to as Oxisols in ST and as Ferralosols, Plinthosols, Acrisols and Lixisols in WRB.

Table 1. Statistics of the soil properties (0–20 cm) regarding main soil types in the study area.

| Soil Property | Argosols | | | | Cambosols | | | | Ferrosols | | | |
|---------------------------------------|------------------|-------|--------|--------------------|-----------|-------|--------|--------------------|-----------|-------|--------|--------------------|
| | Min ^a | Mean | Max | Standard Deviation | Min | Mean | Max | Standard Deviation | Min | Mean | Max | Standard Deviation |
| BS (%) | 6.02 | 56.32 | 143.33 | 26.62 | 6.27 | 42.05 | 104.83 | 20.29 | 4.09 | 34.12 | 171.49 | 22.82 |
| CaCO ₃ (g/kg) | 0.62 | 24.44 | 149.87 | 17.99 | 1.23 | 25.12 | 121.09 | 15.76 | 0.20 | 30.86 | 166.62 | 21.19 |
| CEC (cmol(+) kg ⁻¹) | 4.84 | 15.64 | 32.16 | 2.09 | 2.71 | 14.50 | 36.56 | 2.05 | 0.82 | 12.50 | 36.38 | 2.43 |
| Clay (%) | 8.29 | 26.52 | 56.41 | 4.48 | 5.46 | 24.42 | 52.32 | 4.81 | 2.95 | 25.97 | 56.41 | 4.61 |
| Silt (%) | 17.04 | 41.88 | 68.81 | 4.69 | 3.58 | 40.59 | 70.17 | 5.03 | 4.89 | 36.68 | 67.44 | 5.59 |
| Fe _d (g kg ⁻¹) | 1.63 | 21.62 | 63.22 | 6.32 | 4.26 | 19.02 | 63.22 | 4.23 | 0.61 | 20.00 | 63.22 | 5.33 |
| pH | 4.08 | 5.80 | 8.59 | 0.77 | 4.13 | 5.51 | 8.70 | 0.70 | 3.69 | 5.16 | 8.63 | 0.55 |
| SOC (g kg ⁻¹) | 4.91 | 17.57 | 75.71 | 4.92 | 3.69 | 15.85 | 58.88 | 3.52 | 1.68 | 16.61 | 75.71 | 3.36 |
| TK (g kg ⁻¹) | 3.58 | 16.42 | 40.16 | 3.31 | 4.54 | 16.53 | 43.55 | 2.70 | 0.60 | 14.98 | 43.55 | 3.24 |
| TN (g kg ⁻¹) | 0.38 | 1.66 | 6.02 | 0.33 | 0.18 | 1.41 | 4.12 | 0.27 | 0.10 | 1.37 | 6.78 | 0.28 |
| TP (g kg ⁻¹) | 0.10 | 0.56 | 2.77 | 0.13 | 0.07 | 0.48 | 2.79 | 0.11 | 0.04 | 0.47 | 2.75 | 0.09 |

^a Min: minimum; Max: maximum; BS: base saturation; Fe_d: concentration of free iron oxides; SOC: soil organic carbon; TK: total potassium; TN: total nitrogen; TP: total phosphorus.

2.2. Soil Sampling and Measurement

From 2009–2019, the National Soil Series Survey (NSSS) was carried out to investigate the spatial pattern of soil types across China. The soil sampling sites were chosen according to the strata of land uses, parent materials and soil types. A detailed description of this survey can be found in [32,41]. A total of 122 samples of genetic B horizons were selected from the NSSS database [41], in which clear evidence of clay eluviation and illuviation can be found according to the detailed soil profile description. These horizons satisfied all the required characteristics of the LAC ferric horizon (i.e., the soil texture, soil color and concentration of free iron oxides (Fe_d)) except the value of the CEC_{clay} [6].

The silt (0.002–0.05 mm) and clay contents (<0.002 mm), soil organic carbon (SOC), pH, Fe_d and CEC_{soil} were recorded, as was a detailed description of the pedogenetic characteristics of each layer. These profiles were classified as Ferralosols, Ferrosols and Argosols according to the CST [6] (Figure 1). The main reason why we selected the samples of the B horizons was that CEC_{clay} values were required to determine the soil types of these profiles. The soil textures mainly included clay, silty clay, silty clay loam, sandy clay loam, clay loam and loam (Figure 2) [43].

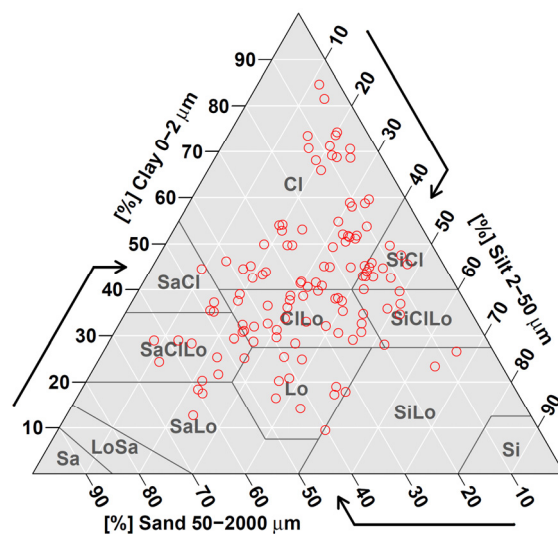


Figure 2. Soil texture classes of the involved samples based on the USDA soil texture triangle. Cl: clay; SiCl: silty clay; SiClLo: silty clay loam; SaCl: sandy clay; SaClLo: sandy clay loam; ClLo: clay loam; Si: silt; SiLo: silt loam; Lo: loam; Sa: sand; LoSa: loamy sand; SaLo: sandy loam. This figure is produced with R package soiltexture [44].

Various methods have been developed for CEC determination [45,46]. Many of these methods entail practical difficulties and are experimentally tedious and time-consuming [46]. The most common determination method might be the ammonium acetate (pH 7) displacement method, which has been widely used in different soil classification systems (Tables S1–S3). Measurement of the effective cation exchange capacity (ECEC) is required if the soil pH is less than 5.5, as significant quantities of exchangeable Al^{3+} may be present [11,46]. The ECEC values might be less than the CEC_{soil} [1], and both the CEC and ECEC might be required for acidic soil classification (Tables S1 and S3). In this study, the CEC_{soil} was determined by the ammonium acetate method (1 M NH_4OAc) at a pH of 7.0 [47]. Specifically, the CEC of clay (CEC_{clay}) and silt (CEC_{silt}) were analyzed based on the clay (<2 μm) and silt (2–50 μm) fractions, respectively. The carbonate, iron and SOM were removed for the clay and silt fractions by using $NaOAc$, $Na_2S_2O_4$ - $Na_3C_6H_5O_7$ and H_2O_2 solutions [11], respectively. A detailed introduction of the analytical procedure is provided by [11]. The clay and silt contents were determined by the pipette method [3]. The SOC, pH and Fe_d were measured by the Walkley-Black wet oxidation method [48], the potentiometer method (soil:water = 1:2.5) and the phenanthroline colorimetry method [47], respectively.

SOM contributes the major portion to the negative variable charge [1,49,50]. Therefore, to quantify the effects of negative variable charges on CEC_{soil} and CEC_{clay} measurements, the CEC of mineral fractions (<2 mm) was also analyzed by removing the SOM, and this value was referred to as the CEC_{Min} .

2.3. Environmental Variables

Recent PTF studies have suggested that the combination of environmental variables and soil data can improve PTFs [34,51]. Therefore, ten environmental variables were collected from [41] and considered as covariates. Terrain attributes were generated from the Shuttle Radar Topography Mission digital elevation model (SRTM DEM) [52] using SAGA-8 GIS software (<http://saga-gis.org/>; accessed on 11 April 2023), including elevation, slope, multiresolution ridge top flatness index (MRRTF), topographic wetness index (TWI) and stream power index (SPI). Climatic variables (i.e., MAP and MAT) were obtained from the WorldClim2 [40]. Categorical variables including the land use, parent material and soil type were based on the field descriptions. If the CEC_{clay} values were significantly different under the categorical variables, these variables were transformed into dummy variables to indicate the presence or absence of each type.

Four covariate datasets were produced based on the available soil data and environmental variables in this study:

- Dataset 1: Relatively accessible soil data (i.e., clay, silt, pH, SOC, Fe_d and CEC_{soil}), as these attributes are commonly available in soil databases and have been frequently used in modeling the CEC_{soil} [18,21,24,53];
- Dataset 2: All soil data (i.e., dataset 1 and CEC_{silt} and CEC_{Min});
- Dataset 3: Relatively accessible soil data and ten environmental variables; and
- Dataset 4: All soil data and ten environmental variables.

Predictor selection can extract relevant information and important features from variables and thus benefit prediction accuracy [54]. Therefore, predictors were considered as potential covariates if they were significantly correlated with soil properties of interest ($p < 0.05$) and would not involve multicollinearity [55]. Regarding the number of available predictors, predictors whose variance inflation factors were greater than 2.5 were removed to avoid excessive multicollinearity [56].

2.4. Prediction Methods

The advantage of machine learning techniques for CEC_{soil} modeling has been widely confirmed [18,23,57]. In contrast to PTFs that can be expressed as general equations [24], machine learning models rely on learning information from data, rather than on pre-determined equations, and can suitably quantify the nonlinear behaviors of soil data. Eight predictive models were used to estimate the CEC_{clay} based on the four covariate datasets, in-

cluding four machine learning models (ANN, DBN, SVR and RF), MLR and three published PTFs [3,38,39] (Figure 3).

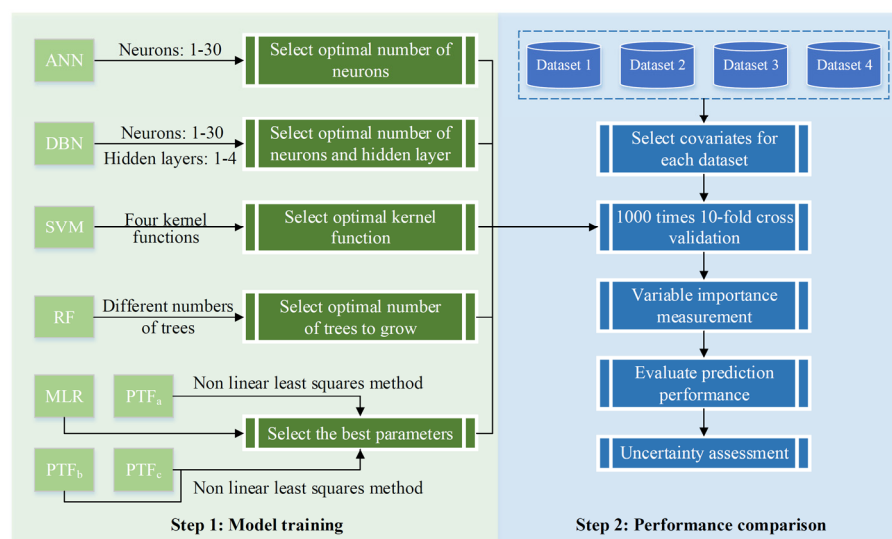


Figure 3. Workflow of the CEC_{clay} PTF fitting. PTF_a is the conventional estimation method based on the division of the CEC_{soil} by the clay content. PTF_b and PTF_c are the published PTFs in [38] (in the Brazilian soil classification system) and [39], respectively.

Inspired by biological neural networks, an ANN has artificial neurons and can model the nonlinear relationships between inputs and outputs by weighting these neurons. The resilient backpropagation algorithm with or without weight backtracking was considered to optimize the ANN model training, in which three layers were included, namely, the input, hidden and output layers. To pursue a robust prediction, the hyperbolic tangent and logistic sigmoid activation functions were applied for the ANN algorithms.

Deep learning models with multiple neural networks have been proposed by [58] and are receiving ever-growing attention in the field of machine learning. DBN techniques can extract deep features of samples through the training of several restricted Boltzmann machines and may achieve the accurate mapping of nonlinear features. This technique has not been attempted for CEC_{clay} PTF fitting to the best of our knowledge. In this study, a DBN with four hidden layers was trained due to limited soil data, in which the number of neurons ranged from 1 to 30. For each covariate dataset, a total of 10,000 DBN models was calibrated.

As an ensemble learning algorithm, an RF can create a large number of decision trees for classification or regression [59]. Based on bootstrap sampling, an RF is robust and less sensitive to overfitting. This method is usually adopted as a predictive tool but not a descriptive tool for quantifying the relationships between a variable of interest and covariates.

An SVR may generate a reliable estimate for regression analysis by minimizing the upper-bound generalization errors. This technique has been frequently considered for PTF fitting [23,60]. In this study, four kernel functions (i.e., linear, polynomial, radial basis and sigmoid) were calibrated for SVR model training, and the most accurate function was adopted.

The conventional estimation method based on the division of the CEC_{soil} by the clay content was adopted for comparison and was referred to as PTF_a. The second CEC_{clay} PTF (PTF_b) was published in the Brazilian soil classification system [38] based on soil data from a national soil survey. PTF_b was calculated from the CEC_{soil} by subtracting the contribution of the SOC as follows:

$$CEC_{clay} = CEC_{soil} - \frac{(4.5 \times SOC) \times 10}{Clay} \quad (1)$$

where the *SOC* and *Clay* are the soil organic carbon concentration (g kg^{-1}) and clay content (%), respectively.

A nonlinear CEC_{soil} PTF (PTF_c) was considered [39], which was based on a dataset with clay contents ranging from 2% to 78% that was consistent with our dataset (Figure 2). PTF_c could account for the interactions between clay and SOC:

$$\text{CEC}_{\text{clay}} = a + b \times \text{Clay} + c \times \text{Clay} \times \text{SOC} \quad (2)$$

where a , b and c are regression coefficients, and *Clay* and *SOC* are clay content (%) and soil organic carbon concentration (g kg^{-1}), respectively.

2.5. Variable Importance Measurement

The mean decrease in accuracy (MDA) of the RF models and the regression coefficients of the MLR models were utilized to measure the relative importance of the adopted predictors. The MDA was calculated by permuting the out-of-bag samples generated by the trees of the RF models [59]. The predictors for the MLR models were normalized by the Z-score method, of which the mean and standard deviation were 0 and 1, respectively. Therefore, the predictors were on the same magnitude, and their regression coefficients could be used to indicate the relative importance [61,62]. The greater the magnitude of the MDAs and regression coefficients, the more important the examined variable.

2.6. Statistical Analysis

A one-way analysis of variance (ANOVA) was performed to identify the effects of the land use, parent material and soil type on the CEC_{clay} , followed by a least-significant-difference (LSD) test ($p < 0.05$), which were implemented with R packages *stats* [63] and *agricolae* [64], respectively. The extraction of environmental variables at sampling sites was achieved with ArcGIS 10.2 (ESRI Inc., Redlands, CA, USA). A Pearson correlation was performed to depict the linear relationships between the soil properties and environmental variables. A paired t-test analysis will be performed to examine the difference of predictions from two methods with almost same prediction accuracy. The Pearson correlation and paired t-test analysis were conducted with R package *stats* [63]. Data processing, descriptive analysis, correlation analysis between the soil properties and PTF fitting were carried out with R (version 3.6.0, <https://cran.r-project.org/>, accessed on 14 May 2022). The nonlinear regression was fitted by the Levenberg-Marquardt nonlinear least squares method with the *minpack.lm* R package (version 1.2-1) [65]. The predictive methods, MLR, ANN, DBN, SVR and RF, were implemented with the R packages *stats* [63], *RSNNS* [66], *h2o* [67], *e1071* [68] and *randomForest* [69], respectively.

2.7. Model Validation

Regarding the limited number of soil samples, a 10-fold cross validation was used to evaluate the predictive models in terms of the root mean squared error (RMSE) and coefficient of determination (R^2). The 10-fold cross validation was performed 1000 times for the 8 models and PTFs, and the mean values of the validation results were used. The standard deviations of the RMSE and R^2 values were calculated to account for the uncertainty involved in the random sample splitting and model training. The 1000 iterations of model training were enough, as the mean values of the validation indices did not change if the models or PTFs were trained further.

3. Results

3.1. Soil Properties

The descriptive statistics of the soil properties are shown in Table 2.

Table 2. Descriptive statistics of the soil properties of the B horizons.

| Soil Property | Min | 25th Percentile | Mean | Median | 75th Percentile | Max | Standard Deviation | Skewness | Coefficient of Variation |
|---|------|-----------------|-------|--------|-----------------|--------|--------------------|----------|--------------------------|
| CEC _{clay} (cmol(+) kg ⁻¹) | 7.51 | 17.17 | 20.93 | 20.76 | 25.08 | 32.79 | 5.67 | 0.11 | 27.09 |
| CEC _{silt} (cmol(+) kg ⁻¹) | 0.48 | 2.25 | 4.96 | 3.59 | 6.66 | 21.25 | 3.85 | 1.65 | 77.62 |
| CEC _{soil} (cmol(+) kg ⁻¹) | 5.12 | 9.16 | 12.44 | 11.61 | 14.84 | 27.13 | 4.50 | 1.02 | 36.17 |
| CEC _{Min} (cmol(+) kg ⁻¹) | 6.88 | 10.85 | 14.55 | 14.09 | 17.31 | 29.83 | 4.89 | 0.84 | 33.61 |
| pH | 3.73 | 4.69 | 5.23 | 5.00 | 5.58 | 7.65 | 0.78 | 1.11 | 14.91 |
| SOC (g kg ⁻¹) | 1.40 | 2.72 | 4.72 | 3.80 | 5.49 | 14.05 | 2.88 | 1.56 | 61.02 |
| Fe _d (g kg ⁻¹) | 9.11 | 42.59 | 62.72 | 56.36 | 83.96 | 142.56 | 26.36 | 0.51 | 42.03 |
| Clay (%) | 9.48 | 30.80 | 40.96 | 39.90 | 49.75 | 71.92 | 14.82 | 0.35 | 36.18 |
| Silt (%) | 8.36 | 21.89 | 31.22 | 31.00 | 39.70 | 66.10 | 11.62 | 0.35 | 37.22 |

In general, the values of the coefficient of variation revealed moderate variation. The mean value of the silt content (31.22%) was less than that of the clay content (40.96%), while the mean value of the CEC_{clay} (20.93 cmol(+) kg⁻¹) was much greater than that of the CEC_{silt} (4.96 cmol(+) kg⁻¹). The B horizons in this study area were characterized by low SOC concentrations (4.72 g kg⁻¹) and CEC_{soil} values (12.44 cmol(+) kg⁻¹) but high Fe_d levels (62.72 g kg⁻¹). The paired *t*-test analysis showed that CEC_{soil} was significantly differed with CEC_{Min} ($p < 0.001$). The mean value of the CEC_{soil} (12.44 cmol(+) kg⁻¹) was slightly less than that of the CEC_{Min} (14.55 cmol(+) kg⁻¹).

The CEC_{clay} was significantly correlated with the CEC_{soil} ($r = 0.580$) and CEC_{Min} ($r = 0.525$), while the CEC_{soil} was strongly positively correlated with the CEC_{Min} ($r = 0.920$) (Tables 3 and S4). The clay content was positively correlated with the CEC_{Min} ($r = 0.237$) but was negatively correlated with the CEC_{clay} ($r = -0.263$) and CEC_{silt} ($r = -0.337$). These correlations were generally consistent with the results of previous studies on humid soils [53], whereas the correlation between the clay and CEC_{soil} was weaker than those of a global soil dataset [18].

Table 3. Pearson correlation coefficients between soil properties.

| | CEC _{silt} | CEC _{soil} | CEC _{Min} | pH | SOC | Fe _d | Silt | Clay |
|---------------------|---------------------|---------------------|--------------------|----------|--------|-----------------|----------|-----------|
| CEC _{clay} | 0.272 ** | 0.580 ** | 0.525 ** | 0.121 * | -0.067 | -0.283 ** | 0.424 ** | -0.263 ** |
| CEC _{silt} | 1 | 0.600 ** | 0.459 ** | 0.068 | 0.073 | 0.140 | 0.108 | -0.337 ** |
| CEC _{soil} | | 1 | 0.920 ** | 0.279 ** | 0.048 | 0.258 ** | 0.160 | 0.065 |
| CEC _{Min} | | | 1 | 0.238 ** | 0.004 | 0.298 ** | 0.085 | 0.237 ** |
| pH | | | | 1 | -0.001 | 0.048 | 0.225 * | -0.067 |
| SOC | | | | | 1 | 0.188 * | 0.032 | -0.171 |
| Fe _d | | | | | | 1 | -0.130 | 0.541 ** |
| Silt | | | | | | | 1 | -0.352 ** |

* Significant at the 0.05 level. ** Significant at the 0.01 level.

3.2. Model Training

ANOVA analysis showed that the CEC_{clay} was significantly different under agroecosystem (i.e., upland) and natural ecosystems (i.e., grassland and forest). CEC_{clay} did not significantly differ relative to the soil type and parent material type. Therefore, the land use was considered as the covariate, in which grassland and forest were combined together. Four ANN algorithms with one hidden layer were compared for the four datasets (Figures 4 and S1). The logistic sigmoid activation function-based networks (ANN2 and ANN4) yielded less errors than did the other functions (Figure 4). We also evaluated the neural networks with two hidden layers (Figures 5 and S2). However, the prediction accuracy was not improved, and thus, the ANN model with one hidden layer was adopted.

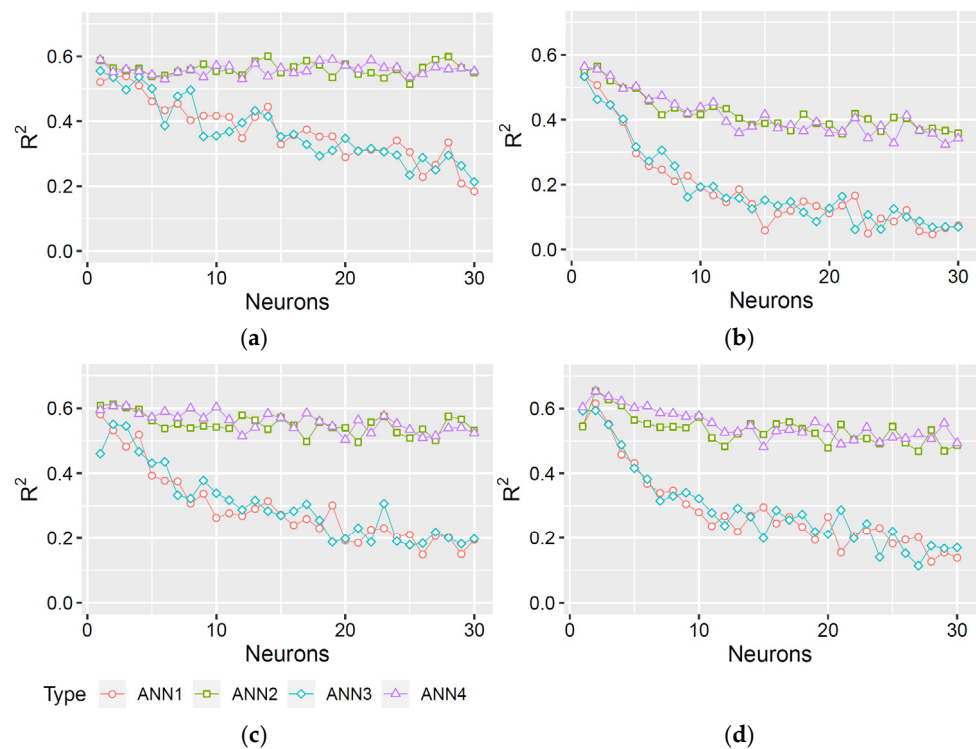


Figure 4. R^2 values of ANN models with one hidden layer for dataset 1 (a), 2 (b), 3 (c) and 4 (d), of which the number of neurons of hidden layer ranged from 1 to 30. ANN1 and ANN2 refer to the resilient backpropagation algorithm with weight backtracking using the hyperbolic tangent and logistic sigmoid activation functions, respectively. ANN3 and ANN4 refer to the resilient backpropagation algorithm without weight backtracking using the hyperbolic tangent and logistic sigmoid activation functions, respectively.

The prediction accuracy of the DBN with one hidden layer increased with the increase of the number of neurons and became stable when the number of neurons was about 30 (Figure S3). Regarding the high computation cost, the DBN model was trained with four hidden layers for dataset 1 (Figure 6), in which the number of neurons of each hidden layer ranged from 1 to 30. After the comparison of the prediction accuracy, a deep neural network with the structure of (30, 24, 27, 12) was selected.

SVR models based on the linear kernel function outperformed SVR models based on the other three kernel types in terms of greater R^2 and lower RMSE values (Figure 7), which suggested that a linear decision boundary benefited the separation of feature points. It was inferred that the relationships between the CEC_{clay} and covariates could well be quantified by linear regression.

The mean squared errors of RF models rapidly decreased with increasing numbers of trees from 1 to 50 (Figure 8). The errors became relatively stable from 50 to 300, even if some fluctuations could be observed when the number of trees was greater than 300. Thus, the number of trees in the RF models was set to 300.

For MLR, we conducted a residual analysis for homogeneity of variance (Figures S4–S7). The residuals were normally distributed via Q-Q plots. There was no obvious distinct pattern of the residuals versus the fitted values, indicating a linear relationship between the CEC_{clay} and the employed predictors. Exceptions were the single cases that were outside of Cook's distance (Figures S6 and S7) when using environmental variables as covariates.

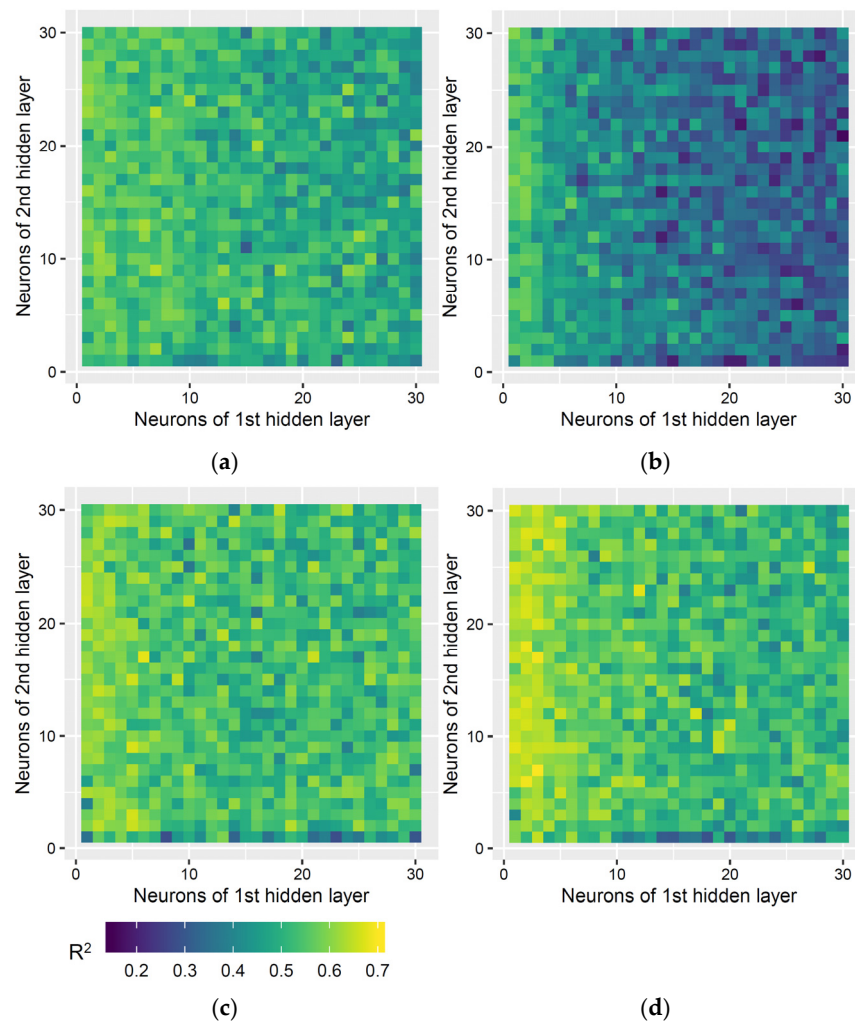


Figure 5. R² values of ANN models with two hidden layers for dataset 1 (a), 2 (b), 3 (c) and 4 (d), in which the number of neurons of the first and second hidden layers ranged from 1 to 30.

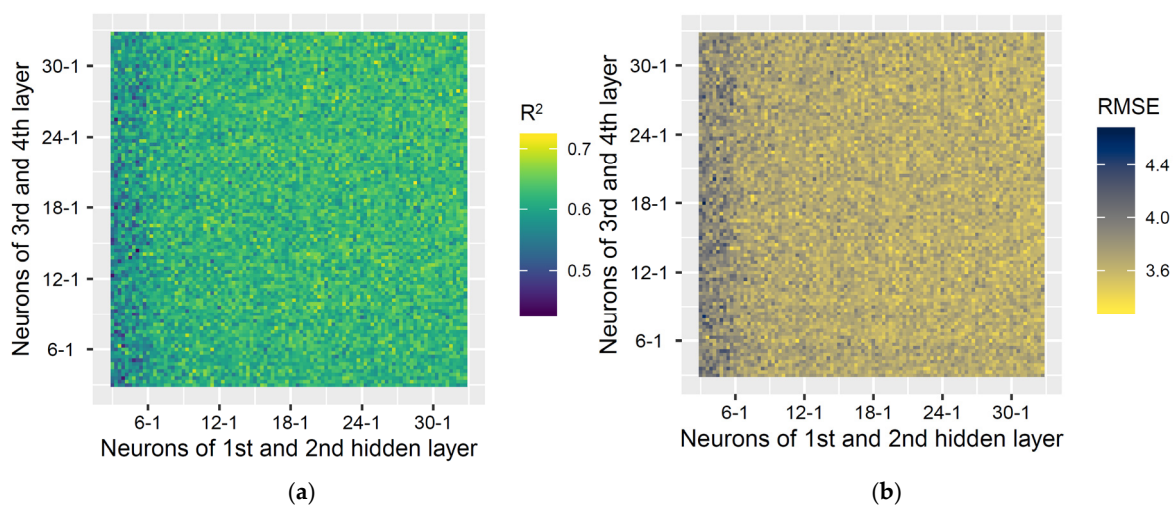


Figure 6. R² (a) and RMSE (b) values of DBN models with four hidden layers for dataset 1. The numbers before and after a dash on the x-axis denote the numbers of neurons in the first and second hidden layers, respectively, which ranged from 1 to 30. The numbers of neurons in the third and fourth layers are shown on the y-axis.

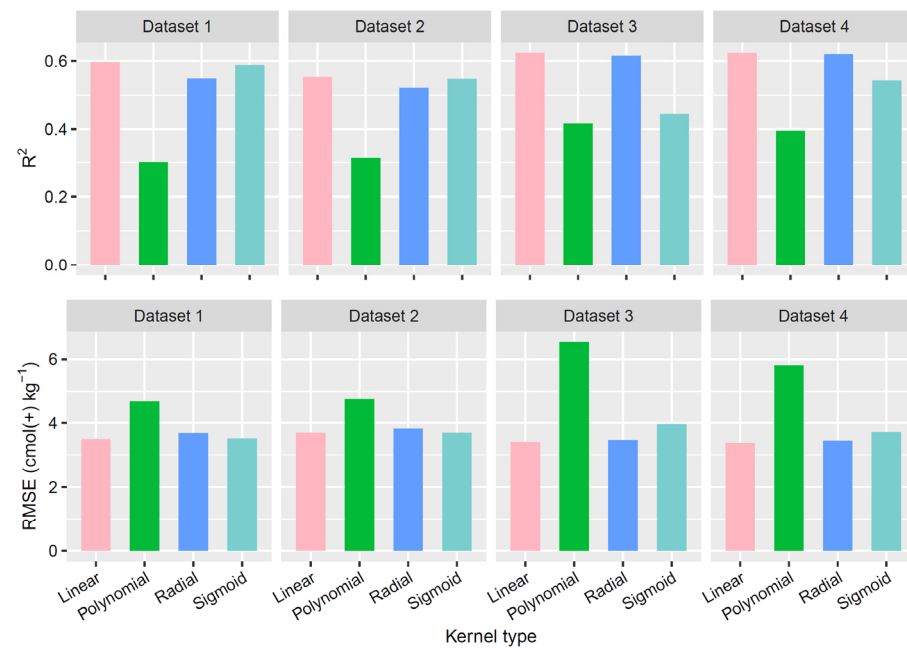


Figure 7. R^2 and RMSE of the SVR models for the four datasets based on different kernel types.

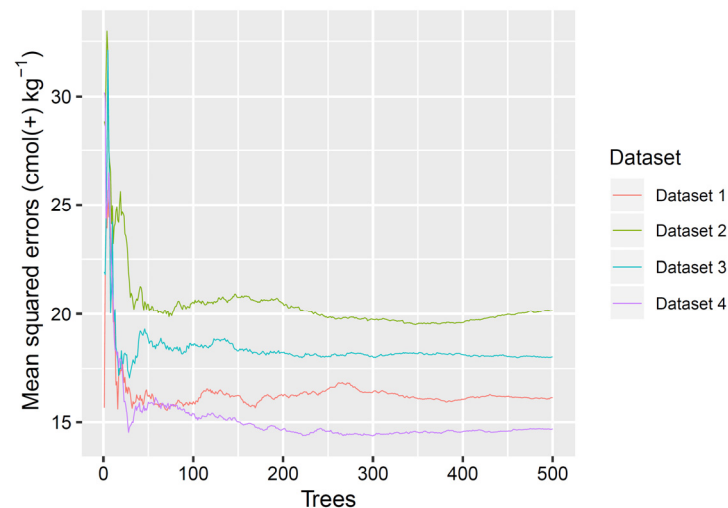


Figure 8. Plots of the mean squared errors of the RF models for the four datasets.

3.3. Variable Importance

The relative importance of the considered covariates was measured by a nonlinear model (i.e., RF) and linear regression (i.e., MLR) in terms of the MDA and regression coefficients, respectively (Figure 9). The magnitude of a regression coefficient accounted for the importance of a variable, regardless of the positive and negative values. Overall, the relative importance of different covariates was similar in both modeling cases. The CEC_{soil} was the most important variable, except for dataset 2, for which the most important variable was the CEC_{Min} . Environmental variables played moderate roles in PTF fitting, and the land use was more important than were most of the soil properties (Figure 9g,h).

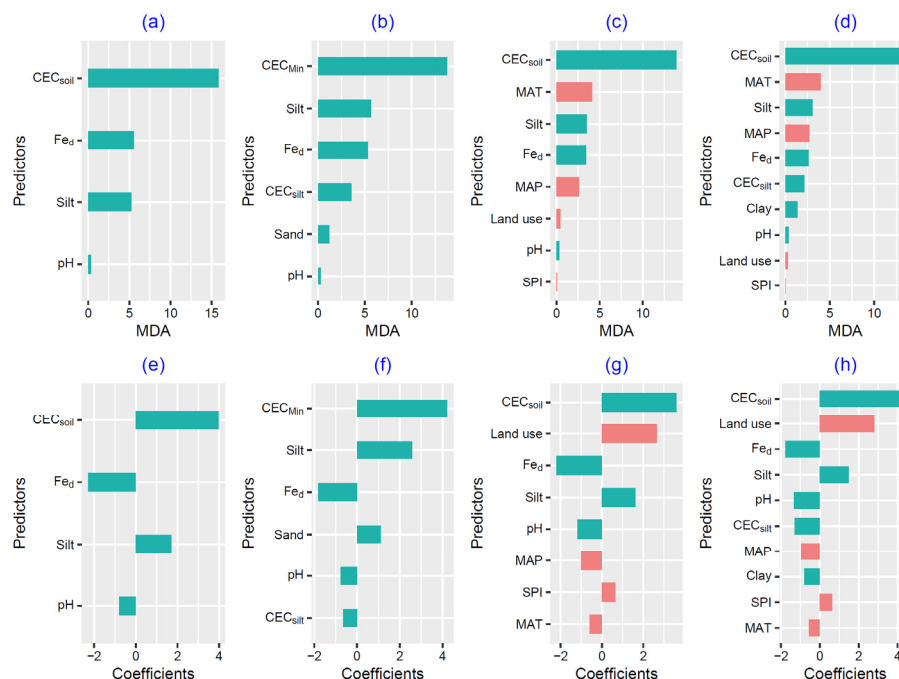


Figure 9. Relative importance of the employed covariates for datasets 1 (a,e), 2 (b,f), 3 (c,g) and 4 (d,h). The values of the mean decrease in accuracy (MDA) (a–d) were derived from RF models. The regression coefficients (e–h) were calculated by MLR models. Bars in sea green and red denote the soil properties and environmental variables, respectively. Land use is a dummy variable representing the presence or absence of upland. CEC_{Min}: CEC of mineral fractions (<2 mm); CEC_{silt}: CEC of the silt fraction; CEC_{soil}: CEC of the fine earth fraction (<2 mm); Fe_d: concentration of free iron oxides; MAP: mean annual precipitation; MAT: mean annual air temperature; SPI: stream power index.

3.4. Performance Comparison

Based on the optimized parameters above and selected predictors (Figure 9), eight models and PTFs were executed 1000 times and evaluated in terms of the R² and RMSE (Table 4).

Table 4. Performance assessment (R² and RMSE) of eight considered models and PTFs. For each prediction case, the mean values and standard deviations of R² and RMSE based on 1000 runs are shown.

| Dataset | ANN | DBN | SVR | RF | MLR | PTF _a | PTF _b | PTF _c |
|----------------|-------------|-------------|-------------|-------------|-------------|------------------|------------------|------------------|
| R ² | | | | | | | | |
| Dataset 1 | 0.63 ± 0.03 | 0.63 ± 0.02 | 0.57 ± 0.02 | 0.59 ± 0.02 | 0.65 ± 0.02 | 0.41 ± 0.03 | 0.24 ± 0.03 | 0.15 ± 0.04 |
| Dataset 2 | 0.59 ± 0.03 | 0.58 ± 0.03 | 0.52 ± 0.02 | 0.55 ± 0.02 | 0.63 ± 0.02 | 0.41 ± 0.02 | 0.25 ± 0.04 | 0.14 ± 0.02 |
| Dataset 3 | 0.54 ± 0.06 | 0.64 ± 0.04 | 0.61 ± 0.02 | 0.62 ± 0.02 | 0.70 ± 0.02 | 0.41 ± 0.03 | 0.25 ± 0.04 | 0.15 ± 0.03 |
| Dataset 4 | 0.64 ± 0.04 | 0.64 ± 0.03 | 0.61 ± 0.02 | 0.63 ± 0.02 | 0.71 ± 0.02 | 0.41 ± 0.02 | 0.25 ± 0.05 | 0.14 ± 0.03 |
| RMSE | | | | | | | | |
| Dataset 1 | 3.56 ± 0.15 | 3.90 ± 0.20 | 3.87 ± 0.06 | 3.87 ± 0.07 | 3.46 ± 0.05 | 22.78 ± 0.41 | 16.75 ± 0.08 | 5.50 ± 0.04 |
| Dataset 2 | 3.74 ± 0.12 | 4.07 ± 0.19 | 4.65 ± 0.07 | 4.00 ± 0.07 | 3.64 ± 0.06 | 22.62 ± 0.43 | 16.75 ± 0.08 | 5.48 ± 0.04 |
| Dataset 3 | 5.06 ± 1.67 | 3.75 ± 0.22 | 3.64 ± 0.10 | 3.78 ± 0.06 | 3.29 ± 0.06 | 22.64 ± 0.44 | 16.74 ± 0.08 | 5.49 ± 0.03 |
| Dataset 4 | 3.55 ± 0.20 | 3.75 ± 0.16 | 3.61 ± 0.09 | 3.73 ± 0.05 | 3.21 ± 0.08 | 22.64 ± 0.34 | 16.74 ± 0.08 | 5.48 ± 0.04 |

The predictive accuracy of machine learning models and MLR was far superior to those of existing PTFs, i.e., PTF_a, PTF_b and PTF_c. Meanwhile, the performance was generally improved when using environmental variables as covariates (Figure 10 and Table 4), in which the RMSEs approximately decreased by 2–10%. MLR models produced more accurate results than those of other methods, with R² values ranging from 0.63 to 0.71. The prediction accuracy of the SVR was same as that of MLR when not using the environmental variables. A paired *t*-test analysis showed that predictions based on the

SVR significantly differed from those based on MLR ($p < 0.05$). Four fitted equations of MLR using all observations were expressed as follows, in which the predictors were not normalized:

$$\left\{ \begin{array}{l} \text{Dataset 1: } CEC_{clay} = 16.108 + 0.881 \times CEC_{soil} - 0.086 \times Fe_d - 1.019 \times pH + 0.147 \times Silt \\ \text{Dataset 2: } CEC_{clay} = 9.798 + 0.862 \times CEC_{Min} - 0.166 \times CEC_{silt} - 0.069 \times Fe_d - 0.977 \times pH + 0.07 \times Sand + 0.221 \times Silt \\ \text{Dataset 3: } CEC_{clay} = 26.10 + 0.804 \times CEC_{soil} - 0.083 \times Fe_d + 2.662 \times Landuse - 0.0003 \times MAP - 0.002 \times MAT \\ \quad - 1.523 \times pH + 0.139 \times Silt + 0.00005 \times SPI \\ \text{Dataset 4: } CEC_{clay} = 27.575 - 0.337 \times CEC_{silt} + 0.978 \times CEC_{soil} - 0.054 \times Clay - 0.067 \times Fe_d + 2.805 \times Landuse \\ \quad - 0.0003 \times MAP - 0.002 \times MAT - 1.7 \times pH + 0.127 \times Silt + 0.00005 \times SPI \end{array} \right. \quad (3)$$

where CEC_{soil} is CEC of the fine earth fraction ($\text{cmol}(+) \text{kg}^{-1}$), CEC_{silt} is CEC of the silt fraction ($\text{cmol}(+) \text{kg}^{-1}$), CEC_{Min} is CEC of mineral fractions ($\text{cmol}(+) \text{kg}^{-1}$), Fe_d is the concentration of free iron oxides (g kg^{-1}), pH is soil acidity, $Silt$ and $Sand$ are clay content and sand content (%), respectively, $Landuse$ is a dummy variable representing the presence or absence of upland, MAP is mean annual precipitation (mm), MAT is mean annual air temperature ($^{\circ}\text{C}$), and SPI is stream power index.

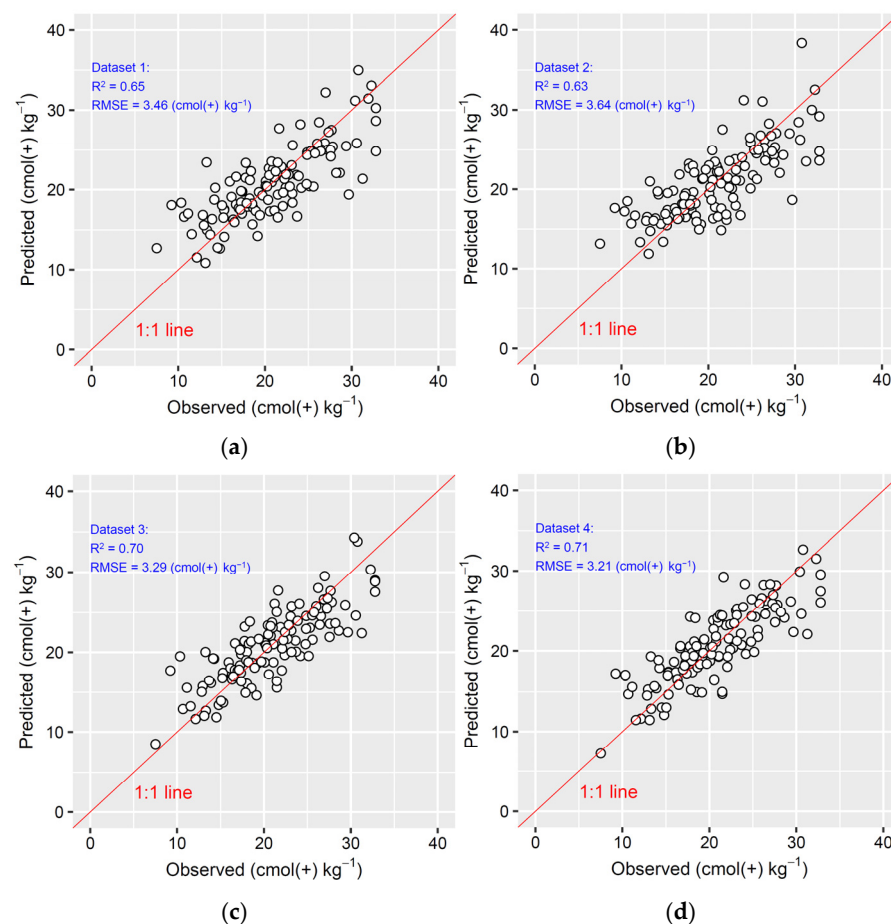


Figure 10. Scatter plots of the observed and predicted CEC_{clay} values based on MLR for datasets 1 (a), 2 (b), 3 (c) and 4 (d).

Notably, both the PTFs defined in ST [3] (i.e., PTF_a) and the Brazilian soil classification system [38] (i.e., PTF_b) performed worse than did other methods. R^2 values suggested that PTF_a and PTF_b were inferior, and the simply dividing of CEC_{soil} by clay content (PTF_a) greatly overestimated the CEC_{clay} (Figure 11a). Furthermore, we fitted a linear model to improve the current calculation method: $CEC_{clay} = 15.31 + 15.90 \times (CEC_{soil}/Clay)$ with R^2

of 0.38 (Figure 11). This equation was recommended for B horizons in subtropical regions for simplification when soil variables in Equation (3) were limited.

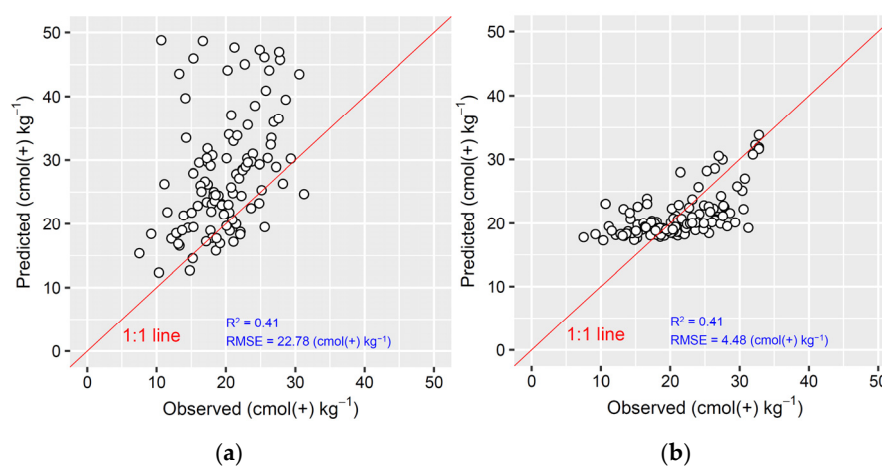


Figure 11. Scatter plots of the observed and predicted CEC_{clay} values based on PTF_a (a) and revised formula of $15.31 + 15.90 \times PTF_a$ (b).

4. Discussion

4.1. Performance of PTF Models

The machine learning techniques achieved promising accuracy in CEC_{clay} prediction (Table 4). However, explicit equations were unavailable, and these models should be trained by certain observations for other areas, which may limit the application of machine learning techniques for PTF fitting in practice. We found that RF outperformed SVR for all datasets (Table 4), which was in agreement with similar study [70]. Overall, the MLR outperformed other considered machine learning techniques and nonlinear models (Table 4), which did not concur with related studies [22,23]. This could be possibly ascribed to the limited soil samples in this study ($n = 122$). These results not only raised concerns regarding these published PTFs' limitations but also implied that a linear regression might be sufficient for modeling the CEC_{clay} of the B horizon in humid subtropical regions, which accounted for 56–65% of the variation. Therefore, the first hypothesis should be rejected. The accuracy was in agreement with the proposed CEC_{soil} PTFs, such as the fitted PTFs in Iran, with R^2 values of 0.59–0.60 [71] and 0.48–0.60 [20]. Furthermore, the PTFs for CEC_{silt} , CEC_{soil} and CEC_{Min} were also fitted (Supplementary Results), of which R^2 ranged from 0.21 to 0.86 (Table S5).

In contrast to the machine learning and MLR techniques, the published PTFs (PTF_a and PTF_b) failed to accurately predict the CEC_{clay} . This suggested that the contributions of other soil properties (i.e., Fe_d and silt) to the CEC_{soil} were far from negligible in South China, as well as in semiarid regions [72]. As a new remarkable technique, deep learning failed to improve accuracy (Table 4). This result could be attributed to the heterogeneous features of the soil properties, and a hybrid modeling technique should be explored by extracting the multi-scale features from a larger dataset [73].

Compared with PTF fitting at fixed sampling depth increments [19,21,24], the soil samples used in this study were collected from genetic horizons with variable thicknesses, which concurred with other studies conducted at the national scale [18,61,74]. The soil properties in genetic horizons are normally homogeneous compared to those collected at fixed depth increments [75].

There are several limitations of the PTFs training. More soil properties and environmental variables should be collected to pursue a PTF with higher accuracy than current methods (Tables 2 and 4). Furthermore, PTFs and machine learning models might be reliant on land use types, especially for study area at a large scale [70]. Due to data limitations, the performance of PTFs regarding land use types was not tested in this study, in which

only 122 samples were analyzed from genetic B horizons rather than from the whole soil profiles. Future works should focus on the CEC of the sand fraction, the ECEC of the clay fraction ($<2 \mu\text{m}$) and the assessment of the accompanying uncertainty of PTF fitting. These shortcomings should be considered in the near future when extra experimental soil data are collected. We also conducted additional experiments by partitioning the soil dataset at the SOC break of 3 g kg^{-1} [74]. Due to the limited soil points, a leave-one-out cross validation was performed. Overall, the stratification did not improve the estimate accuracy. CEC_{clay} data based on laboratory measurements are usually very limited. An independent validation will be required to evaluate the generalization of the proposed PTFs at various geographical locations (Equation (3)).

4.2. Importance of Predictors

Different from the CEC_{soil} PTFs fitted in the United States [74], Denmark [61] and Spain [72], the clay and SOC were not selected as effective predictors in this case (Figure 9). The clay contents were significantly negatively correlated with the CEC_{clay} ($r = -0.263$), but the correlation was less than that between the silt and CEC_{clay} ($r = 0.352$). Therefore, silt was frequently selected as the covariate (Figure 9). The CEC_{soil} exhibited a high correlation with the CEC_{clay} ($r = 0.58$), which could be ascribed to the main contribution of the clay minerals to the CEC_{soil} in humid soils, as vermiculite, kaolinite and hematite were the main clay minerals from north to south in the current study area [15]. The Fe_d played a vital role in the prediction scenarios based on soil properties (Figure 9). In general, iron oxides usually exist in the clay fraction or are strongly cemented with clays [76,77] and thus greatly affect the stabilization of the soil structure in terms of the aggregate stability and clay dispersion. Notably, some soil properties that are not readily obtainable, that is, the CEC_{silt} and CEC_{Min} , were also selected as helpful predictors (Figure 9b,f).

The warm and moist climate in South China strengthens the soil weathering and base cation leaching. Thus, in addition to the soil properties, climatic variables (i.e., MAP and MAT) were included in the models as important predictors, suggesting that the second hypothesis can be accepted. Specifically, terrain attributes were dispensable in predicting the CEC_{clay} due to the low correlations (Table S4). This could be explained by the slight effect of the topography on deep soil evolution.

Significant differences were found in the CEC_{clay} regarding land use. The inclusion of land use as predictor obviously benefited the prediction (Table 4 and Figure 9). It was suggested that soil variability under uplands was different from that under other land uses. Interestingly, the relative importance of land use was more evident with the MLR model than with the RF model (Figure 9), which may imply that the CEC_{clay} could be well described as a linear function based on the soil properties and soil-forming factors (Table 4 and Equation (3)).

4.3. Determination of the CEC_{clay}

Our study area was mainly characterized by acidic soils (Table 2). Thus, the effects of variable charges (or pH-dependent charges) on the CEC_{soil} and CEC_{clay} should be discussed, as variable charges might be neutralized by the standard method, that is, the ammonium acetate method ($1 \text{ M NH}_4\text{OAc}$) at a pH of 7.0 [11]. It is well known that organic matter materials and iron and aluminum oxides account for major portions of the negative and positive variable charges, respectively [49]. The CEC_{soil} is a measure of the quantity of negative charges that indicates the cations retained by electrostatic forces. The CEC of iron oxide concentrates is very low [78], even though iron oxides have relatively large surface areas [1]. Therefore, the contribution of negative variable charges to the CEC_{soil} or CEC_{clay} can be quantified by analyzing the correlation between the CEC_{soil} and CEC_{Min} , which was referred to as the CEC of the fine earth fraction ($<2 \text{ mm}$) after removing the SOM (Figure 12). SOM removal led to a slight increase of the CEC (Table 2), which concurred with the results of [49,78]. This suggested that variable charges of SOM did not make a major contribution to the CEC, possibly due to the low SOC concentrations of subsoils

(Table 2). It was also suggested that some charges could be blocked by the interaction of the SOM with the clay [49]. The high correlation ($r = 0.92$) showed evidence that variable charges did not greatly affect the measurement, and thus, the information on the CEC_{soil} and CEC_{clay} was credible.

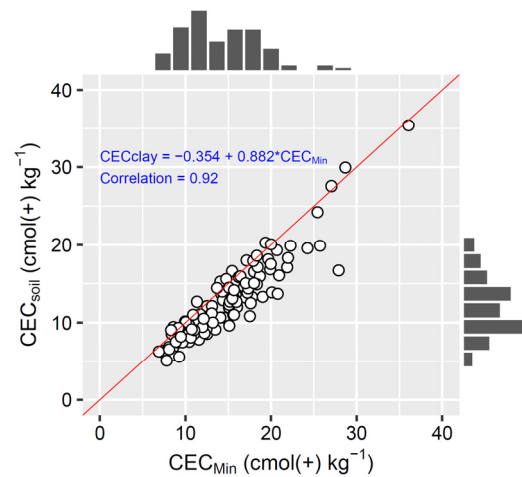


Figure 12. Relationships between the CEC_{Min} and CEC_{soil} , with the histograms exhibited at the top and right, respectively.

4.4. Implications of Using the CEC_{clay} for Soil Identification

PTF_a overestimated the CEC_{clay} with a mean error of $14.42 \text{ cmol}(+) \text{ kg}^{-1}$, and PTF_b produced a negative CEC_{clay} with a mean error of $-14.91 \text{ cmol}(+) \text{ kg}^{-1}$. The CEC_{clay} is usually adopted as a diagnostic criterion for soil taxonomy purposes (Tables S1–S3). It could be concluded that the ratio-based method (PTF_a) may affect the accuracy of the soil type allocation, especially when a pedon satisfies all required diagnostic characteristics except the CEC_{clay} [3–5].

With the aid of the developed PTF (Equation (3)), the value of the CEC_{clay} could be corrected to update the soil types of the NSSS database across China in turn [41]. We examined the CEC_{clay} and its influence on the soil taxonomy in Guangdong Province [79], in which many soil profiles were classified as Ferralosols. Argosols (which are mainly referred to as Alfisols in ST) and Cambosols (which are referred to as Inceptisols in ST) [3] with hues of 5YR or that were more intensely red and had $Fe_d \geq 14 \text{ g kg}^{-1}$ or DCB-extractable iron $\geq 40\%$ of the total iron were mainly considered (Table 5).

Here, four profiles that were classified as Argosols (the Liangtian and Jinji series) and Cambosols (the Datuo and Dengta series) are illustrated (Figure S8). Their diagnostic surface horizons were ochric epipedons. The diagnostic subsurface horizons of the Argosols and Cambosols were argic and cambic horizons, respectively. After updating the CEC_{clay} , 5 Argosols soil series and 4 Cambosols soil series satisfied the requirement of LAC ferric horizons ($CEC_{clay} < 24 \text{ cmol}(+) \text{ kg}^{-1}$) and were classified as Ferrosols. The use of the proposed equations improved the CST and seemed to be promising for soil classification in other areas. The fitted equation for dataset 1 (Equation (3)) might be helpful when other predictors are unavailable.

Table 5. Information on the soil series in Guangdong Province, of which the soil type could be classified as Ferrosols.

| Original Soil Suborder (CST) | Soil Series | Location | Layer (cm) | Hue | Fe _d (g kg ⁻¹) | Fe Freeness (%) | CEC _{clay} Based on PTF _a (cmol(+) kg ⁻¹) | Revised CEC _{clay} (cmol(+) kg ⁻¹) |
|------------------------------|---------------|---------------------------|------------|-------|---------------------------------------|-----------------|---|---|
| Udic Argosols | Dingbao | 22°19'25" N, 111°01'44" E | 32–96 | 7.5YR | 55.0 | 53.5 | 28.1 | 22.15 |
| | Jinji | 22°11'43" N, 12°28'41" E | 31–120 | 2.5Y | 24.7 | 65.7 | 27.2 | 16.30 |
| | Liangtian | 23°33'21" N, 115°50'49" E | 13–57 | 7.5YR | 64.3 | 61.2 | 28.1 | 23.67 |
| | Shangzhongben | 25°06'20" N, 113°31'54" E | 10–40 | 7.5YR | 75.6 | 70.7 | 27.2 | 22.50 |
| | Wenfu | 24°42'40" N, 116°11'19" E | 16–22 | 2.5YR | 40.2 | 61.3 | 26.6 | 14.54 |
| Udic Cambosols | Beidou | 23°49'43" N, 116°07'43" E | 15–55 | 5Y | 46.3 | 68.3 | 28.1 | 23.50 |
| | Dengta | 24°00'37" N, 114°46'57" E | 11–23 | 2.5YR | 42.2 | 62.9 | 25.2 | 21.55 |
| | Datuo | 24°33'35" N, 115°55'42" E | 14–29 | 10YR | 52.0 | 72.7 | 27.9 | 23.33 |
| | Xiajiashan | 23°15'41" N, 116°11'13" E | 9–25 | 10YR | 27.8 | 47.3 | 25.5 | 19.55 |

5. Conclusions

Eight models and PTFs were evaluated for CEC_{clay} prediction based on a 10-fold cross validation. The simply dividing CEC_{soil} by clay content greatly overestimated the CEC_{clay}, with a mean error of 14.42 cmol(+) kg⁻¹. MLR outperformed the other methods, with R² of 0.63–0.71 and RMSE of 3.21–3.64 cmol(+) kg⁻¹. The prediction accuracy of the SVR was the same as that of MLR when not using the environmental variables. The use of environmental variables obviously improved the model fit. Given certain calibration samples, machine learning techniques are promising for use in establishing an accurate model through fitting. For simplification, the MLR PTFs are recommended in practice, as an explicit equation and an enhanced prediction performance can be attained. We propose using the new PTFs to generate more accurate CEC_{clay} values with acceptable time and cost investments. Notably, the proposed PTFs should be validated further if they are used for taxonomic classification in other regions.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy14112671/s1>, Supplementary Results; Figure S1: RMSE values of ANN models with 1 hidden layer for datasets 1 (a), 2 (a), 3 (c) and 4 (d), in which the number of neurons in the hidden layer ranged from 1 to 30. ANN1 and ANN2 refer to the resilient backpropagation algorithm with weight backtracking using hyperbolic tangent and logistic sigmoid activation functions, respectively. ANN3 and ANN4 refer to the resilient backpropagation algorithm without weight backtracking using hyperbolic tangent and logistic sigmoid activation functions, respectively; Figure S2: RMSE values of ANN models with 2 hidden layers for datasets 1 (a), 2 (b), 3 (c) and 4 (d), in which the number of neurons in the first and second hidden layer ranged from 1 to 30; Figure S3: R² (a) and RMSE (b) values of DBN models with 1 hidden layer for dataset 1; Figure S4: Residuals distribution check of dataset 1; Figure S5: Residuals distribution check of dataset 2; Figure S6: Residuals distribution check of dataset 3; Figure S7: Residuals distribution check of dataset 4; Figure S8: Representative pedons of four soil series, of which the soil type should be referenced to Ferrosols: Liangtian (a), Jinji (b), Datuo (c), and Dengta (d). The land use type of Liangtian and Datuo is forest, and that of Jinji and Dengta is upland; Table S1: Main soil types and diagnostic horizons with the Soil Taxonomy, in which the values of the CEC_{clay} (by NH₄OAc pH 7) and ECEC are required; Table S2: Reference soil groups and qualifiers in the World Reference Base for

Soil Resources (WRB), in which the values of the CEC_{clay} (by NH_4OAc pH 7) are required; Table S3: Diagnostic subsurface horizons and diagnostic characteristics in the Chinese Soil Taxonomy (CST), in which the values of the CEC_{clay} (by NH_4OAc pH 7) and ECEC are required; Table S4: Correlation coefficients between soil properties and environmental variables; Table S5: Performance assessment (R^2 and RMSE) of multiple linear regression for the PTFs of CEC of silt (CEC_{silt}), CEC of the fine earth fraction (<2 mm) (CEC_{soil}) and the CEC of mineral fractions (CEC_{Min}). For each prediction case, the mean values and standard deviations of R^2 and RMSE based on 100 runs are shown.

Author Contributions: Conceptualization, J.Z. and Z.-X.S.; methodology, J.Z. and Z.-X.S.; software, J.Z.; validation, J.Z.; formal analysis, Z.-X.S.; investigation, Z.-X.S.; resources, Z.-X.S.; data curation, Z.-X.S.; writing—original draft preparation, J.Z.; writing—review and editing, Z.-X.S.; visualization, J.Z. and Z.-X.S.; supervision, J.Z. and Z.-X.S.; project administration, J.Z. and Z.-X.S.; funding acquisition, Z.-X.S. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by grants from the National Natural Science Foundation of China (No. 42277285), “Xing Liao Talent Plan” Youth Top Talent Support Program (XLYC2203085), the Natural Science Foundation of Jiangsu Province (No. BK20241070), and the Start-up Fund for New Talented Researchers of Nanjing Vocational University of Industry Technology (No. YK23-05-01).

Data Availability Statement: Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors would like to expressly acknowledge the contributions of all colleagues in China to the establishment of the soil series database. Our acknowledgments also extend to the anonymous reviewers for their constructive reviews of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arnold, R.; Eswaran, H.; Meyer, R. *Proceedings of a Symposium on Low Activity Clay (LAC) Soils*; Technical Monograph No. 14; U.S. Department of Agriculture: Las Vegas, NV, USA, 1984.
2. Demattê, J.A.M.; Nanni, M.R.; Formaggio, A.R.; Epiphanyo, J.C.N. Spectral reflectance for the mineralogical evaluation of Brazilian low clay activity soils. *Int. J. Remote Sens.* **2007**, *28*, 4537–4559. [[CrossRef](#)]
3. Soil Survey Staff. *Keys to Soil Taxonomy*, 12th ed.; USDA: Washington, DC, USA, 2014.
4. IUSS Working Group WRB. *World Reference Base for Soil Resources 2014*; Update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. World Soil Resources Reports, No. 106; FAO: Rome, Italy, 2015.
5. Gong, Z.T. *In Theory, Methodology and Application of Chinese Soil Taxonomy*; Science Press: Beijing, China, 1999. (In Chinese)
6. CSTC (Chinese Soil Taxonomic Classification Research Group). *Chinese Soil Taxonomy*; Science Press: Beijing, China; New York, NY, USA, 2001.
7. Caner, L.; Bourgeon, G.; Toutain, F.; Herbillon, A.J. Characteristics of non-allophanic Andisols derived from low-activity clay regoliths in the Nilgiri Hills (Southern India). *Eur. J. Soil Sci.* **2000**, *51*, 553–563. [[CrossRef](#)]
8. Prasetyo, B.H.; Suharta, N. Properties of low activity clay soils from South Kalimantan. *Indones. Soil Clim. J.* **2004**, *22*, 26–39. [[CrossRef](#)]
9. Láng, V.; Fuchs, M.; Szegi, T.; Csorba, Á.; Michéli, E. Deriving World Reference Base Reference Soil Groups from the prospective Global Soil Map product—A case study on major soil types of Africa. *Geoderma* **2016**, *263*, 226–233. [[CrossRef](#)]
10. Hengl, T.; de Jesus, J.M.; Heuvelink, G.B.M.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.Y.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)]
11. Soil Survey Staff. *Kellogg Soil Survey Laboratory Methods Manual*, version 5.0; Burt, R., Soil Survey Staff, Eds.; Soil Survey Investigations Report No. 42; U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA, 2014.
12. Zhang, M.; He, Z.; Milson, M.J. Chemical and physical characteristics of red soils from Zhejiang province, Southern China. In *The Red Soils of China*; Wilson, M.J., Ed.; Kluwer Academic Publishers: Norwell, MA, USA, 2004.
13. D’Angelo, B.; Bruand, A.; Qin, J.; Peng, X.; Hartmann, C.; Sun, B.; Hao, H.T.; Rozenbaum, O.; Muller, F. Origin of the high sensitivity of Chinese red clay soils to drought: Significance of the clay characteristics. *Geoderma* **2014**, *223–225*, 46–53. [[CrossRef](#)]
14. Li, Q.K.; Xiong, Y. *Soil of China*, 2nd ed.; Science Press: Beijing, China, 1990. (In Chinese)
15. Gong, Z.T.; Huang, R.J.; Zhang, G.L. *Soil Geography in China*; Science Press: Beijing, China, 2014. (In Chinese)
16. Schnitzer, M. Contribution of organic matter to the cation exchange capacity of soils. *Nature* **1965**, *207*, 667–668. [[CrossRef](#)]
17. Peinemann, N.; Amiotti, N.M.; Zalba, P.; Villamil, M.B. Effect of clay minerals and organic matter on the cation exchange capacity of silt fractions. *J. Plant Nutr. Soil Sci.* **2000**, *163*, 47–52. [[CrossRef](#)]
18. Khaledian, Y.; Brevik, E.C.; Pereira, P.; Cerdà, A.; Fattah, M.A.; Tazikeh, H. Modeling soil cation exchange capacity in multiple countries. *Catena* **2017**, *158*, 194–200. [[CrossRef](#)]

19. Shekofteh, H.; Ramazani, F.; Shirani, H. Optimal feature selection for predicting soil CEC: Comparing the hybrid of ant colony organization algorithm and adaptive network-based fuzzy system with multiple linear regression. *Geoderma* **2017**, *298*, 27–34. [CrossRef]
20. Khodaverdiloo, H.; Momtaz, H.; Liao, K.H. Performance of soil cation exchange capacity pedotransfer function as affected by the inputs and database size. *Clean-Soil Air Water* **2018**, *46*, 1700670. [CrossRef]
21. Sulieman, M.; Saeed, I.; Hassaballa, A.; Rodrigo-Comino, J. Modeling cation exchange capacity in multi geochronological-derived alluvium soils: An approach based on soil depth intervals. *Catena* **2018**, *167*, 327–339. [CrossRef]
22. Seyedmohammadi, J.; Esmaeelnejad, L.; Ramezanpour, H. Determination of a suitable model for prediction of soil cation exchange capacity. *Model. Earth Syst. Environ.* **2016**, *2*, 156. [CrossRef]
23. Shiri, J.; Keshavarzi, A.; Kisi, O.; Iturraran-Viveros, U.; Bagherzadeh, A.; Mousavi, R.; Karimi, S. Modeling soil cation exchange capacity using soil parameters: Assessing the heuristic models. *Comput. Electron. Agric.* **2017**, *135*, 242–251. [CrossRef]
24. Liao, K.; Xu, S.; Zhu, Q. Development of ensemble pedotransfer functions for cation exchange capacity of soils of Qingdao in China. *Soil Use Manag.* **2015**, *31*, 483–490. [CrossRef]
25. Kaya, N.S.; Dengiz, O. Assessment of the neutrosophic Fuzzy-AHP and predictive power of some machine learning approaches for maize silage soil quality. *Comput. Electron. Agric.* **2024**, *226*, 109446. [CrossRef]
26. Huang, Y.; Song, X.; Wang, Y.P.; Canadell, J.G.; Luo, Y.; Ciaia, P.; Chen, A.P.; Hong, S.B.; Wang, Y.G.; Tao, F.; et al. Size, distribution, and vulnerability of the global soil inorganic carbon. *Science* **2024**, *384*, 233–239. [CrossRef]
27. Shahabi, M.; Ghorbani, M.A.; Naganna, S.R.; Kim, S.; Hadi, S.J.; Inyurt, S.; Farooque, A.A.; Yaseen, Z.M. Integration of multiple models with hybrid artificial neural network-genetic algorithm for soil cation-exchange capacity prediction. *Complexity* **2022**, *2022*, 3123475. [CrossRef]
28. Saidi, S.; Ayoubi, S.; Shirvani, M.; Azizi, K.; Zhao, S. Digital mapping of soil phosphorous sorption parameters (PSPs) using environmental variables and machine learning algorithms. *Int. J. Digit. Earth* **2023**, *16*, 1752–1769. [CrossRef]
29. Wang, L.; Liu, D.; Sun, Y.; Zhang, Y.; Chen, W.; Yuan, Y.; Hu, S.; Li, S. Machine learning-based analysis of heavy metal contamination in Chinese lake basin sediments: Assessing influencing factors and policy implications. *Ecotoxicol. Environ. Saf.* **2024**, *283*, 116815. [CrossRef]
30. Emamgholizadeh, S.; Bazoobandi, A.; Mohammadi, B.; Ghorbani, H.; Sadeghi, M.A. Prediction of soil cation exchange capacity using enhanced machine learning approaches in the southern region of the Caspian Sea. *Ain Shams Eng. J.* **2023**, *14*, 101876. [CrossRef]
31. Sarkar, A.; Maity, P.P.; Ray, M.; Kundu, A. Inclusion of fractal dimension in machine learning models improves the prediction accuracy of hydraulic conductivity. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 4043–4067. [CrossRef]
32. Song, X.D.; Yang, F.; Wu, H.Y.; Zhang, J.; Li, D.C.; Liu, F.; Zhao, Y.G.; Yang, J.L.; Ju, B.; Cai, C.F.; et al. Significant loss of soil inorganic carbon at the continental scale. *Natl. Sci. Rev.* **2022**, *9*, nwab120. [CrossRef]
33. Song, X.D.; Alewell, C.; Borrelli, P.; Panagos, P.; Huang, Y.Y.; Wang, Y.; Wu, H.Y.; Yang, F.; Yang, S.H.; Sui, Y.Y.; et al. Pervasive soil phosphorus losses in terrestrial ecosystems in China. *Glob. Chang. Biol.* **2024**, *30*, e17108. [CrossRef]
34. Akpa, S.I.C.; Ugbaje, S.U.; Bishop, T.F.A.; Odeh, I.O.A. Enhancing pedotransfer functions with environmental data for estimating bulk density and effective cation exchange capacity in a data-sparse situation. *Soil Use Manag.* **2016**, *32*, 644–658. [CrossRef]
35. Zhao, D.; Li, N.; Zare, E.; Wang, J.; Triantafyllis, J. Mapping cation exchange capacity using a quasi-3d joint inversion of EM38 and EM31 data. *Soil Tillage Res.* **2020**, *200*, 104618. [CrossRef]
36. Ulusoy, Y.; Tekin, Y.; Tümsavaş, Z.; Mouazen, A.M. Prediction of soil cation exchange capacity using visible and near infrared spectroscopy. *Biosyst. Eng.* **2016**, *152*, 79–93. [CrossRef]
37. Zhao, X.Z.T.; Arshad, M.; Li, N.; Zare, E.; Triantafyllis, J. Determination of the optimal mathematical model, sample size, digital data and transect spacing to map CEC (Cation exchange capacity) in a sugarcane field. *Comput. Electron. Agric.* **2020**, *173*, 105436. [CrossRef]
38. Klamt, E.; Kauffman, J.H. *The Brazilian System of Soil Classification*; Research Report; ISRIC: Wageningen, The Netherlands, 1985.
39. McBratney, A.B.; Minasny, B.; Cattle, S.R.; Vervoort, R.W. From pedotransfer functions to soil inference systems. *Geoderma* **2002**, *109*, 41–73. [CrossRef]
40. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
41. Song, X.D.; Wu, H.Y.; Ju, B.; Liu, F.; Yang, F.; Li, D.C.; Zhao, Y.G.; Yang, J.L.; Zhang, G.L. Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma* **2020**, *363*, 114145. [CrossRef]
42. Sun, Z.X.; Jiang, Y.Y.; Wang, Q.B.; Sun, F.J.; Zhang, M.G.; Owens, P.R.; Libohova, Z. A comparative analysis between local soils and dust deposition on snow in Shenyang, China and implications on loess-paleosols evolution. *Geoderma* **2019**, *342*, 34–41. [CrossRef]
43. Schoeneberger, P.J.; Wysocki, D.A.; Benham, E.C.; Soil Survey Staff. *Field Book for Describing and Sampling Soils*, version 3.0; Natural Resources Conservation Service, National Soil Survey Center: Lincoln, NE, USA, 2012.
44. Moeys, J. soiltexture: Functions for Soil Texture Plot, Classification and Transformation. R Package Version 1.5.1. 2018. Available online: <https://CRAN.R-project.org/package=soiltexture> (accessed on 19 July 2023).
45. Ammann, L.; Bergaya, F.; Lagaly, G. Determination of the cation exchange capacity of clays with copper complexes revisited. *Clay Miner.* **2005**, *40*, 441–453. [CrossRef]

46. Jaremko, D.; Kalembasa, D. A comparison of methods for the determination of cation exchange capacity of soils. *Ecol. Chem. Eng. S* **2014**, *21*, 487–498. [CrossRef]
47. Zhang, G.L.; Gong, Z.T. *Soil Survey Laboratory Methods*; Science Press: Beijing, China, 2012. (In Chinese)
48. Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon and organic matter. In *Methods of Soil Analysis: Part 2: Chemical and Microbiological Properties*; Agronomy Monograph; Page, A.L., Miller, R.H., Keeney, D., Eds.; ASA and SSSA: Madison, WI, USA, 1982; Volume 9, pp. 539–579.
49. Tan, K.H.; Dowling, P.S. Effect of organic matter on CEC due to permanent and variable charges in selected temperate region soils. *Geoderma* **1984**, *32*, 89–101. [CrossRef]
50. Strahm, B.D.; Harrison, R.B. Mineral and organic matter controls on the sorption of macronutrient anions in variable-charge soils. *Soil Sci. Soc. Am. J.* **2007**, *71*, 1926–1933. [CrossRef]
51. Zolfaghari, Z.; Mosaddeghi, M.R.; Ayoubi, S. ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran. *Soil Use Manag.* **2015**, *31*, 142–154. [CrossRef]
52. Jarvis, A.; Reuter, H.I.; Nelson, A.; Guevara, E. Hole-Filled SRTM for Globe Version 4. Available from CGIAR-CSI SRTM 90 m Database. 2008. Available online: <http://srtm.csi.cgiar.org> (accessed on 11 September 2022).
53. Oorts, K.; Vanlauwe, B.; Merckx, R. Cation exchange capacities of soil organic matter fractions in a Ferric Lixisol with different organic matter inputs. *Agric. Ecosyst. Environ.* **2003**, *100*, 161–171. [CrossRef]
54. Zhang, F.; Yang, X. Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sens. Environ.* **2020**, *251*, 112105. [CrossRef]
55. Dobarco, M.R.; Cousin, I.; Bas, C.L.; Martin, M.P. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma* **2019**, *336*, 81–95. [CrossRef]
56. Johnston, R.; Jones, K.; Manley, D. Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behavior. *Qual. Quant.* **2018**, *52*, 1957–1976. [CrossRef] [PubMed]
57. Emamgolizadeh, S.; Bateni, S.M.; Shahsavani, D.; Ashrafi, T.; Ghorbani, H. Estimation of soil cation exchange capacity using genetic expression programming and multivariate adaptive regression splines. *J. Hydrol.* **2015**, *529*, 1590–1600. [CrossRef]
58. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
59. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
60. Bayat, H.; Asghari, S.; Rastgou, M.; Sheykhzadeh, G.R. Estimating Proctor parameters in agricultural soils in the Ardabil plain of Iran using support vector machines, artificial neural networks and regression methods. *Catena* **2020**, *189*, 104467. [CrossRef]
61. Krogh, L.H.; Breuning, M.; Greve, H.M. Cation-exchange capacity pedotransfer functions for Danish soils. *Acta Agric. Scand. Sect. B-Soil Plant Sci.* **2000**, *50*, 1–12. [CrossRef]
62. Zhang, G.; Liu, X.; Lu, S.; Zhang, J.; Wang, W. Occurrence of typical antibiotics in Nansi Lake's inflowing rivers and antibiotic source contribution to Nansi Lake based on principal component analysis-multiple linear regression model. *Chemosphere* **2020**, *242*, 125269. [CrossRef]
63. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: <https://www.R-project.org/> (accessed on 16 March 2023).
64. de Mendiburu, F. *Agricolae: Statistical Procedures for Agricultural Research*. R Package Version 1.3-5. 2021. Available online: <https://CRAN.R-project.org/package=agricolae> (accessed on 1 April 2023).
65. Elzhov, T.V.; Mullen, K.M.; Spiess, A.N.; Bolker, B. *minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds*. R Package Version 1.2-1. 2016. Available online: <https://CRAN.R-project.org/package=minpack.lm> (accessed on 14 May 2022).
66. Bergmeir, C.; Benitez, J.M. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS. *J. Stat. Softw.* **2012**, *46*, 1–26. [CrossRef]
67. LeDell, E.; Gill, N.; Aiello, S.; Fu, A.; Candel, A.; Click, C.; Kraljevic, T.; Nykodym, T.; Aboyou, P.; Kurka, M.; et al. *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*. R Package Version 3.30.0.1. 2020. Available online: <https://CRAN.R-project.org/package=h2o> (accessed on 14 May 2022).
68. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R Package Version 1.7-2. 2019. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 14 May 2022).
69. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
70. Mishra, G.; Sulieman, M.M.; Kaya, F.; Francaviglia, R.; Keshavarzi, A.; Bakhshandeh, E.; Loum, M.; Jangir, A.; Ahmed, I.; Elmobarak, A.; et al. Machine learning for cation exchange capacity prediction in different land uses. *Catena* **2022**, *216*, 106404. [CrossRef]
71. Bayat, H.; Davatgar, N.; Jalali, M. Prediction of CEC using fractal parameters by artificial neural networks. *Int. Agrophysics* **2014**, *28*, 143–152. [CrossRef]
72. Caravaca, F.; Albaladejo, A.L.J. Organic matter, nutrient contents and cation exchange capacity in fine fractions from semiarid calcareous soils. *Geoderma* **1999**, *93*, 161–176. [CrossRef]

73. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]
74. Seybold, C.A.; Grossman, R.B.; Reinsch, T.G. Predicting cation exchange capacity for soil survey using linear models. *Soil Sci. Soc. Am. J.* **2005**, *69*, 856–863. [[CrossRef](#)]
75. Parras-Alcántara, L.; Lozano-García, B.; Brevik, E.C.; Cerdá, A. Soil organic carbon stocks assessment in Mediterranean natural areas: A comparison of entire soil profiles and soil control sections. *J. Environ. Manag.* **2015**, *155*, 219–228. [[CrossRef](#)]
76. Hu, G.C.; Zhang, M.K. Mineralogical evidence for strong cementation of soil particles by iron oxides. *Chin. J. Soil Sci.* **2002**, *33*, 25–27. (In Chinese)
77. Martín-García, J.M.; Sánchez-Marañón, M.; Calero, J.; Aranda, V.; Delgado, G.; Delgado, R. Iron oxides and rare earth elements in the clay fractions of a soil chronosequence in southern Spain. *Eur. J. Soil Sci.* **2016**, *67*, 749–762. [[CrossRef](#)]
78. Ketrot, D.; Suddhiprakarn, A.; Kheoruenromne, I.; Singh, B. Interactive effects of iron oxides and organic matter on charge properties of red soils in Thailand. *Soil Res.* **2013**, *51*, 222–231. [[CrossRef](#)]
79. Lu, Y. *Guangdong Volume of Soil Series of China*; Science Press: Beijing, China, 2017. (In Chinese)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.