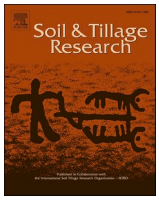
Soil & Tillage Research 244 (2024) 106225

Contents lists available at ScienceDirect

Soil & Tillage Research

journal homepage: www.elsevier.com/locate/still



Current limitations and future research needs for predicting soil precompression stress: A synthesis of available data

Lorena Chagas Torres [a,*], Attila Nemes [b,c], Loraine ten Damme [d], Thomas Keller [a,e]

[a] Swedish University of Agricultural Sciences, Department of Soil & Environment, Box 7014, SE-75007 Uppsala, Sweden
[b] Norwegian Institute of Bioeconomy Research, Division of Environment and Natural Resources, Frederik A. Dahls vei 43, N-1431 Ås, Norway
[c] Norwegian University of Life Sciences, Faculty of Environmental Sciences and Natural Resource Management, P.O. Box 5003, N-1432 Ås, Norway
[d] Aarhus University, Department of Agroecology, Blichers Allé 20, P.O. Box 50, DK-8830 Tjele, Denmark
[e] Agroscope, Department of Agroecology & Environment, Reckenholzstrasse 191, CH-8046 Zürich, Switzerland

ARTICLE INFO

Keywords:
Soil compaction
Soil compressive properties database
Oedometer
Pedotransfer function
Random forest

ABSTRACT

Precompression stress, compression index, and swelling index are used for characterizing the compressive behavior of soils, and are essential soil properties for establishing decision support tools to reduce the risk of soil compaction. Because measurements are time-consuming, soil compressive properties are often derived through pedotransfer functions. This study aimed to develop a comprehensive database of soil compressive properties with additional information on basic soil properties, site characteristics, and methodological aspects sourced from peer-reviewed literature, and to develop random forest models for predicting precompression stress using various subsets of the database. Our analysis illustrates that soil compressive properties data primarily originate from a limited number of countries. There is a predominance of precompression stress data, while little data on compression index or recompression index are available. Most precompression stress data were derived from the topsoils of conventionally tilled arable fields, which is not compatible with knowledge that subsoil compaction is a serious problem. The data compilation unveiled considerable variations in soil compression test procedures and methods for calculating precompression stress across different studies, and a concentration of data at soil moisture conditions at or above field capacity. The random forest models exhibited unsatisfactory predictive performance although they performed better than previously developed models. Models showed slight improvement in predictive power when the underlying data were restricted to a specific precompression stress calculation method. Although our database offers broader coverage of precompression stress data than previous studies, the lack of standardization in methodological procedures complicates the development of predictive models based on combined datasets. Methodological standardization and/or functions to translate results between methodologies are needed to ensure consistency and enable data comparison, to develop robust models for precompression stress predictions. Moreover, data across a wider range of soil moisture conditions are needed to characterize soil mechanical properties as a function of soil moisture, similar to soil hydraulic functions, and to develop models to predict the parameters of such soil mechanical functions.

1. Introduction

Soil compaction caused by agricultural field traffic is recognized as one of the most severe threats to soil quality in modern agriculture (FAO and ITPS, 2015; Stolte et al., 2016), and the negative impacts of soil compaction on crop productivity and soil functions are well documented (Nawaz et al., 2013; Graves et al., 2015; Alaoui and Diserens, 2018). The adverse impact of soil compaction, estimated to affect millions of hectares of arable land worldwide (Schjønning et al., 2016; Schneider and Don, 2019; Hu et al., 2021) is expected to worsen in the coming years due to the ongoing trend towards the use of heavier machinery, thereby increasing mechanical stress on the soil (Schjønning et al., 2015; Keller et al., 2019). Given the persistence of soil compaction for years or decades, particularly in the subsoil (Berisso et al., 2012), the environmental and economic consequences of soil compaction are long-lasting, affecting not only farmers but also the society at large. Restoring the structure of compacted soil through mechanical methods such as deep loosening is both energy and cost-intensive, and creates unstable soil

* Corresponding author.
E-mail address: lorena.chagas.torres@slu.se (L.C. Torres).

https://doi.org/10.1016/j.still.2024.106225
Received 18 April 2024; Received in revised form 24 June 2024; Accepted 30 June 2024
Available online 2 August 2024

conditions often with the risk of recompaction (Chamen et al., 2015; Spoor, 2006). Therefore, preventing soil compaction is the most effective approach to preserve soil functionality.

Understanding a soil's ability to resist compaction and determining stress limits to prevent compaction requires knowledge of the compressive behavior of the soil, which is characterized by a stress-strain curve. This curve represents the changes in the soil volume as a function of the applied stress and is commonly measured on cylindrical soil samples in confined uniaxial compression equipment, also referred to as an oedometer. The stress-strain curve provides three important soil compression properties: the swelling or recompression index, the compression index, and the precompression stress (Fig. 1). The compression index is considered an indicator of the soil's susceptibility to compaction or its resistance to compression (Imhoff et al., 2004), while the swelling or recompression index is used as a measure of rebound and as an indicator of the soil's mechanical resilience (Stone and Larson, 1980). The precompression stress has been considered as an indicator of the historical stress to which the soil had previously been subjected and of its load support capacity (Lebert and Horn, 1991).

According to theory, soil deformation is assumed to be elastic and recoverable if the stress applied is smaller than precompression stress. However, if the load applied produces a higher stress than precompression stress, it results in plastic deformation, leading to permanent soil compaction (Lebert and Horn, 1991). Thus, soil compaction caused by agricultural field traffic could in principle be prevented by ensuring that the exerted stress does not exceed the given soil's precompression stress. The precompression stress concept has been applied in various studies to derive soil strength and evaluate the risks of soil compaction (e.g. van den Akker, 2004; Lamandé et al., 2018; Kuhwald et al., 2022).

The labor-intensive nature of performing soil compression tests limits the application of soil precompression stress across a wide range of soil texture and soil moisture conditions. Consequently, precompression stress is often derived through pedotransfer functions (PTFs) (e.g. Lebert and Horn, 1991; Imhoff et al., 2004; Rücknagel et al., 2012; Severiano et al., 2013; Schjønning and Lamandé, 2018; Schjønning et al., 2023). These functions typically rely on readily available soil properties such as soil bulk density, soil organic carbon content, soil texture, and soil water content. However, some authors, such as Lebert and Horn (1991), also used less accessible predictors, including angle of internal friction, cohesion, and saturated hydraulic conductivity, complicating the use of PTFs due the data demand. Additionaly, these functions are typically developed for specific geographic regions, based on few data, restricting their broader applicability. Furthermore, when these functions are tested on independent data, they often exhibit relatively poor predictive performance.

Machine learning algorithms, such as random forest, have been employed by the soil science community to improve the prediction of soil properties that involve complex, non-linear relationships, which traditional statistical methods like multiple linear regression may not capture effectively. These applications include predicting soil organic carbon stock (Sarkodie et al., 2023), soil pH (Makungwe et al., 2021), soil bulk density (Palladino et al., 2022), soil hydraulic properties (Zhang et al., 2021), as well as digital soil mapping (van der Westhuizen et al., 2023). Furthermore, random forest models have been used to identify which co-variables influence soil water retention and flow (Koestel and Jorda, 2014; Gao et al., 2018). To predict soil precompression stress, random forest algorithms have been applied by Ebrahimzadeh et al. (2023) with soil properties and remote sensing data as predictors.

In this study, our aim was to develop a database of soil compressive properties with additional information on soil properties, site characteristics, and methodological aspects sourced from peer-reviewed literature. We present a synthesis of the available data, and develop predictive models, using the random forest technique to predict precompression stress from readily available soil and environmental information. Based on these analyses, we highlight current limitations to the development of predictive models, identify key knowledge gaps, and provide recommendations to advance future research.

2. Soil compressive properties database

2.1. Data collection

We searched published peer-reviewed journal articles in the databases Web of Science and Scopus in February 2022. The search terms used in the topic (title, abstract, keywords) were "soil precompression stress", "soil compression index", "soil compaction index", "soil recompression index", "soil swelling index", "soil precompaction stress" and "preconsolidation pressure". A total of 1235 publications were found. These references were added to the citation management application Endnote Web for removing duplicates and exported to the VOSviewer bibliometric mapping software (van Eck and Waltman, 2010) to create a network visualization of the most common terms used in the

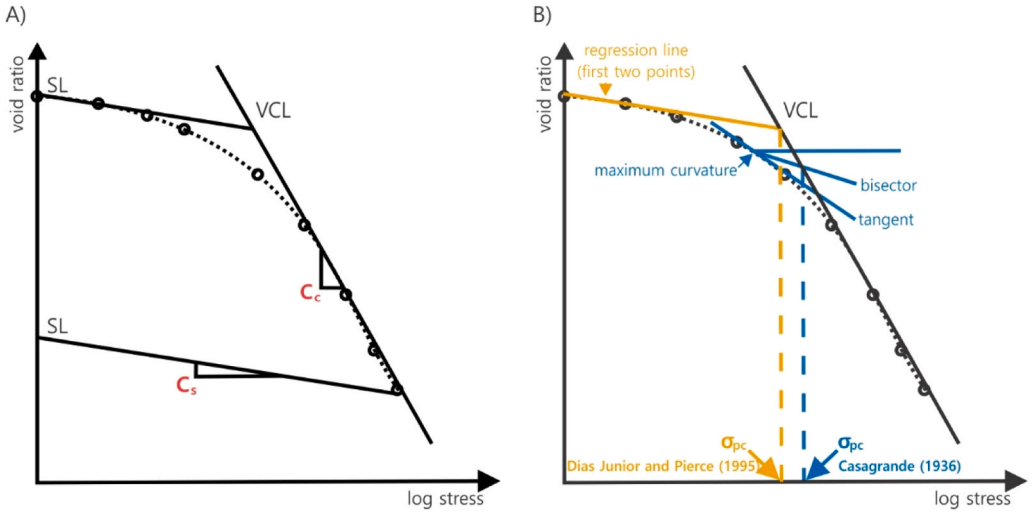


Fig. 1. Example of soil compression curve, expressed in terms of void ratio as a function of the logarithm of applied stress. Soil deformation is plastic, irreversible along the virgin compression line (VCL) and elastic and reversible along a swelling line (SL). The slope of the VCL is termed compression index, Cc; the slope of the SL is termed swelling index, Cs (A). Determination of precompression stress (σ_{pc}) according the Casagrande (1936) and Dias Junior and Pierce (1995) methods (B). Adapted from Keller et al. (2011).

selected studies. After removing duplicates (437 studies), the references were exported to the Rayyan software (Ouzzani et al., 2016) for screening by title and abstract based on the following pre-defined criteria: i) the article is peer-reviewed, with full text available, ii) soil compressive properties were derived from laboratory confined uniaxial compression tests on undisturbed soil samples, and iii) descriptive information about the soil(s) studied was provided. After these restrictions, 296 papers were selected for full-text reading. Language restriction was not applied during the selection of the studies, but only studies written in the Latin alphabet were selected. After a full-paper review, we identified 127 papers where the data on soil compressive properties were reported in numerical format or legible graphical format and considered suitable for inclusion in the database. Details of the data source papers, including the title, the authors, and the publication year are given in Supplementary Table S1.

For each study, we systematically recorded data on soil compressive properties, i.e. compression index, swelling index, and precompression stress, as well as descriptive data categorized into four groups: (1) static soil properties (e.g. soil texture), (2) soil moisture status indicators (i.e., water content and/or matric potential), (3) site characteristics, and (4) methodological aspects. In total, we compiled 4644 individual data entries.

In a number of cases, important information that was not presented in the paper was obtained directly from the authors. If more than one paper reported the same experiment, we selected the paper that provided more detailed information. However, we held on to both papers if they provided complementary information. For studies that compared several methods to calculate soil precompression stress, we only collected the precompression stress data calculated by the Casagrande (1936) method, which is widely accepted and used as a standard method.

The WebPlotDigitizer software (Rohatgi, 2015) was used to extract data from figures in original publications. The collected data were converted to the same unit (e.g., precompression stress in kPa, water content in kg kg^{-1}, matric potential in hPa, soil organic carbon in g kg^{-1}, etc.) to allow comparing data of different studies. The following calculations were performed to standardize the data when deemed necessary: i) in studies where only soil organic matter (SOM) was reported, the soil organic carbon (SOC) was obtained assuming that SOC was 58 % of SOM, ii) when only total soil porosity was provided, the soil bulk density was calculated assuming a soil particle density of 2.65 Mg m^{-3}, iii) when possible, soil water content data expressed on a volumetric basis were converted to gravimetric basis by dividing it by the corresponding soil bulk density, and iv) all texture data were standardized to the USDA classification system, which defines the silt/sand boundary at 50 μm, using a "k-nearest neighbor" type algorithm based on Nemes et al. (1999).

A simplified climatic classification of the study locations was performed using the latitude of each location. Areas located between the latitudes 0° - 23°, 23° - 35°, and 35° - 66° (north or south) were considered under tropical, subtropical, and temperate climate, respectively. Table 1 shows a more detailed description of the data compiled in this study. The complete database is published by Torres et al. (2023) and can be accessed at https://zenodo.org/records/10060810. The rights of use are defined by a creative commons license (CC BY 4.0).

Table 1
Overview of the included data in the soil compressive properties database.

Variable	Unit	Available	Missing	Description
Precompression stress	kPa	4674	69	-
Compression index	-	600	4143	-
Swelling index	-	149	4594	-
Clay	%	4566	177	-
Silt	%	4410	333	-
Sand	%	4410	333	-
Texture class	-	4410	333	Soil textural classification according USDA
Soil organic carbon	g kg^{-1}	2465	2278	-
Soil matric potential	-	2099	2644	Soil water matric potential expressed by log before loading
Soil water content	kg kg^{-1}	3056	1687	Gravimetric soil water content before loading
Bulk density	Mg m^{-3}	3528	1215	Soil bulk density before loading
Depth		4336	407	
Land use	-	4630	113	Land use classified into four categories: arable, forest, grassland, and native vegetation. The latter includes forest, grassland, and savanna
Tillage system	-	2900	521	Tillage system for arable soils, classified as "conventional" and "conservation"
Climate	-	4618	125	Climatic regions classified as temperate, tropical, subtropical
Calculation method of σ_{pc}	-	4642	101	Calculation method for estimating precompression stress (σ_{pc})
Minimum stress	kPa	4496	247	Minimum stress applied in soil compression tests
Maximum stress	kPa	4496	247	Maximum stress applied in soil compression tests
Stress application procedure	-	4707	36	Stress application type during soil compression tests classified as: 1=stepwise stress 2=one sample per stress 3=strain controlled
Number of steps	-	4073	553	Number of steps in stepwise stress application procedure
Loading time	min	1896	2736	Time for load application in each step in stepwise stress application and one sample per stress procedure
D/H sample ratio	-	3903	840	Ratio between diameter and height of the soil cores
Soil compression curve components	-	3901	842	Component of the soil compression curve related to the soil packing state: soil bulk density, void ratio, and strain.

2.2. Data analysis

We used descriptive statistics to summarize and diagnose the collected data. Initially, we assessed the distribution of data entries on soil compressive properties, i.e. compression index, swelling index, and precompression stress. Due to the limited availability of data on soil compression index and swelling index, subsequent analyses focused on precompression stress data. Specifically, we examined the geographical distribution and the representation of soil properties and conditions, land use, and soil management, the diversity of methodological procedures during uniaxial compression tests, and calculation methods employed to obtain precompression stress data.

2.3. Overview of entries included in the database

The 127 published studies included in the database cover the publication period from 1992 to 2021 (Supplementary Table S1). Data availability for soil compressive properties and descriptive data reported in each paper exhibited high variability. This resulted in a heterogeneous database with varying dataset sizes for different variables, as shown by available and missing data in Table 1.

2.3.1. Soil and site properties

In 98.5 % of data entries, precompression stress data were reported. However, only around 13 % and 3 % of the total data entries represent soil compression index and soil swelling index data, respectively. Our analysis included soils from 20 countries, with a significant concentration of the studies originating from Brazil (approx. 52 %), followed by Germany, Switzerland, Sweden, and Denmark collectively contributing with 32 % of the studies. More than 74 % of data entries are from Brazil, followed by Switzerland, Germany and Sweden, together accounting for approximately 19 % of data entries. Hence, more than 90 % of all data are from four countries only (Fig. 2). In temperate regions, Inceptisols and Alfisols dominated, together accounting for about 67 % of the data. In tropical and subtropical regions, predominantly represented by Brazilian data, Oxisols comprised approximately 57 % of the entries, followed by Ultisols with 20 %.

The contents of clay, silt, and sand ranged from 5 to more than 80 %, covering a broad range of soil textures (Fig. 3). Clay was the predominant textural class in the database, accounting for 41 % of the entries, followed by sandy clay loam, sandy loam, and silty loam with 20 %, 12 %, and 9 % of the entries, respectively (Fig. 3). The predominance of the texture class clay is mainly due to data from Brazil, where 95 % of the soils with clay texture originated. When examining the data of soil textural distribution from different climatic regions, it became apparent that the climatic regions represented distinct soil textural types. Soils from tropical and subtropical regions are characterized by very low silt content, whereas soils from temperate regions generally have greater silt content, occupying a distinctly different area of the USDA soil texture triangle (Fig. 4).

In most of the reviewed publications, several samples are associated with the same combination of clay, silt and sand content. As a result, different values of precompression stress exist for the same values of sand, silt, and clay, which explains the smaller number of data points displayed in the textural triangle (Fig. 4) compared to the number of data entries shown in Fig. 3.

The majority of soil bulk density values ranged from 1.2 to 1.7 Mg m^{-3}, and a great part of the soil samples (97 % of entries with available SOC data) had a soil organic carbon content of less than 50 g kg^{-1} (Fig. 3). The majority of data (72 %) are from arable soils, with 62 % of those data from conventionally-tilled fields, i.e. ploughed fields. Managed forest and grassland are the second and third most common land use type represented in the database, accounting for approx. 17 % and 6 %, respectively. Native vegetation accounts for approx. 3.5 % of the data. Approximately 71 % of the entries are from the top 30 cm of a soil (Fig. 3).

2.3.2. Soil conditions at the start of the compression tests

The initial soil water content was reported for 64 % of the precompression data entries, whereas only 44 % provided information on initial matric potential, expressed in our study by $pF=\log10(\psi\ [hPa])$, where ψ represents the soil matric potential. Only 18.8 % of the data entries had both the soil water content and the soil water matric potential. The gravimetric water content ranged from 0.02 to 0.89 kg kg^{-1},

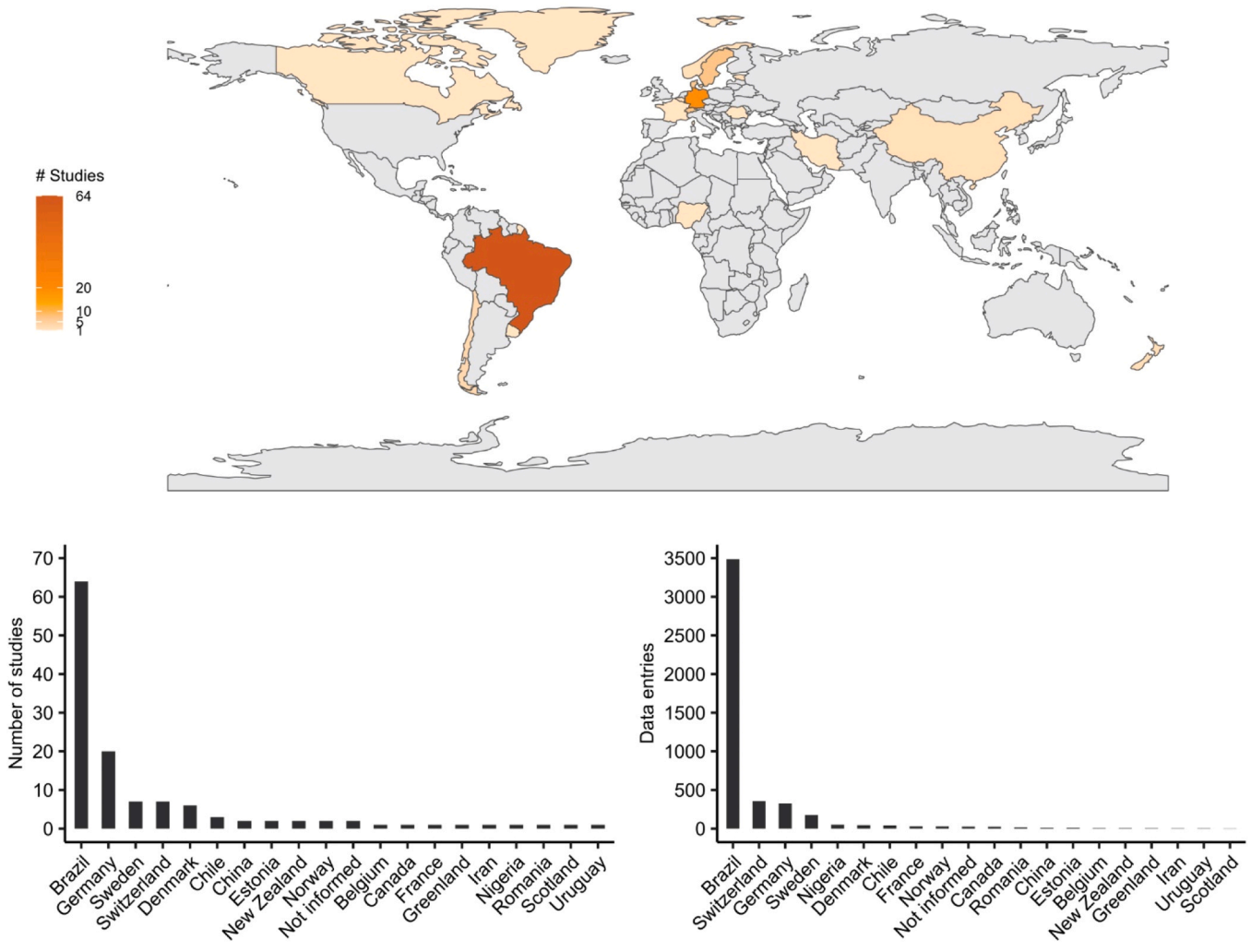


Fig. 2. Geographical distribution of studies and data entries for precompression stress included in the database.

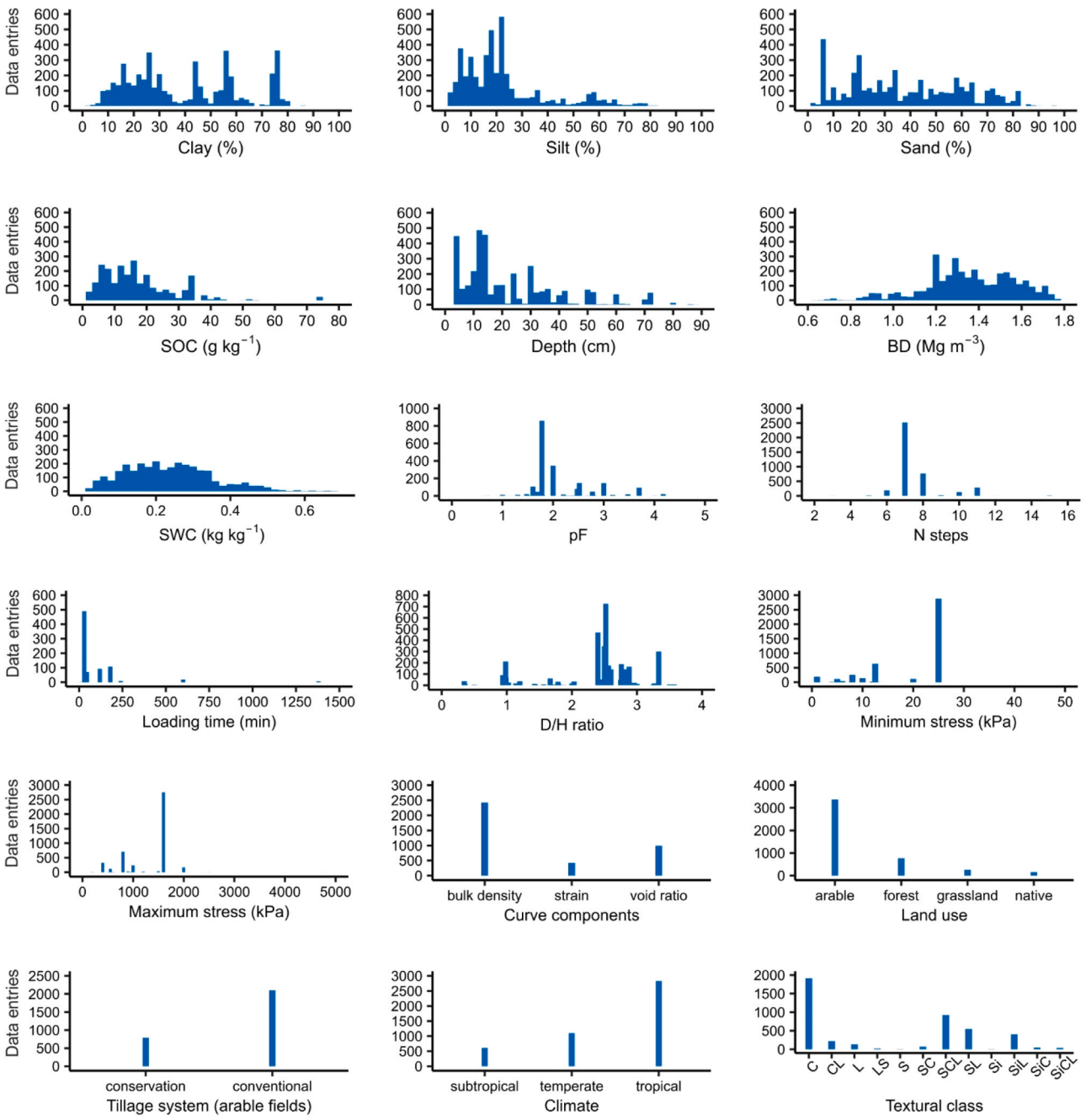


Fig. 3. Data distribution of supplementary variables for precompression stress data. BD: bulk density, SWC: soil water content, pF: log10(ψ [hPa]), N steps: number of steps in stepwise stress application procedure, D/H ratio: ratio between diameter and height of the soil cores, curve components: component of the soil compression curve related to the soil packing state, and textural class C: clay, CL: clay loam, L: loam, LS: loamy sand, S: sand, SC: sandy clay, SCL: sandy clay loam, SL: sandy loam, Si: silt, SiL: silt loam, SiC: silty clay, SiCL: silty clay loam.

while the matric potential ranged from saturation to wilting point (*p*F~4.2). The most frequent initial matric potential was *p*F 1.8 (-60 hPa), representing 42 % of the *p*F data entries, followed by *p*F 2 (-100 hPa), which accounted for 17 % of entries (Fig. 3). Approximately 83 % of the precompression stress values were measured on soil samples that were wetter than *p*F 2.5 (-330 hPa). These values often represent what is considered an approximation of matric potential at "field capacity" in various countries (e.g. *p*F 1.8 in Germany and Switzerland, *p*F 2.0 in most other European countries, and *p*F 2.5 in the US.).

2.3.3. *Methodological procedures during uniaxial compression tests*

The procedure used for stress application on the soil samples was mainly the stepwise stress application method (98 % of data entries), while the constant strain rate method was applied in only 1.6 % of the entries. In the stepwise stress application method, the number of steps ranged from 3 to 16, with most sequences having 7 steps. The loading time ranged from 0.5 min to 1380 min, with 5 min and 30 min being prevalent, representing 15.5 % and 10.5 % of the data, respectively (Fig. 3). For around 50 % of the precompression stress data entries, the stress was applied on soil samples until 90 % of the maximum deformation was reached. For these samples, no loading time was reported. The minimum stress applied during the compression tests was between 1 and 50 kPa, while the maximum applied stress ranged from 200 kPa to 4800 kPa. The most frequently applied maximum stresses were 1600 kPa with approximately 60 % of data entries, and 800 kPa, representing 15 % of the data (Fig. 3).

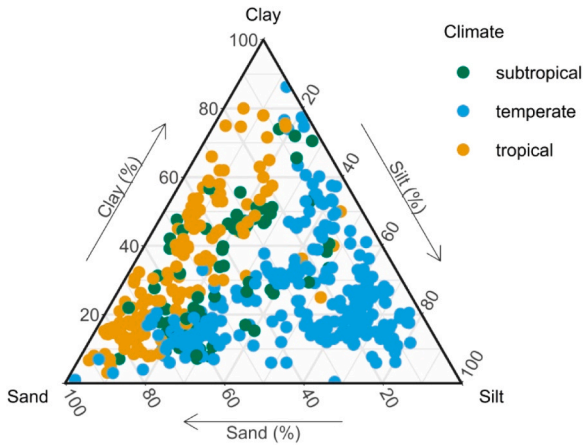


Fig. 4. Distributions of climatic regions in the texture triangle of the USDA system.

2.3.4. *Compression curve representation and precompression stress calculation methods*

The component of the compression curve relating to the soil packing state was most commonly represented by soil bulk density (approx. 52 %), followed by void ratio (approx. 22 %) and strain (approx. 9 %), which refers to the change in the soil sample height (settlement) relative to its original height (Fig. 3). The stress component of the compression curve was represented in a logarithmic form in the entirety of the entries.

The database comprised eight different methods for calculating precompression stress from the stress-strain curve (Table 2). The methods described by Dias Junior and Pierce, (1995) and Casagrande (1936) were most commonly used, representing 52 % and 36 % of the data entries, respectively. The method described by Lamandé et al. (2017) showed the lowest number of data entries, but it is also the most recently published method. The Casagrande (1936) method has been applied in studies across different climatic regions, including temperate, tropical, and subtropical areas. Other methods were associated with certain regions. For example, the Dias Junior and Pierce (1995) method was predominantly used in studies from the tropical region (Table 2).

Across all data, precompression stress ranged from approximately 5 to 1250 kPa (Fig. 5). Among the different precompression stress calculation methods, most showed similar distributions of precompression stress, with the majority of the entries with values for precompression stress smaller than 200 kPa. However, the Dias Junior and Pierce (1995) method showed a data concentration at higher values of precompression stress compared to those obtained with other methods (Fig. 5).

The initial moisture status was expressed in different units for the two most represented methods. The entries in which the Casagrande (1936) method was used, moisture status was predominantly expressed in terms of the soil's matric potential, while 90 % of the entries in which the method of Dias Junior and Pierce (1995) was used had it expressed

Table 2
Methods for calculation of precompression stress (σ_{pc}) and data entries for each climatic region included in the database.

σ_{pc} calculation method	Data entries		
	Temperate	Tropical	Subtropical
Casagrande (1936)	874	552	253
Dias Junior and Pierce (1995)	15	2029	277
Gregory et al. (2006)	103	210	0
ABNT (1990)	0	33	85
Casini (2012)	70	0	0
Culley and Larson (1987)	25	0	0
O'Sullivan and Robertson (1996)	0	10	0
Lamandé et al. (2017)	2	0	0

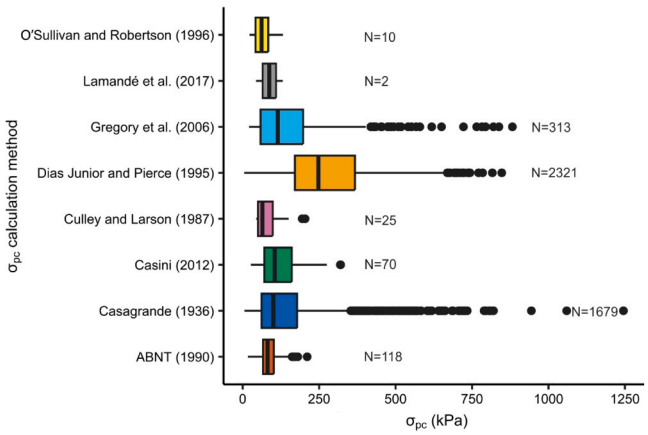


Fig. 5. Precompression stress (σ_{pc}) data entries through different methods of calculation represented by boxplots. The median is represented by the line in the box, the edges of the box indicates the quartiles, and the whiskers represents the variability of the data outside the upper and lower quartiles. Points that are located beyond the whiskers are often termed "outliers.".

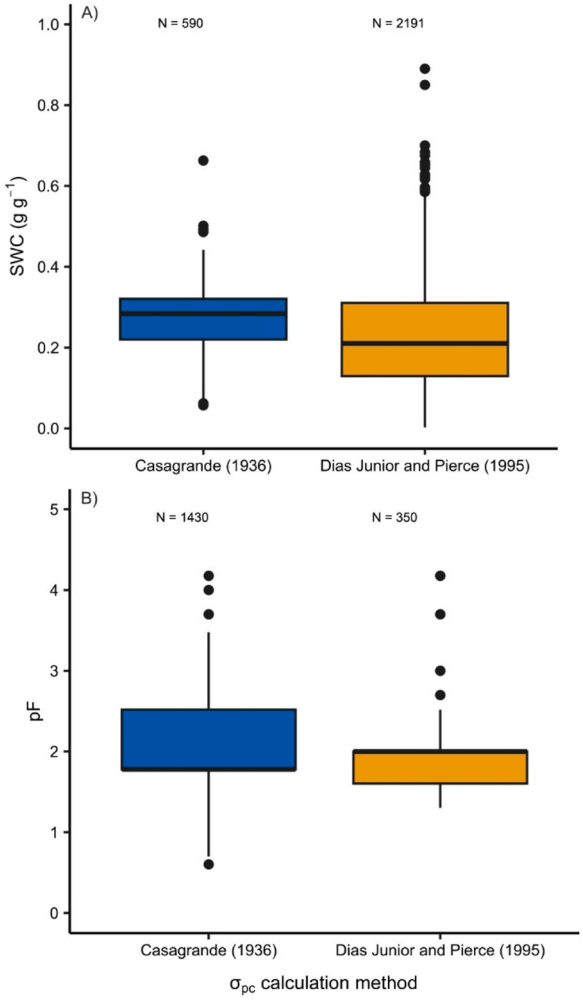


Fig. 6. Distribution of soil water content (A) and matric potential (B) data through the two most common methods for precompression stress (σ_{pc}) calculation in the database. N indicates the number of entries. The median is represented by the line in the box, the edges of the box indicates the quartiles, and the whiskers represents the variability of the data outside the upper and lower quartiles. Points that are located beyond the whiskers are often termed "outliers.".

as soil water content (Fig. 6).

3. Random forest models for predicting soil precompression stress

3.1. Random forest algorithm development

Random forest (RF) is a machine learning methodology designed for regression and classification problems, constituting an ensemble of classification and regression trees developed on randomized subsets of the data (Breiman, 2001). To develop each individual tree model, a subsample of data is selected randomly from the calibration dataset for which the algorithm finds the best set of predictor variables that help reduce variance in the output response within each tree node, and maximize the variance between nodes. In the RF approach, the number of trees and the number of predictor variables used are tuning parameters, i.e. parameters to be optimized. Due to the use of multiple trees and random sampling, RF is less susceptible to issues such as overfitting in the training process. Therefore, RF leads to robust results. In this study we used the *randomForest* package (Liaw and Wiener, 2002) in the R statistical software (version 4.3.2) to build the RF models.

Random forest models were developed using various subsets of the database, as detailed below. Each subset was randomly divided into training data (ca. 70 % of the data) and test data (ca. 30 % of the data) through a random sampling approach. When dividing the data, we considered that samples with identical soil properties were typically assessed in multiple steps of applied stress, which means that such observations in the database are co-dependent. Therefore, we randomized the data with the constraint that samples from the same study could not be in both the training and the test data set. As a result, the targeted 70–30 % ratio could not always be accurately met, but the actual proportions are being reported for each derived model (Table 3).

For each training dataset, a specific set of predictors (Table 3) was chosen to formulate the random forest model. The database contains numerous potential predictors, but because of missing data, there is a trade-off between the number of predictors and the size of the usable data set. Therefore, our choice of input variables was advised by their expectable influence on the estimated variable, and the intention to maximize data support to the resulting model.

Initially, we developed two distinct RF models, referred to here as RF1 and RF2, using largely identical predictors (clay, silt, sand, SOC, BD, SWC, *p*F, soil depth, method of calculation, and climate), with the only difference being that RF1 includes data entries with soil water content and RF2 includes data entries with soil matric potential as characterization of initial soil moisture status. Due to the limited overlap in the available soil moisture indicators, incorporating both such indicators would have resulted in a critically small number of selected samples.

Additional models, RF3 and RF4, were developed based on more selective subsets. In order to test our approach on more homogeneous data subsets, we limited the data to two subsets containing samples for distinct climatic regions (temperate vs. subtropical and tropical) and we limited the datasets to a single, dominant calculation method within each such subset. Due to data availability, this selection also defined two different soil moisture indicators to be used. As a result, the RF3 model was developed using data from subtropical and tropical regions from studies that adopted the precompression stress calculation method developed by Dias Junior and Pierce (1995) and used soil water content as moisture status indicator, while the RF4 model was built using data from the temperate region that adopted the Casagrande (1936) method for calculating precompression stress and used soil matric potential as moisture status indicator. Description of the subset and predictors to train and validate the random forest models are presented in Table 3. The distribution of data used to develop and train the random forest models are available in Supplementary material (Supplementary Figs. S1-S8).

The described set of predictors was used to train the RF model for each subset independently. The RF models were calibrated and internally assessed using 10-fold cross-validation. Evaluation metrics, such as the coefficient of determination (R^2) and root mean squared error (RMSE) were averaged across the 10 validation folds. The *mtry* hyperparameter, which indicates the number of variables considered at each split, was optimized based on minimizing the RMSE obtained on the validation portion of the data. The number of trees was set to 300. The results for RF models calibration are presented in Table 4.

To evaluate the importance of variables in the RF models, we used the increase in node purity as criterion that accounts for reduction in impurity in each tree node that results from splitting a particular variable. Variables that cause the greatest reduction in impurity are considered the most important.

Table 3
Subset and predictors details to train and validate the Random forest models.

Model	Subset description	Predictors	Training data	Test data
RF1	All database	Sand, silt, clay, organic carbon, bulk density, *water content*, depth, method of calculation, climate region, and land use	962 (16)*	249 (8)
RF2	All database	Sand, silt, clay, organic carbon, bulk density, *matric potential*, depth, method of calculation, climate region, and land use	552 (28)	204 (12)
RF3	Data from tropical and subtropical regions that applied Dias Junior and Pierce (1995) method for calculation of precompression stress	Sand, silt, clay, organic carbon, bulk density, *soil water content*, depth, and land use	491 (8)	491 (4)
RF4	Data from temperate region that applied Casagrande (1936) method for calculation of precompression stress	Sand, silt, clay, organic carbon, bulk density, *matric potential*, depth, and land use	285 (12)	40 (6)

* The number in parentheses indicates the quantity of studies from which the data originated.

3.2. Model evaluation using independent data

We evaluated the performance of the RF models on the independent test data set for each model using RMSE and R^2 as evaluation criteria. Recognizing the sensitivity of these metrics to outliers, we also incorporated the median absolute percentage error (MAPE) as an additional metric to evaluate the model performance. In addition to evaluating the models derived herein, we compared the performance of our models with some existing models that were developed from soils of similar climatic regions and that used a comparable set of input variables. Therefore, the RF3 model was compared with the pedotransfer function

Table 4
The optimized random forest parameter (*mtry*), mean for the squared correlation coefficient (R^2) and root mean square error (RMSE) across the 10-fold cross-validation.

Model	*mtry*	$R^2 \pm$ std	RMSE $\pm$ std
RF1	4	0.86± 0.007	56.90 ± 1.41
RF2	2	0.60± 0.006	34.35 ± 0.33
RF3	3	0.85 ± 0.009	62.04 ± 1.76
RF4	2	0.54 ± 0.005	31.98± 0.32

proposed by Imhoff et al. (2004), and the RF4 model was compared with the model proposed by Schjønning and Lamandé (2018).

In general, best predictive performance was observed for those models that used soil water content as predictor, i.e. RF1 and RF3, explaining 42 % and 48 % of the variability in precompression stress, respectively (Fig. 7A and C). Models RF2 and RF4 used soil matric potential as a predictor and explained 27 % and 43 % of the variability in precompression stress (Fig. 7 B and D). Although the RF1 and RF3 model exhibited the highest R^2 value for the independent data, they recorded the highest RMSE (87.21 and 90.96 kPa). The smaller MAPE values of the RF1 and RF3 models (26.7 % and 18.3 %) compared to those of the RF2 and RF4 models (29.0 % and 35.1 %) also signal their stronger predictive capability (Fig. 7).

3.3. Variable importance for the precompression stress predictions

Variable importance ranking is presented in Fig. 8. In the RF1 and RF3 models that are dominated by soils from (sub-) tropical regions, the soil water content and silt content were the most important variables, as well as the calculation methodology in case of RF1. In RF3, methodology was not a variable as the data was restricted to one method, the Dias Junior and Pierce (1995) method. All other variables had much less significance in these two models. In RF2, climatic region, soil matric potential and soil organic carbon content were the three most dominant variables, but interestingly, when the data were restricted to the temperate climatic region (RF4), the significance of soil bulk density and soil depth became amplified and the relative importance of soil matric potential reduced substantially.

3.4. Comparison of prediction performance of RF models and existing pedotransfer functions

When applying the Imhoff (2004) model to the test data used to validate the RF3 model, a considerable number of negative estimates for precompression stress were obtained (Fig. 9A). This indicates a limitation in this model's applicability. When examined closer, the soil moisture range of the data used to develop the Imhoff et al. (2004) model was limited to the range of 0.08–0.28 kg kg^{-1}, whereas our test data set contained samples with initial water contents up to 0.6 kg kg^{-1}. In some cases such extrapolations do not yield obvious discrepancies, but in this case the formulation of the Imhoff (2004) model hints why it is not applicable beyond its original data range. The model uses a coefficient with a large negative number to multiply the soil moisture value with (*c.f.* −773.057 in Table 3 of Imhoff et al., 2004), which can explain the very low (negative) predictions for samples of high water content.

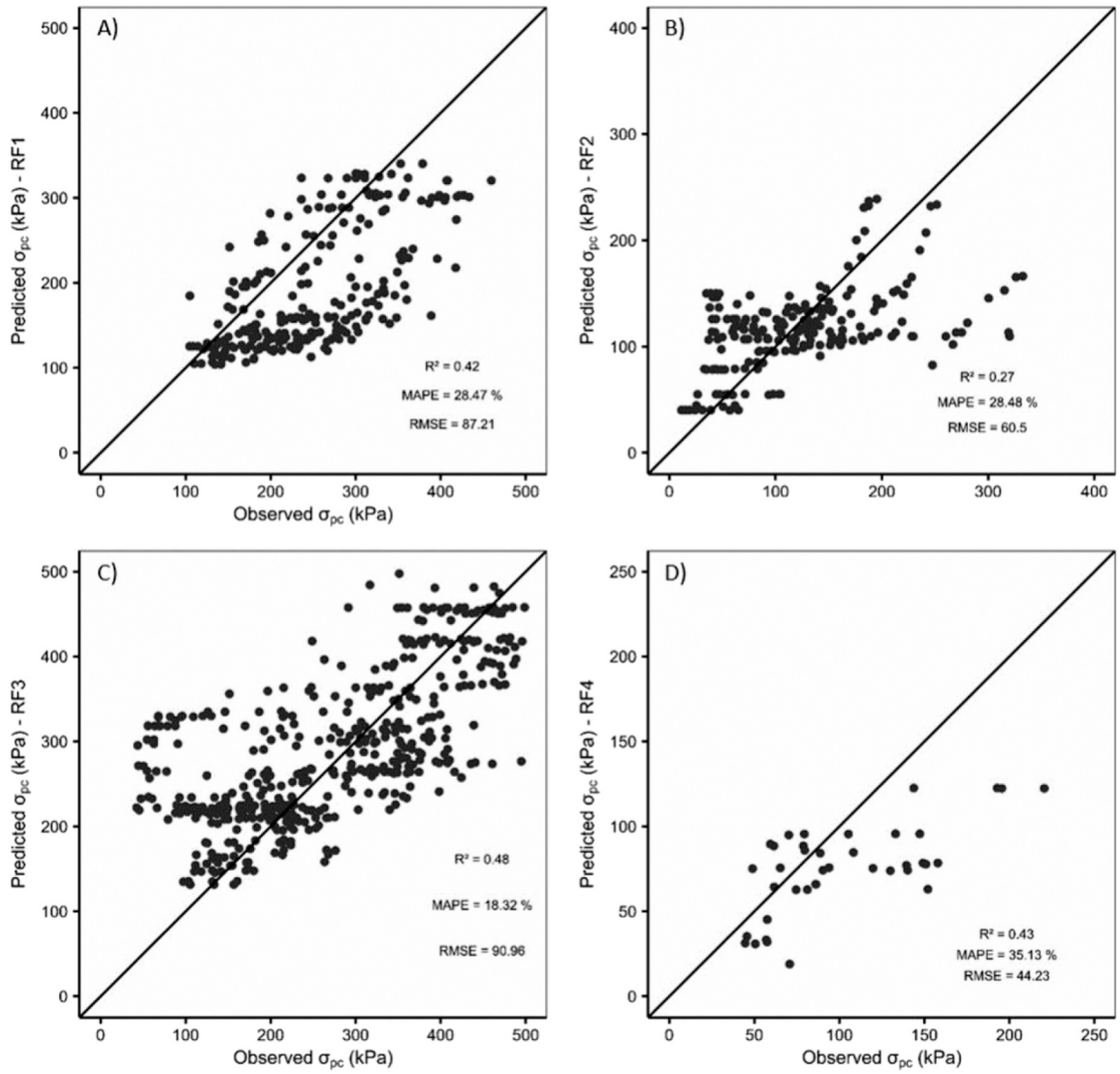


Fig. 7. Predicted precompression stress (σ_{pc}) by Random Forest models RF1 (A), RF2 (B), RF3 (C), and RF4 (D) versus measured σ_{pc}.

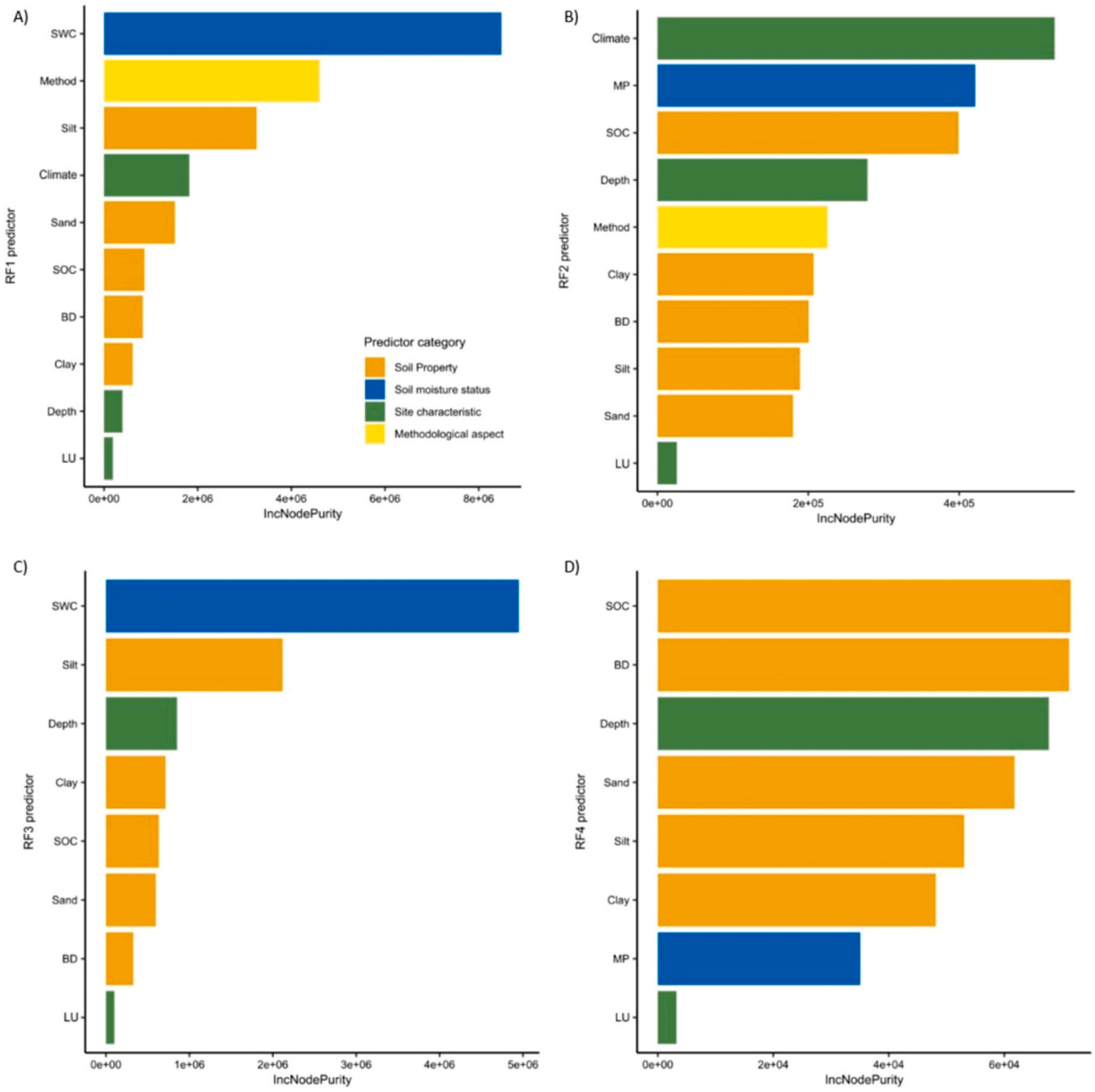


Fig. 8. Relative importance of the predictors for modeling precompression stress by random forest models, RF1 (A), RF2 (B), RF3 (C), and RF4 (D). SWC: soil water content, MP: soil matric potential, SOC: soil organic carbon, BD: soil bulk density, LU: land use.

The RF4 model (Fig. 7D) outperformed the pedotransfer function developed by Schjønning and Lamandé (2018) (cf. Figs. 7D and 9B). While the comparison involved selecting models developed using data from similar climatic regions and with similar input variables, it is important to note that differences in methodological approaches between our test data and the data used to develop these models may impact their predictive power. For instance, our test data for RF3 and RF4 models used the Dias Junior and Pierce (1995) and Casagrande (1936) methods for estimating precompression stress, respectively. In contrast, the Imhoff et al. (2004) model used for comparison with RF3 utilized the Casagrande (1936) method, while the Schjønning and Lamandé (2018) model used for comparison with RF4 utilized the Lamandé et al. (2017) method for estimating precompression stress.

4. Discussion

4.1. Data collection of soil compressive properties

4.1.1. Data gaps for most part of the world

The extensive data compilation predominantly features information from only a few countries, namely Brazil, Germany, Switzerland, Sweden, and Denmark. This highlights a considerable underrepresentation or even absence of data from many parts of the world. The high number of studies from Brazil (64 papers included) can be partly attributed to the inclusion of publications in Portuguese (24 studies). We considered all peer-reviewed papers written with the Latin alphabet, but we only found papers in English and Portuguese that fulfilled all selection criteria. The large number of data entries for precompression stress from Brazilian studies (Fig. 2) is due to several studies presenting data of precompression stress from single soil samples for varying levels of soil

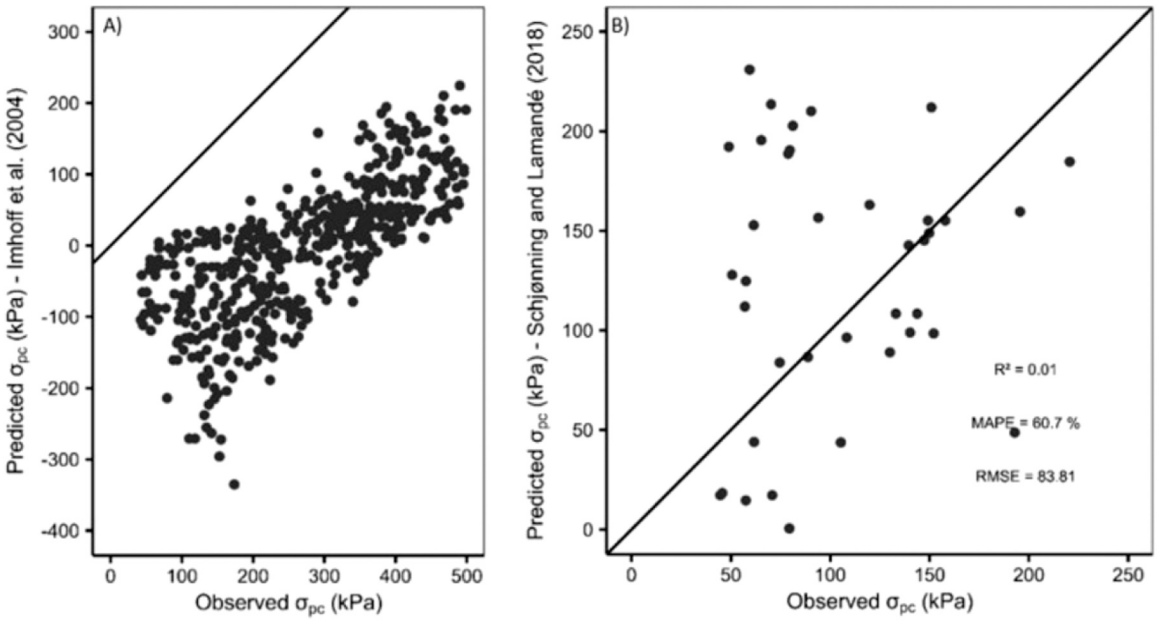


Fig. 9. Predicted precompression stress (σ_{pc}) by the models proposed by A) Imhoff et al. (2004), and B) Schjønning and Lamandé (2018) versus measured σ_{pc}.

water content, obtained by air drying soil samples over a range of time (see e.g. Araujo-Junior et al., 2011; Guimarães Júnnyor et al., 2019 - Table S1). Other studies have primarily reported mean values of precompression stress for specific soil matric potentials or for field moisture conditions (see e.g. Keller et al., 2004; Pesch et al., 2020;Supplementary Table S1), which results in a smaller number of data entries (i.e., precompression stress and corresponding matric potential or corresponding water content).

4.1.2. *Lack of data entries for compression index and recompression index*

In our literature search, we found much more data on precompression stress than on compression index or recompression index. While compression and recompression indexes are relevant for understanding the soil resistance and resilience to compression forces (Imhoff et al., 2004; Kuan et al., 2007), for characterizing the soil stress-strain behavior and hence for modelling soil deformation, the precompression stress represents a limit of mechanical stress above which soil deformation becomes permanent (Lebert and Horn, 1991). The dominance of precompression stress data in our dataset can be attributed to its direct applicability in risk assessment for preventing soil compaction. Consequently, scientists have prioritized quantifying this soil property over compression and recompression indexes, despite these indexes being derived from the same stress-strain curve.

The strong prevalence of precompression stress data from topsoils (0–30 cm soil depth) of conventionally tilled arable fields could reflect a negligence of the soil compaction threat in other tillage systems, e.g. reduced tillage and no-till, in other ecosystems such as forests and grassland, and of deeper soil layers (subsoil compaction) (Fig. 2). The latter observation is particularly interesting, and at the same time perplexing and worrying. While topsoil compaction is not negligible, subsequent tillage practices can, to some extent, mitigate compaction effects. In contrast, there is documented evidence of long-term (decades to centuries) persistence of subsoil compaction and associated negative consequences on soil functioning (Berisso et al., 2012; Keller et al., 2019). Consequently, information on precompression stress seems less relevant for conventionally tilled topsoils than for subsoils. Our compilation and analyses of published information reveals a critical need for obtaining precompression stress data from subsoil layers in order to provide guidance for avoiding subsoil compaction.

4.1.3. *Methodological inconsistences*

Various studies employed different procedures to perform soil compression tests, as evident in our data compilation (Fig. 2). It is well known that methodological aspects significantly impact soil deformation during compression tests, thus influencing the magnitude of compressive properties derived from them. Previous studies have reported the effects of soil-cylinder wall friction on vertical stress during compression tests (Koolen, 1974) and its interaction with sample dimensions (Lima and Keller, 2019). Lebert et al. (1989) demonstrated that precompression stress increases with loading time, and that this influence is higher in more fine-textured soils.

Besides the influence of procedures to perform compression tests, consistent findings, including those by Keller et al. (2004), Cavalieri et al. (2008) and De Pue et al. (2020), highlighted that different calculation methods yield varying values for precompression stress. Furthermore, discrepancies in precompression stress values have been observed when different variables are used to represent soil packing state in compression curves. The use of void ratio and strain yields the same precompression stress value, given their linear relationship. However, the use of bulk density to represent soil structural changes during compression leads to higher precompression stress values (Mosaddeghi et al., 2003; Rücknagel et al., 2010).

The lack of standardization in the methodological procedures for obtaining precompression stress undermines the usability of precompression stress and difficulties in the development of models for prediction of precompression stress based on combined datasets. Since a complete standardization may be considered difficult to achieve for practical reasons, given the variability in equipment and resources among different soil laboratories, conversion functions that translate results from one method to another would be of great value.

The predominance of data for moist to wet initial soil conditions (*p*F 1.8–2.5) is linked to the fact that soil is most prone to compaction under moist conditions. For this reason, it is most important to have information on precompression stress in such moist conditions. It is important to recognize that soil mechanical properties are a function of soil moisture, much like soil hydraulic properties. To accurately estimate this relationship, we argue that data at the "dry end" are equally essential. We may illustrate this in analogy with soil hydraulic functions: if we only measured water retention at *p*F values lower than 2.5, we would not be able to predict or parameterize the complete soil water

retention function. Similarly, the lack of data at the dry end for precompression stress results in incomplete information and hinders our understanding of soil mechanical behavior across a wider soil moisture range.

To address this limitation, we propose exploring the development and parameterization of precompression stress functions, e.g. $\sigma_{pc} = f(\psi)$, similarly to soil water retention functions (i.e. soil water content as a function of matric potential). This approach could provide a more comprehensive understanding compared to attempting to predict precompression stress for single matric potential.

Most of the entries included in the database developed and presented here have expressed precompression stress as a function of soil water content to quantify the effect of soil moisture on precompression stress. Although some studies have proposed that understanding the mechanical behavior of the soil may be better achieved by considering both matric potential and degree of saturation, based on the effective stress theory, rather than soil water content that is confounded with textural differences (Berli et al., 2015; Schjønning et al., 2023), we acknowledge that equilibrating soil samples at a broader range of matric potentials is more labor and cost-intensive. The procedure of obtaining a wider range of soil moisture by drying soil samples over various time intervals, as commonly performed in many Brazilian studies yields precompression stress data for a broad range of soil moisture but only provides data for soil water content. Obtaining similar datasets for both precompression stress and matric potential might be impractical due to the laborious and time-consuming nature of equilibrating numerous samples to different matric potentials. This is because it is not possible to reuse a sample once it has been used in a compression test. We suggest that the installation of microtensiometers in the soil samples would facilitate the collection of more data points for precompression stress across a wider range of matric potential, following a similar approach to determining soil hydraulic properties by the evaporation method (e.g. Hohenbrink et al., 2023).

Moreover, considering the interplay between mechanical stresses and hydraulic properties during loading, the installation of microtensiometers during oedometer tests is recommended to monitor changes in pore water pressure throughout the loading process. This enables adjustments to loading times to prevent excessive pore water pressure in less conductive soils, and helps to better understand soil deformation during compression, aiding in the interpretation of precompression stress values. Only a few studies, such as Horn et al. (2023) and Faloye et al. (2021), have considered the effect of pore water pressure on soil compressive behavior. The importance of considering the time-dependency of load application for more realistic predictions of compressive properties has been demonstrated by Peth et al. (2010).

4.2. Random forest models for precompression stress predictions

The better performance of the RF models that incorporated soil water content as a predictor, i.e. RF1 and RF3, is likely attributed to the wider variation in soil water content in the training data, ranging from wet to very dry soil conditions (Supplementary Figs. S1 and S5), since the soil water content emerged as the most important predictor for the models. We assume that the relatively high RMSE is partially attributable to the inclusion of data with drier initial conditions, resulting in greater precompression stress values associated with greater estimation errors. The poorer correspondence (R^2) shown by the RF2 and RF4 models could be attributed to the limited variation in available matric potential values within the training data, i.e. concentration of data measured at a single matric potential value of pF=1.8 (Supplementary Figs. S3 and S7).

The importance of the calculation method in predicting precompression stress shown in RF1 and RF2 models aligns with existing studies (Cavalieri et al., 2008; De Pue et al., 2020), consistently emphasizing the impact of calculation methods on precompression stress results. This highlights the challenge of developing robust models for predicting precompression stress based on an international database, as methodological variations introduce confounding factors. This also limits us in exploring and delineating the effect of e.g. environmental covariates, such as climate and topography.

Refining models by limiting the underlying data to those obtained by a specific precompression stress calculation method, i.e., RF3 and RF4, showed a small enhancement in predictive power compared to models RF1 and RF2 (Fig. 7). This is an indication to what extent a methodologically homogeneous data set may yield improved results. The trade-off is a limited data set that presents a limited statistical power. Unfortunately, the prevalent precompression stress calculation method and soil moisture status indicator (water content vs. matric potential) differed among the delineated climatic regions, which did not allow us to fully explore all influential factors independently of each other.

In general, the developed RF models emphasized the importance of soil moisture (expressed by matric potential or soil water content), as a key-predictor for estimating precompression stress. This observation aligns with findings from previous studies (Imhoff et al., 2004; Schjønning and Lamandé, 2018; Schjønning et al., 2023). The exception observed in the case of the RF4 model, where soil matric potential exhibited lower importance in predictions, could be attributed to the data used to train the model, which was heavily dominated by a single value of matric potential, pF=1.8 (Supplementary Fig. S7).

The relevance of the climate region in the models, especially in RF2, could be related to the predominance of specific soil types in each of these climate regions: Oxisols in the sub/tropical regions, and Inceptisols and Alfisols in the temperate region. Oxisols, being very old soils, differ from temperate soils by containing kaolinite (a 1:1 clay mineral) as well as iron (Fe) and aluminum (Al) oxides, which affects their resistance to deformation under external loads. The effect of the soil structure associated with the clay mineralogy on the soil precompression stress was shown by Ajayi et al. (2009a), (2009b).

Our models identified other important soil properties as predictors, such as soil texture (clay, silt, sand), soil organic carbon content, and bulk density (Fig. 8). Although our study does not aim to define the mechanisms by which these variables affect precompression stress, existing studies provide insights into their impacts. For instance, soil texture can influence soil mechanical strength, with clay-rich soils typically exhibiting higher precompression stress due to their cohesive nature (Gregory et al., 2006; Saffih-Hdadi et al., 2009). However, in Brazilian Oxisols studied by Severiano et al. (2013), soil precompression stress decreased with increasing clay content. This was attributed to their granular structure due to microaggregation, which promotes greater macroporosity. The increased macroporosity, directly related to the clay content, results in fewer contact points compared to other soil structures, leading to lower resistance to soil deformation. Several studies (e.g. Mosaddeghi et al., 2003) have found that higher bulk density lead to greater precompression stress. Soil organic carbon positively affects aggregate stability, thereby enhancing soil strength. However, soil organic carbon also improves soil water retention, which may increase soil susceptibility to compaction by decreasing precompression stress (Pereira et al., 2007).

4.2.1. RF models versus existing models in literature

The models developed in this study presented better performance than the two selected models from the literature. Although our selection of an independent test data set ensured that no soils were present in the training and test data sets simultaneously, the database-dependent preferential behavior of a new model cannot be ruled out. Historic models are often trained on data of different range/domain, and therefore may show biased behavior when tested on independent data of a new source. This highlights the need to pre-assess whether a predictive model is used on data that is included within the data domain (e.g., within the soil texture and moisture range) of the historic data behind the model.

5. Conclusion

Despite the existence of numerous studies on soil compressive properties, their geographical distribution remains limited. Moreover, our data compilation revealed the urgent need to determine pre-compression stress of subsoil layers to prevent subsoil compaction. While our database covers a broader range of precompression stress data compared to other studies, the low standardization in methodological procedures, calculation methods, and data reporting poses challenges in combining data from different laboratories. Hence, while technically our RF models gave more accurate predictions than existing pedotransfer functions, their performance is deemed to be unsatisfactory. The need for methodological standardization or functions to translate results between methodologies is urgent to ensure consistency, facilitate data comparison, and enable the development of robust and generalizable models for more accurate precompression stress predictions. Therefore, achieving consensus on specific boundary conditions of soil compression tests, including loading time, applied stress level, sample dimension, and calculation methods, along with proper description of methodology could lead to a more unified approach in studying soil compressive properties. There is a concentration of available precompression stress data measured at soil moisture conditions at or above field capacity, signaling that relevant research has prioritized conditions relevant to compaction risk. It is important to realize, however, that soil mechanical properties are functions of soil moisture, much like soil hydraulic functions, and understanding of soil mechanical properties and parameterizing soil mechanical functions requires data across a wider soil moisture range, including the "dry end".

Declaration of Competing Interest

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

Acknowledgements

This study was partially funded by Formas (grant no. 2022-00544), the Swedish Farmeŕs Foundation for Agricultural Research (Stiftelsen Lantbruksforskning, SLF) through grant no. O-22-23-740, and through the Conference of European Directors of Roads (CEDR) Transnational Road Research Programme "Soil" (project "ROADSOIL"). We gratefully acknowledge the authors who contributed additional data from their studies. We thank Alena Holzknecht (SLU, Sweden) for her assistance with data collection for the database construction. The first author (Lorena C. Torres) acknowledges Anderson R. da Silva (Goiano Federal Institute, Brazil) for his initial guidance in developing the random forest models.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.still.2024.106225.

References

ABNT - Associação Brasileira de Normas Técnicas, 1990. NBR 12007: Ensaio de adensamento unidimensional. Rio de Janeiro.

Ajayi, A.E., Dias Junior, M., de, S., Curi, N., Gontijo, I., Araujo-Junior, C.F., Vasconcelos Júnior, A.I., 2009b. Relation of strength and mineralogical attributes in Brazilian latosols. Soil Tillage Res 102, 14–18. https://doi.org/10.1016/j.still.2008.05.013.

Ajayi, A.E., Dias Junior, M. de S., Curi, N., Araujo Junior, C.F., Souza, T.T.T., Inda Junior, A.V., 2009a. Strength attributes and compaction susceptibility of Brazilian Latosols. Soil Tillage Res 105, 122–127. https://doi.org/10.1016/j.still.2009.06.004.

Alaoui, A., Diserens, E., 2018. Mapping soil compaction – A review. Curr. Opin. Environ. Sci. Health 5, 60–66. https://doi.org/10.1016/j.coesh.2018.05.003.

Araujo-Junior, C.F., Dias Junior, M.S., Guimarães, P.T.G., Alcântara, E.N., 2011. Capacidade de suporte de carga e umidade crítica de um latossolo induzida por diferentes manejos. Rev. Bras. Cienc. Solo 35, 115–131. https://doi.org/10.1590/S0100-06832011000100011.

Berisso, F.E., Schjønning, P., Keller, T., Lamandé, M., Etana, A., De Jonge, L.W., Iversen, B.V., Arvidsson, J., Forkman, J., 2012. Persistent effects of subsoil compaction on pore size distribution and gas transport in a loamy soil. Soil Tillage Res 122, 42–51. https://doi.org/10.1016/j.still.2012.02.005.

Berli, M., Casini, F., Attinger, W., Schulin, R., Springman, S.M., Kirby, J.M., 2015. Compressibility of undisturbed silt loam soil-measurements and simulations. Vadose Zone J. 14, 1–11. https://doi.org/10.2136/vzj2014.10.0153.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Casagrande, A., 1936. The determination of pre-consolidation load and its practical significance. In: Proc. Int. Conf. Soil Mechanics and Foundation Engineering, 22–26 June, Vol. 3. Harvard University, Cambridge, Mass., USA, pp. 60–64.

Casini, F., 2012. Deformation induced by wetting: a simple model. Can. Geotech. J. 49, 954–960. https://doi.org/10.1139/t2012-054.

Cavalieri, K.M.V., Arvidsson, J., da Silva, A.P., Keller, T., 2008. Determination of precompression stress from uniaxial compression tests. Soil Tillage Res 98, 17–26. https://doi.org/10.1016/j.still.2007.09.020.

Chamen, W.C.T., Moxey, A.P., Towers, W., Balana, B., Hallett, P.D., 2015. Mitigating arable soil compaction: a review and analysis of available cost and benefit data. Soil Tillage Res 146, 10–25. https://doi.org/10.1016/j.still.2014.09.011.

Culley, J.L.B., Larson, W.E., 1987. Susceptibility to compression of a clay loam Haplaquoll. Soil Sci. Soc. Am. J. 51, 562–567. https://doi.org/10.2136/sssaj1987.03615995005100030002x.

De Pue, J., Di Emidio, G., Bezuijen, A., Cornelis, W.M., 2020. Functional evaluation of the various calculation methods for precompression stress. Soil Use Manag 36, 459–469. https://doi.org/10.1111/sum.12558.

Dias Junior, M.S., Pierce, F.J., 1995. A simple procedure for estimating preconsolidation pressure from soil compression curves. Soil Technol. 8, 139–151. https://doi.org/10.1016/0933-3630(95)00015-8.

Ebrahimzadeh, G., Yaghmaeian Mahabadi, N., Bayat, H., MatinFar, H.R., 2023. Estimating pre-compression stress in agricultural soils: integrating spectral indices and soil properties through machine learning. Comput. Electron. Agric. 215, 108393 https://doi.org/10.1016/j.compag.2023.108393.

Faloye, O.T., Ajayi, A.E., Zink, A., Fleige, H., Dörner, J., Horn, R., 2021. Effective stress and pore water dynamics in unsaturated soils: Influence of soil compaction history and soil properties. Soil Tillage Res 211. https://doi.org/10.1016/j.still.2021.104997.

FAO, ITPS, 2015. Status of the world's soil resources (SWSR) – main report. Food and agriculture Organization of the United Nations and Intergovernmental Technical Panel on soils, Rome, Italy.

Gao, M., Li, H.Y., Liu, D., Tang, J., Chen, Xingyuan, Chen, Xi, Blöschl, G., Ruby Leung, L., 2018. Identifying the dominant controls on macropore flow velocity in soils: a meta-analysis. J. Hydrol. 567, 590–604. https://doi.org/10.1016/j.jhydrol.2018.10.044.

Graves, A.R., Morris, J., Deeks, L.K., Rickson, R.J., Kibblewhite, M.G., Harris, J.A., Farewell, T.S., Truckle, I., 2015. The total costs of soil degradation in England and Wales. Ecol. Econ. 119, 399–413. https://doi.org/10.1016/j.ecolecon.2015.07.026.

Gregory, A.S., Whalley, W.R., Watts, C.W., Bird, N.R.A., Hallett, P.D., Whitmore, A.P., 2006. Calculation of the compression index and precompression stress from soil compression test data. Soil Tillage Res 89, 45–57. https://doi.org/10.1016/j.still.2005.06.012.

Guimarães Júnnyor, W.S., Diserens, E., Maria, I.C., Araujo-Junior, C.F., Farhate, C.V.V., Souza, Z.M., 2019. Prediction of soil stresses and compaction due to agricultural machines in sugarcane cultivation systems with and without crop rotation. Sci. Total Environ. 681, 424–434. https://doi.org/10.1016/j.scitotenv.2019.05.009.

Hohenbrink, T.L., Jackisch, C., Durner, W., Germer, K., Iden, S.C., Kreiselmeier, J., Leuther, F., Metzger, J.C., Naseri, M., Peters, A., 2023. Soil water retention and hydraulic conductivity measured in a wide saturation range. Earth Syst. Sci. Data 15, 4417–4432. https://doi.org/10.5194/essd-15-4417-2023.

Horn, R., Fleige, H., Mordhorst, A., Dörner, J., 2023. Structure dependent changes in pore water pressure due to stress application and consequences on the effective stress. Soil Tillage Res 231. https://doi.org/10.1016/j.still.2023.105719.

Hu, W., Drewry, J., Beare, M., Eger, A., Müller, K., 2021. Compaction induced soil structural degradation affects productivity and environmental outcomes: A review and New Zealand case study. Geoderma 395, 115035. https://doi.org/10.1016/j.geoderma.2021.115035.

Imhoff, S., Da Silva, A.P., Fallow, D., 2004. Susceptibility to Compaction, Load Support Capacity, and Soil Compressibility of Hapludox. Soil Sci. Soc. Am. J. 68, 17–24. https://doi.org/10.2136/sssaj2004.1700.

Keller, T., Arvidsson, J., Dawidowski, J.B., Koolen, A.J., 2004. Soil precompression stress II. A comparison of different compaction tests and stress-displacement behaviour of the soil during wheeling. Soil Tillage Res 77, 97–108. https://doi.org/10.1016/j.still.2003.11.003.

Keller, T., Lamandé, M., Schjønning, P., Dexter, A.R., 2011. Analysis of soil compression curves from uniaxial confined compression tests. Geoderma 163, 13–23. https://doi.org/10.1016/j.geoderma.2011.02.006.

Keller, T., Sandin, M., Colombi, T., Horn, R., Or, D., 2019. Historical increase in agricultural machinery weights enhanced soil stress levels and adversely affected soil functioning. Soil Tillage Res 194, 104293. https://doi.org/10.1016/j.still.2019.104293.

Koestel, J., Jorda, H., 2014. What determines the strength of preferential transport in undisturbed soil under steady-state flow? Geoderma 217–218, 144–160. https://doi.org/10.1016/j.geoderma.2013.11.009.

Koolen, A.J., 1974. A method for soil compactibility determination. J. Agric. Eng. Res. 19, 271–278. https://doi.org/10.1016/0021-8634(74)90066-3.

Kuan, H.L., Hallett, P.D., Griffiths, B.S., Gregory, A.S., Watts, C.W., Whitmore, A.P., 2007. The biological and physical stability and resilience of a selection of Scottish soils to stresses. Eur. J. Soil Sci. 58, 811–821. https://doi.org/10.1111/j.1365-2389.2006.00871.x.

Kuhwald, M., Kuhwald, K., Duttmann, R., 2022. Spatio-Temporal High-Resolution Subsoil Compaction Risk Assessment for a 5-Years Crop Rotation at Regional Scale. Front. Environ. Sci. 10, 823030 https://doi.org/10.3389/fenvs.2022.823030.

Lamandé, M., Schjønning, P., Labouriau, R., 2017. A novel method for estimating soil precompression stress from uniaxial confined compression tests. Soil Sci. Soc. Am. J. 81, 1005–1013. https://doi.org/10.2136/sssaj2016.09.0274.

Lamandé, M., Greve, M.H., Schjønning, P., 2018. Risk assessment of soil compaction in Europe – Rubber tracks or wheels on machinery. Catena 167, 353–362. https://doi.org/10.1016/j.catena.2018.05.015.

Lebert, M., Horn, R., 1991. A method to predict the mechanical strength of agricultural soils. Soil Tillage Res 19, 275–286. https://doi.org/10.1016/0167-1987(91)90095-F.

Lebert, M., Burger, N., Horn, R., 1989. Effects of dynamic and static loading on compaction of structured soils. In: Larson, W.E., Blake, G.R., Allmaras, R.R., Voorhees, W.B., Gupta, S. (Eds.), Mechanics and Related Processes in Structured Agricultural Soils. NATO ASI Series E, Applied Science 172. Kluwer Academic Publishers, Dordrecht, pp. 73–80.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R. N. 2, 18–22.

Lima, R.P., Keller, T., 2019. Impact of sample dimensions, soil-cylinder wall friction and elastic properties of soil on stress field and bulk density in uniaxial compression tests. Soil Tillage Res 189, 15–24. https://doi.org/10.1016/j.still.2018.12.021.

Makungwe, M., Chabala, L.M., Chishala, B.H., Lark, R.M., 2021. Performance of linear mixed models and random forests for spatial prediction of soil pH. Geoderma 397, 115079. https://doi.org/10.1016/j.geoderma.2021.115079.

Mosaddeghi, M.R., Hemmat, A., Hajabbasi, M.A., Alexandrou, A., 2003. Pre-compression stress and its relation with the physical and mechanical properties of a structurally unstable soil in central Iran. Soil Tillage Res 70, 53–64. https://doi.org/10.1016/S0167-1987(02)00120-4.

Nawaz, M.F., Bourrié, G., Trolard, F., 2013. Soil compaction impact and modelling. A review. Agron. Sustain. Dev. 33, 291–309. https://doi.org/10.1007/s13593-011-0071-8.

Nemes, A., Wösten, J.H.M., Lilly, A., Oude Voshaar, J.H., 1999. Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma 90, 187–202. https://doi.org/10.1016/S0016-7061(99)00014-2.

O'Sullivan, M.F., Robertson, E.A.G., 1996. Critical state parameters from intact samples of two agricultural soils. Soil Tillage Res. 39, 161–173. https://doi.org/10.1016/S01671987(96)01068-9.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan-a web and mobile app for systematic reviews. Syst. Rev. 5, 1–10. https://doi.org/10.1186/s13643-016-0384-4.

Palladino, M., Romano, N., Pasolli, E., Nasta, P., 2022. Developing pedotransfer functions for predicting soil bulk density in Campania. Geoderma 412, 115726. https://doi.org/10.1016/j.geoderma.2022.115726.

Pereira, J.O., Défossez, P., Richard, G., 2007. Soil susceptibility to compaction by wheeling as a function of some properties of a silty soil as affected by the tillage system. Eur. J. Soil Sci. 58, 34–44. https://doi.org/10.1111/j.1365-2389.2006.00798.x.

Pesch, C., Lamande, M., de Jonge, L.W., Norgaard, T., Greve, M.H., Moldrup, P., 2020. Compression and rebound characteristics of agricultural sandy pasture soils from South Greenland. Geoderma 380, 114608. https://doi.org/10.1016/j.geoderma.2020.114608.

Peth, S., Rostek, J., Zink, A., Mordhorst, A., Horn, R., 2010. Soil testing of dynamic deformation processes of arable soils. Soil Tillage Res 106, 317–328. https://doi.org/10.1016/j.still.2009.10.007.

Rohatgi A., 2015. WebPlotDigitizer (Version 3.9). ⟨https://automeris.io/WebPlotDigitizer.html⟩.

Rücknagel, J., Brandhuber, R., Hofmann, B., Lebert, M., Marschall, K., Paul, R., Stock, O., Christen, O., 2010. Variance of mechanical precompression stress in graphic estimations using the Casagrande method and derived mathematical models. Soil Tillage Res 106, 165–170. https://doi.org/10.1016/j.still.2009.11.001.

Rücknagel, J., Christen, O., Hofmann, B., Ulrich, S., 2012. A simple model to estimate change in precompression stress as a function of water content on the basis of precompression stress at field capacity. Geoderma 177–178, 1–7. https://doi.org/10.1016/j.geoderma.2012.01.035.

Saffih-Hdadi, K., Defossez, P., Richard, G., Cui, Y.J., Tang, A.M., Chaplain, V., 2009. A method for predicting soil susceptibility to the compaction of surface layers as a function of water content and bulk density. Soil Tillage Res 105, 96–103. https://doi.org/10.1016/j.still.2009.05.012.

Sarkodie, O.V.Y., Vašát, R., Pouladi, N., Šrámek, V., Sáňka, M., Fadrhonsová, V., Hellebrandová, K.N., Borůvka, L., 2023. Predicting soil organic carbon stocks in different layers of forest soils in the Czech Republic. Geoderma Reg. 34 https://doi.org/10.1016/j.geodrs.2023.e00658.

Schjønning, P., van den Akker, J.J.H., Keller, T., Greve, M.H., Lamandé, M., Simojoki, A., Stettler, M., Arvidsson, J., Breuning-Madsen, H., 2015. Driver-Pressure-State-Impact-Response (DPSIR) analysis and risk assessment for soil compaction-A European perspective. Adv. Agron. 133, 183–237. https://doi.org/10.1016/bs.agron.2015.06.001.

Schjønning, P., van den Akker, J.J.H., Greve, M.H., Lamande, M., Simojoki, A., Stettler, M., Arvidsson, J., Breuning-Madsen, H., 2016. In: J. Tesfai, M., Øygarden, L. (Eds.), Soil Threats in Europe. Publications Office of the European Union, Luxembourg, pp. 69–78. https://doi.org/10.2788/828742.

Schjønning, P., Lamandé, M., 2018. Models for prediction of soil precompression stress from readily available soil properties. Geoderma 320, 115–125. https://doi.org/10.1016/j.geoderma.2018.01.028.

Schjønning, P., Lamandé, M., De Pue, J., Cornelis, W.M., Labouriau, R., Keller, T., 2023. The challenge in estimating soil compressive strength for use in risk assessment of soil compaction in field traffic. Adv. Agron. 178, 61–105. https://doi.org/10.1016/bs.agron.2022.11.003.

Schneider, F., Don, A., 2019. Root-restricting layers in German agricultural soils. Part I: extent and cause. Plant Soil 442, 433–451. https://doi.org/10.1007/s11104-019-04185-9.

Severiano, E.D., Oliveira, G.C., Dias, M.D., Curi, N., Costa, K.A.D., Carducci, C.E., 2013. Preconsolidation pressure, soil water retention characteristics, and texture of Latosols in the Brazilian Cerrado. Soil Res 51, 193–202. https://doi.org/10.1071/SR12366.

Spoor, G., 2006. Alleviation of soil compaction: Requirements, equipment and techniques. Soil Use Manag 22, 113–122. https://doi.org/10.1111/j.1475-2743.2006.00015.x.

Stolte, J., Tesfai, M., Øygarden, L., Kværnø, S., Keizer, J., Verheijen, F., Panagos, P., Ballabio, C., Hessel, R., 2016. Soil threats in Europe. Publications Office of the European Union, Luxemburg. https://doi.org/10.2788/828742.

Stone, J.A., Larson, W.E., 1980. Rebound of five one-dimensionally compressed unsaturated granular soils. Soil Sci. Soc. Am. J. 44, 819–822. https://doi.org/10.2136/sssaj1980.03615995004400040032x.

Torres, L.C., Holzknecht, A., Nemes, A., Keller, T., 2023. SoilCompDB: Global soil compressive properties database. Zenodo v1. https://doi.org/10.5281/zenodo.10060810.

van den Akker, J.J.H., 2004. SOCOMO: A soil compaction model to calculate soil stresses and the subsoil carrying capacity. Soil Tillage Res 79, 113–127. https://doi.org/10.1016/j.still.2004.03.021.

van der Westhuizen, S., Heuvelink, G.B.M., Hofmeyr, D.P., 2023. Multivariate random forest for digital soil mapping. Geoderma 431, 116365. https://doi.org/10.1016/j.geoderma.2023.116365.

van Eck, N.J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84, 523–538. https://doi.org/10.1007/s11192-009-0146-3.

Zhang, L., Zeng, Y., Zhuang, R., Szabó, B., Manfreda, S., Han, Q., Su, Z., 2021. In situ observation-constrained global surface soil moisture using random forest model. Remote Sens 13. https://doi.org/10.3390/rs13234893.