Geoderma 416 (2022) 115783

Contents lists available at ScienceDirect

Geoderma

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

Estimating textural fractions of the USDA using those of the International System: A quantile approach

Eva Corral-Pazos-de-Provens, Ígor Rapp-Arrarás, Juan M. Domingo-Santos *

School of Engineering, University of Huelva, E-21007 Huelva, Spain

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:
Piecewise quantile regression
Textural triangle
USDA silt fraction
Particle size classification systems
Soil texture

ABSTRACT

In soil science, the two most frequently used classification systems for the soil particle size distribution are the schemes by the United States Department of Agriculture (USDA) and the so-called International System (IS), whose difference is the upper particle size limit of the silt fraction, namely, 0.02 mm for the IS and 0.05 mm for the USDA system. The existence of these and other systems creates a disparity that hinders and prevents the use and exchange of soil information worldwide. To solve this problem, it is necessary to devise methodologies for the conversion of textural fractions between the different classification systems. This work focuses on the estimation of the USDA silt fraction from the basic textural fractions (sand, silt and clay) in the IS. Five models are currently available for that purpose: the log-linear interpolation method, the Minasny-McBratney-Bristow regression formula, the Shirazi-Boersma-Johnson interpolation method, the Minasny-McBratney regression formula, and the Padarian-Minasny-McBratney regression formula. The accuracy of some of these methods has already been assessed, but in this work we develop a new methodology, based on a local quantile regression, which improves and enriches this evaluation, providing both the regions of the textural triangle where the predictions of the models are acceptable, and the regions where each model is most appropriate. The data used were taken from the publicly available National Cooperative Soil Survey Soil Characterization Database, from which more than 270,000 soil horizon samples were selected for having valid texture data. The analysis carried out concludes that the Padarian-Minasny-McBratney regression formula is the best model of those evaluated. In addition, the tool developed for the evaluation of the models becomes a new model that provides point estimates of the USDA silt fraction from the basic textural fractions in the IS, with further improvement, compared to the 5 models evaluated, as it also provides a prediction interval for those estimates.

1. Introduction

Soil texture refers to the relative percentages by weight of the mineral particles in a soil sample grouped (after the destruction of their aggregates) by size classes in granulometric fractions. Although there is general agreement in the terminology used to name these fractions (clay, silt, and sand), there is no agreement regarding the limits that define them, which vary according to the objectives pursued (Porta Casanellas et al., 2003, chap. 6).

Different scientific disciplines and countries have developed different classification schemes for the soil particle size distribution, most of which share upper limits of particle diameter values of 0.002 mm and 2 mm for the clay fraction and the sand fraction, respectively. It is in the cutoff point between the silt and sand fractions where there is a greater diversity of values. An excellent review of these classification schemes can be found in Blott and Pye (2012).

The particle-size classification schemes used in soil science differ from country to country (Nemes and Rawls, 2006). A common scheme in

Abbreviations: AIC, Akaike information criterion; IS, International System. The particle-size classification adopted by the International Society of Soil Science; IS-S, Simplified version of IS particle-size classification with a single sand class; LLI, The log-linear Interpolation method to estimate the USDA silt fraction from the basic textural information in the IS-S; MM, The Minasny-McBratney regression formula to estimate the USDA silt fraction from the basic textural information in the IS-S; MMB, The Minasny-McBratney-Bristow regression formula to estimate the USDA silt fraction from the basic textural information in the IS-S; PMM, The Padarian-Minasny-McBratney regression formula to estimate the USDA silt fraction from the basic textural information in the IS-S; PTFs, Pedotransfer functions; SBJ, The Shirazi-Boersma-Johnson interpolation method to estimate the USDA silt fraction from the basic textural information in the IS-S.

* Corresponding author.
E-mail addresses: eva.corral@dcaf.uhu.es (E. Corral-Pazos-de-Provens), igor.rapp@dcaf.uhu.es (Í. Rapp-Arrarás), juan.domingo@uhu.es (J.M. Domingo-Santos).

https://doi.org/10.1016/j.geoderma.2022.115783
Received 24 August 2021; Received in revised form 17 February 2022; Accepted 20 February 2022
Available online 11 March 2022

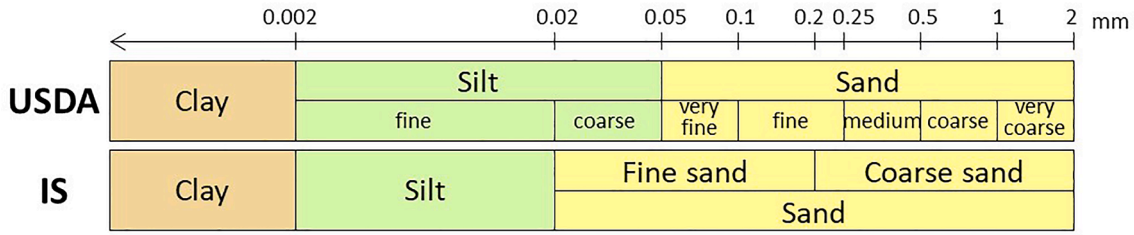


Fig. 1. Particle size limits used in the USDA classification scheme and in the International System (IS) scheme for the different fractions and sub-fractions defined in them.

Europe is the one used in the World Reference Base for soil resources (IUSS Working Group WRB, 2015), which uses the value of 0.063 mm as the cutoff point between the silt and sand fractions. In Russia, in other Eastern European countries, and in China they usually use the Kachinsky's system (Katschinski, 1956), where the limits for the three basic fractions are: under 0.001 mm for clay, 0.001 to 0.05 mm for silt, and 0.05 to 1 mm for sand. But the most frequently used schemes are the so-called International System (IS) and the classification system by the United States Department of Agriculture (USDA) (Porta Casanellas et al., 2003, chap. 6); Fig. 1 shows the size separates used by both systems for fine-earth fraction.

The Swedish chemist and agricultural scientist Atterberg (1905) originally defined the limits of the IS, starting from the upper limit of 2 mm and establishing equal lengths on a decimal logarithmic scale. The International Society of Soil Science (ISSS) officially adopted these limits at its first congress in Washington D.C. in 1927, and established four broad particle-size classes: clay (0–0.002 mm), silt (0.002–0.02 mm), fine sand (0.02–0.2 mm), and coarse sand (0.2–2 mm) (ISSS, 1929). Subsequently, authors of different nationalities simplified this scheme, grouping the two largest fractions, fine sand and coarse sand, in a single class: sand (0.02–2 mm). Hereinafter, we shall refer to the complete scheme as IS and its simplification as IS-S. Currently, this system is used by soil scientists in countries such as Australia, Japan, China, and Spain (Nemes and Rawls, 2006; Richer-de-Forges et al., 2008).

The scheme used by the USDA has its origin in the divisions established by Osborne (1887) using arbitrary particle size limits (Blott and Pye, 2012). Those responsible for the USDA were subsequently modifying the initial limits to make their textural data comparable with the IS data. Thus, in 1938, the boundary between clay and silt was changed (from 0.005 to 0.002 mm) and added into the reports were an extra sieving at 0.2 mm and a pipetting at 0.02 mm (Knight, 1938). Along the same lines, in 1947 the upper limit for sand was shifted from 1 mm to 2 mm (Soil Survey Staff, 1951). The USDA scheme is currently very widely used in many countries outside the US, with the values 0.002, 0.05, and 2 mm as upper limits for the clay, silt, and sand fractions, respectively.

The existence of different classification schemes considerably hinders the exchange and use of soil information, not to mention the confusion caused by the use of the same terms to address different particle size ranges. Efforts have been made to resolve this problem through the development of methodologies allowing the conversion of textural fractions between classification schemes. Thanks to these methodologies, progress can be made in aspects such as the use of pedotransfer functions (PTFs) on an international scale, given that textural fractions are frequently used as input variables in these types of functions (Nemes and Rawls, 2006); the creation of soil databases with information from different countries (Nemes et al., 1999); the comparison between different soil classification systems (Hughes et al., 2017); or any global initiative that provides worldwide soil information.

Most of the published PTFs use, among their input variables, the USDA textural fractions (Nemes and Rawls, 2006). These functions cannot be applied, therefore, in those countries that use the IS, unless practitioners have a model allowing the conversion of their textural fractions to the USDA scheme. On the other hand, although evaluating the texture of a soil sample is a routine analysis, laboratories are often limited to providing only the percentages of sand, silt, and clay. If the information of a greater number of particle-size classes could be made available, then obtaining the equivalences between classification schemes would be simpler and sometimes even immediate.

This study evaluates and compares the performance of the different published models that are used to estimate the USDA textural fractions via the three basic IS-S textural fractions. To do this, an original methodology is developed, based on a local quantile regression, which allows to improve the assessment of the models, and which becomes a new model by itself. Data from more than 270,000 soil horizons are used for the analysis, a number of records one order of magnitude higher than in any previous similar study.

2. Materials

2.1. Input data

This study was carried out by drawing on data from the National Cooperative Soil Survey (NCSS) Soil Characterization Database (NCSS, 2017), which contains soil characterization data from the National Soil Survey Center (NSSC) Kellogg Soil Survey Laboratory (KSSL) and cooperating laboratories, according their standardized methods. This database is freely available for downloading, it includes records from 96 countries around the world, although the majority of the profiles, over 97%, come from the US. More information about the characteristics of this data can be accessed in the provided reference, in various data dictionaries.

This database has been selected, in the first place, because it includes exactly the data needed, namely, the basic USDA textural fractions in addition to the two sub-fractions of silt: the fine silt fraction (0.002–0.02 mm), coinciding with the IS silt fraction, and the coarse silt fraction (0.02–0.05 mm), which is the increment of the silt fraction when converting from the IS to the USDA scheme (see Fig. 1). Also, we are confident that all of these values in the database have been precisely measured, rather than simply estimated. Other reasons for selecting this database have been the reputation of the institution and the large quantity of records available.

After a preliminary review of the database and selection of the records with information regarding the fine silt and coarse silt fractions, we decided to dispense with those records that involved non-standard sample preparation methods (such as soils derived from volcanic materials, for example) and also with those that presented any type of numerical inconsistency (such as the three textural fractions adding up to 99.9% or 100.1% instead of 100.0%, for example). Once the data were filtered, we arranged for calculations regarding a total of 272,360 records.

2.2. Models evaluated

In the body of literature, we found five different ways to estimate the USDA silt fraction (or USDA sand fraction) from the basic textural information in the IS-S: the log-linear interpolation method (Minasny and

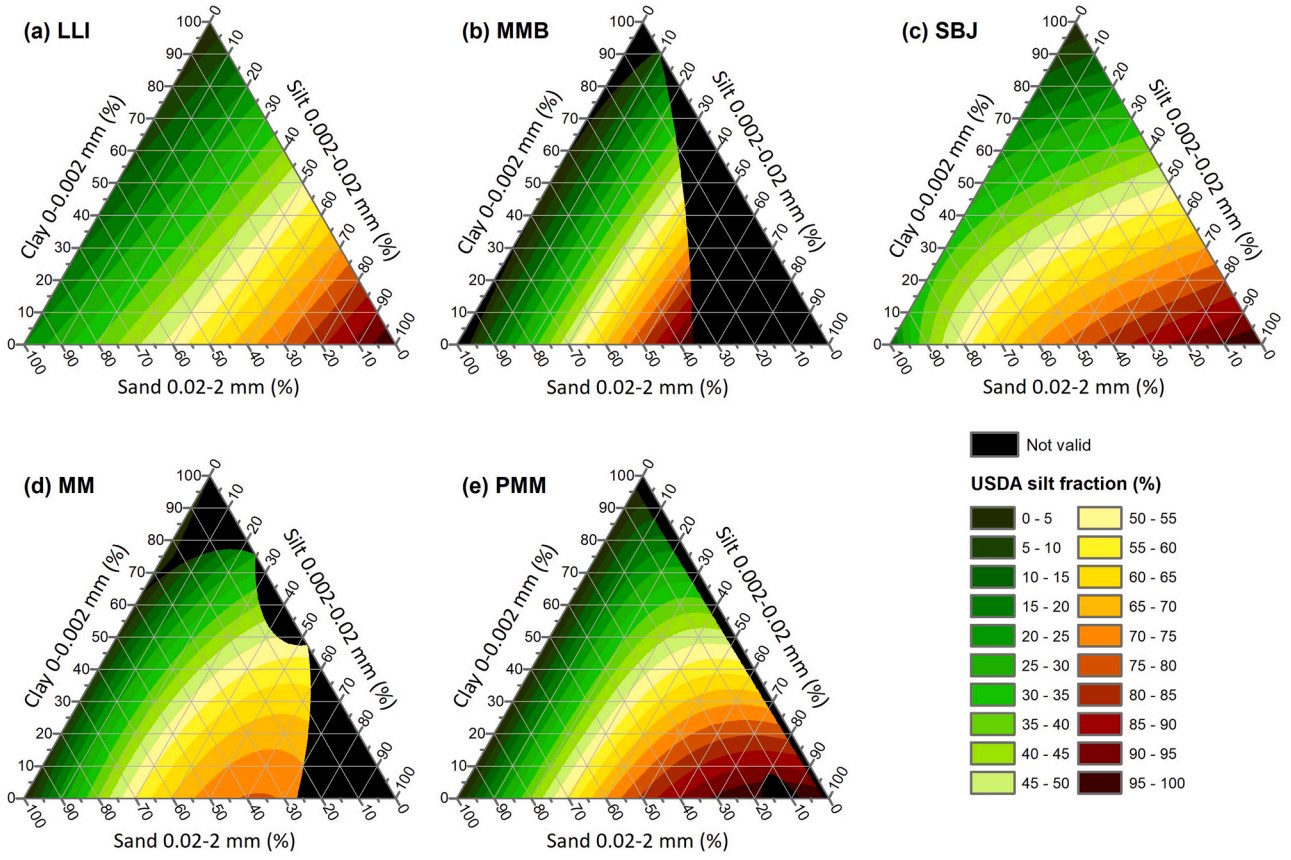


Fig. 2. Models for estimating the USDA silt fraction: (a) the log-linear interpolation method, (b) the Minasny-McBratney-Bristow regression formula, (c) the Shirazi-Boersma-Johnson interpolation method, (d) the Minasny-McBratney regression formula, and (e) the Padarian-Minasny-McBratney regression formula.

McBratney, 2001), the Minasny-McBratney-Bristow regression formula (Minasny et al., 1999), the Shirazi-Boersma-Johnson interpolation method (Shirazi et al., 2001), the Minasny-McBratney regression formula (Minasny and McBratney, 2001), and the Padarian-Minasny-McBratney regression formula (Padarian et al., 2012).

Unless expressly indicated otherwise (e.g., in the graphic representations), it is understood that the sand *d*, silt *t*, and clay *y* fractions are relative proportions expressed in the range from 0 to 1 (both inclusive). This has forced us to transform the numerical values of the fitting parameters of the three aforementioned regression formulas, since their authors expressed them in terms of percentage.

To distinguish between the two soil particle-size classification systems, subscripts have been used: U for the values in the USDA system and I for the values in the IS-S. Thus, t_U and d_U will be, respectively, the silt and sand fractions in the USDA classification system, and t_I and d_I the homonyms in the IS-S classification scheme. The clay fraction does not require a subscript, since it coincides in both classification schemes.

2.2.1. The log-linear interpolation method (LLI)

Log-linear interpolation, that is, linear interpolation in terms of the logarithm of the particle diameter, has frequently been used as a procedure to estimate missing particle size fractions from the cumulative particle-size distribution (e.g., Miao et al., 2008; Nemes et al., 1999; Padarian et al., 2012; Shang, 2013; Shein, 2009; Tietje and Hennings, 1996). In fact, it is the procedure used by the R package "soiltexture" (Moeys, 2018) to adapt soil texture data from one particle size classification system to another. For the case at hand, the method has been used by Minasny and McBratney (2001).

As shown in Appendix A, estimating the USDA silt fraction can be summarized by the following expression:

$$t_U = t_I + 0.19897 \cdot d_I \tag{1}$$

We will refer to this model as LLI.

2.2.2. The Minasny-McBratney-Bristow regression formula (MMB)

In their work on pedotransfer functions (PTFs) for water retention curves, Minasny et al. (1999) developed a polynomial equation to estimate the USDA coarse silt fraction by means of least squares regression. The fitting was made using the granulometric information of 161 Australian soil samples, 146 soil samples reported by Geeves et al. (1995) and 15 soil samples reported by Smettem and Gregory (1996).

Minasny et al. (1999) calculated the USDA silt fraction by adding to the IS silt fraction the value of the USDA coarse silt fraction provided by their equation. As shown in Appendix B, their two-step procedure can be summarized by the following formula:

$$t_U = -0.030607 + 1.3352 \cdot t_I + 0.1127 \cdot d_I + 0.18 \cdot {t_I}^2 + 0.36 \cdot t_I \cdot d_I - 0.11 \cdot {d_I}^2 \tag{2}$$

This model, which will be referred to as MMB, was used by Minasny and McBratney (2000).

2.2.3. The Shirazi-Boersma-Johnson interpolation method (SBJ)

Shirazi et al. (1988) postulated that the cumulative particle-size distribution for fine-earth of a soil sample can be approximated by three segments of lognormal distribution function: one for the sand fraction, another for the silt fraction, and a third for the clay fraction. Based on this theory, Shirazi et al. (2001) described a method to estimate the USDA silt fraction, which can be summarized by the following interpolation function:

$$t_U = t_I + \Phi\left[0.739973 + 0.801030 \cdot \Phi^{-1}(1 - d_I)\right] - 1 + d_I \tag{3}$$

where Φ denotes the standard normal distribution function and Φ^{-1} denotes its inverse (see Appendix C).

We refer to this model as SBJ.

2.2.4. The Minasny-McBratney regression formula (MM)

For estimating the USDA silt fraction, Minasny and McBratney (2001) presented a piecewise polynomial model derived by least squares regression analysis. For the calibration, the authors used 1,210 records from several sources: the Australian data sets reported by Geeves et al. (1995) and Smettem and Gregory (1996) (see section 2.2.2), the unsaturated soil hydraulic database UNSODA (Leij et al., 1996), the Grenoble Catalogue of Soils GRIZZLY (Haverkamp et al., 1997) and the USDA Natural Resources Conservation Service (NRCS) Soil Survey Laboratory data (USDA-NRCS, 1997). The resulting formula can be expressed as

$$t_U = \begin{cases} t_{U,0} & \text{for} \quad t_{U,0} \geq 0 \\ 0.8289 \cdot t_I + 0.0198 \cdot d_I & \text{for} \quad t_{U,0} < 0 \end{cases} \tag{4}$$

where $t_{U,0}$ is given by

$$t_{U,0} = -0.183914 + 2.0971 \cdot t_I + 0.6726 \cdot d_I - 1.42 \cdot {t_I}^2 - 0.49 \cdot {d_I}^2 \tag{5}$$

Minasny and McBratney (2001) tested Eq. (4) with another 400 records from the same source (internal validation) obtaining a mean error (ME) of 0.39% and a root mean square error (RMSE) of 8.01%.

We refer to this model as MM. It has been used by Minasny and McBratney (2007) and Sakaguchi et al. (2014).

2.2.5. The Padarian-Minasny-McBratney regression formula (PMM)

In order to improve the performance of the MM model, Padarian et al. (2012) conducted symbolic regression analysis based on a genetic programming technique. For the calibration, the authors used 52,432 records from the National Soil Characterization database from USDA/NRCS (Soil Survey Staff, 1995) and the resulting formula was

$$t_U = 2.26 \cdot t_I + \frac{5.55 \cdot t_I + 151.3 \cdot {t_I}^2}{0.9966 - 123.6 \cdot t_I - 134.9 \cdot d_I} \tag{6}$$

Padarian et al. (2012) tested their formula with another 52,432 records from the same database (internal validation), obtaining a coefficient R^2 of 82%.

We refer to this model as PMM. It has been used by Hughes et al. (2017).

2.2.6. Representation of the models evaluated on the textural triangle

Fig. 2 displays the predictions of the USDA silt fraction, in percentage, provided by the five models evaluated for each possible combination of sand-silt–clay (IS-S), that is, for each point of the textural triangle constructed according to the IS-S criteria.

We have marked in black color and labeled as "Not valid" those areas of the textural triangle where the models provide non-sense values in physical terms, that is, values of the USDA silt fraction that are lower than values of the IS silt fraction or values higher than the sum of the silt and sand fractions in the IS-S (see Fig. 1); that is, they provide values for the USDA silt fraction that do not meet the following condition:

$$t_I \leq t_U \leq t_I + d_I \tag{7}$$

Both the log-linear interpolation and the Shirazi-Boersma-Johnson interpolation method are monotonic in nature, hence Eq. (1) and Eq. (3) provide values that meet conditions of the Eq. (7) for all points on the textural triangle (see Fig. 2a and c). But the other three models, which are regression formulas in all three cases, do not always meet the condition of Eq. (7). Thus, the MMB model provides non-sense values for 34.9% of the area of the textural triangle (Fig. 2b); for the MM model it is 19.8% (Fig. 2d); and for the PMM model it is 6.3% (Fig. 2e).

In Fig. 2d it can be noted that the region of the textural triangle where Eq. (4) does provide values with sense is made up of two unconnected areas, the largest corresponding to the upper piece of the equation ($t_{U,0} \geq 0$) and the smallest to its lower piece ($t_{U,0} < 0$).

Note also that Fig. 2e is, in essence, equivalent to Fig. 2a of Padarian et al. (2012).

3. Methods

For the development of this work, we have followed the methodology described in Corral-Pazos-de-Provens et al. (2018), with the necessary adaptations for our objective. A summary of this methodology is presented in the next sections.

3.1. Descriptive statistics

3.1.1. Least squares polynomial regression

In order to describe how the USDA silt fraction varies in relation to the three basic textural fractions in the IS–S (sand, silt, and clay), we first applied least squares fits of polynomial functions, namely,

$$t_U = \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{ij} \cdot f_1^i \cdot f_2^j \tag{8}$$

where n is the degree of the polynomial, $a_{i,j}$ stands for the fitting parameters, and f_1 and f_2 represent any two of the three basic textural fractions in the IS–S, taking into account that the inclusion of the third fraction is unnecessary, since the sand, silt and clay fractions are linearly dependent. The number p of parameters $a_{i,j}$ is given by the expression $p = ½ \cdot (n + 1) \cdot (n + 2)$ (Venables and Ripley, 2002, chap. 15). All calculations were performed using the R Stats package (R Core Team, 2020).

3.1.2. Subdivision of the textural triangle

Given that the amount of available data for this study was so large, it was possible to divide the textural triangle into smaller triangles (tiles) in order to analyze their records in turn, that is, to carry out a piecewise statistical analysis using the 272,360 available records. For this, the textural triangle was first divided into four equal equilateral triangles; then, each of them was subdivided into four new equal equilateral triangles, thus initiating an iterative process in which, as the level of subdivision (r) increases, the number of triangles generated (4^r) increases considerably. Each of these generated triangles contains a different number of records, logically, even some empty triangles may appear.

The coefficient of determination for the USDA silt fraction has been calculated, for each level of subdivision, by the following expression,

$$R^2 = 1 - \frac{\sum_{i=1}^{q} \sum_{j=1}^{N_i} (t_U(i,j) - m_i)^2}{\sum_{i=1}^{q} \sum_{j=1}^{N_i} (t_U(i,j) - m_T)^2} \tag{9}$$

where q is the number of non-empty triangles, N_i is the number of data values in the i-th non-empty triangle, $t_U(i,j)$ is the j-th USDA silt fraction in the i-th non-empty triangle, m_i is the average USDA silt fraction in the i-th non-empty triangle, and m_T is the average USDA silt fraction in the original textural triangle. That is, we carried out a piecewise least squares regression (Dagnelie, 2011, chap. 15; James et al., 2013, chap. 7) where the pieces are the q non-empty triangles, in a domain of two explanatory variables.

As the level of subdivision of the textural triangle (r) increases, the coefficient of determination, logically, also increases, but at the same time the number q of parameters increases at a nearly exponential pace. Then, in order to set the size of these triangles that will be the object of subsequent analysis, we have applied the minimization of the Akaike information criterion (AIC) (see, e.g., Burnham and Anderson, 2002, chap. 2). In this instance, this criterion is obtained through the following expression:

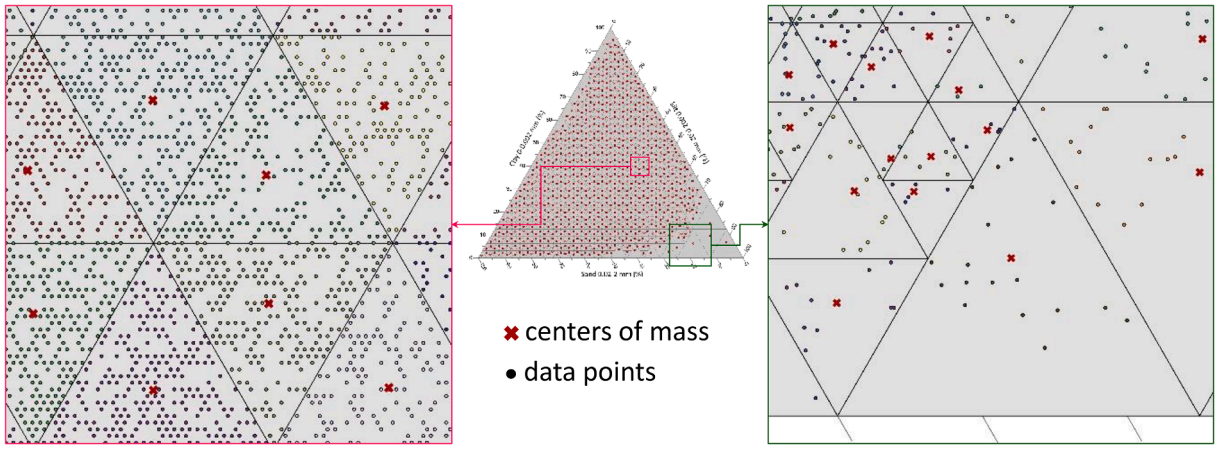


Fig. 3. Detail of the subdivisions (tiles) of the textural triangle and the location of the centers of mass (interpolation nodes) of the set of points of each tile, in order to assign them their statistical measures of the USDA silt fraction and estimate their spatial distribution, using IDW interpolation.

$$\mathrm{AIC} = N \cdot ln\left[\frac{\sum_{i=1}^{q}\sum_{j=1}^{N_i}\left(t_U(i,j) - m_i\right)^2}{N}\right] + 2 \cdot q \tag{10}$$

where N is the total number of data values.

Finally, some triangles resulting from the subdivision were merged to meet the condition that the number of data points contained in each triangle would be not smaller than 3. This last step was necessary because one of the statistical tests used later, the Shapiro-Wilk test, requires a sample size equal to or greater than 3. To do this, the iterative process described above was reverted at the local scale, merging those triangles that contained less than 3 records with their corresponding triangles at the higher subdivision level; in the event that the triangle obtained did not contain 3 records, the fusion was continued towards the next level, and so on until the required minimum was reached.

At the end of the process, q' tiles were generated, all of them equilateral triangles, but of different sizes, and all with 3 or more records. Their analysis will be described within the next sections.

The calculations have been performed using PTC Mathcad Prime 4.0. We have represented the results graphically using ArcGIS Desktop by ESRI: Release 10.8.

3.1.3. Testing for normality

The values of the USDA silt fraction in each of the q' tiles were subjected to the Shapiro-Wilk test of normality, according to the version by Royston (1992). All calculations were performed using the R Stats Package (R Core Team, 2020).

3.1.4. Conditional statistical measures

The conditional value of several statistical measures relative to the USDA silt fraction for each point of the textural triangle was calculated from the information contained in each of the q' tiles generated. Specifically, we have generated the conditional values of the mean m, the standard deviation s, the Fisher's skewness coefficient g, the first quartile Q_1, the median or second quartile Q_2 and the third quartile Q_3.

The steps taken for this have been the following: (i) calculation of the coordinates of the center of mass of the N_i data points belonging to the i–th tile ($i = 1, 2,\ldots, q'$) (see Fig. 3); (ii) calculation of the mean m_i of the N_i values of the USDA silt fraction corresponding to the i–th tile; (iii) location of the value obtained from the mean m_i on its corresponding center of mass; (iv) from these mean values, to extend the information to the whole textural triangle, the inverse distance weighting (IDW) interpolation is used, where the interpolation nodes are the defined centers of mass; (v) repetition of steps (ii) to (iv) for each of the statistical measures considered (see Appendix D).

The expression of the interpolation function used, in terms of the coordinates d_I, t_I and y of an arbitrary point P in the textural triangle, is the following

$$m(d_I, t_I, y) = \frac{\sum_{i=1}^{q'} w_i(d_I, t_I, y) \cdot m_i}{\sum_{i=1}^{q'} w_i(d_I, t_I, y)} \tag{11}$$

where the weighting coefficient w_i for the i–th interpolation node is given by

$$w_i(d_I, t_I, y) = \frac{f(N_i)}{[D_i(d_I, t_I, y)]^u} \tag{12}$$

with $f(N_i)$ being a generic function of the number of records in the i-th tile, D_i is the measured distance on the textural triangle between data point P and the i-th interpolation node, and u is a simple numeric parameter. Among the different functions $f(Ni)$ and the values of the parameter u which are considered, namely, N_i, square root of N_i, log N_i and 1 for $f(N_i)$, and 1, 2, 3 and 4 for u, those with the best performance in terms of least squares will be chosen.

Finally, we generated the interpolation functions for the mean m (d_I, t_I, y), the standard deviation s (d_I, t_I, y), the skewness coefficient g (d_I, t_I, y), the first quartile Q_1 (d_I, t_I, y), the median Q_2 (d_I, t_I, y), and the third quartile Q_3 (d_I, t_I, y), related to the USDA silt fraction for each point in the textural triangle.

It can be seen that the performance of step (ii) for each quartile consists of a piecewise quantile regression where the pieces are the q' tiles.

All calculations were performed using the R Base and R Moments packages (R Core Team, 2020). Interpolated values have been obtained using ArcGIS Desktop by ESRI: Release 10.8.

3.2. Model assessment

A preliminary analysis of the efficiency of the five models based on the R^2 coefficient yields the following results: 49.9% for LLI, 69.9% for MMB, 52.6% for SBJ, 80.8% for MM, and 81.1% for PMM; the analysis was performed by means of PTC Mathcad Prime 4.0. These figures serve to provide a general idea of the performance of the different models; however, they do not allow us to identify the areas of the textural triangle in which each model provides predictions that may or may not be considered acceptable.

Thus, in adapting to our case the criteria established by Corral-Pazos-de-Provens et al. (2018), we judged that a model would be acceptable for estimating the USDA silt fraction at a point P in the textural triangle if it provided, for the said point, a value between those values determined by the interpolating function of the first quartile Q_1 (d_I, t_I, y) and the interpolating function of the third quartile Q_3 (d_I, t_I, y), both included.

Table 1
Coefficients of determination (R^2) for least squares polynomial regression as function of the degree of the polynomial.

n	*p*	R^2 (%)
1	3	76.0
2	6	81.8
3	10	82.1
4	15	82.1

n is the degree of the polynomial; p is the number of parameters of the polynomial, and R^2 is the coefficient of determination.

Table 2
Values of the Akaike information criterion for the piecewise least squares regression as function of subdivision level of the textural triangle.

r	4^r	*q*	R^2 (%)	AIC $\times 10^{-6}$
1	4	4	37.0	−0.98429
2	16	16	68.3	−1,17146
3	64	64	78.1	−1,27249
4	256	255	81.2	−1,31269
5	1024	996	82.1	−1,32444
6	4096	3732	82.4	−1,32422

r is the level of subdivision of the textural triangle; 4^r is the number of triangles generated; q is the number of parameters adjusted (non-empty triangles); R^2 is the coefficient of determination, and AIC is the value of the Akaike information criterion.

As a result of this procedure, for each of the models studied, we were able to establish two regions within the textural triangle, one where the performance of the model in question was acceptable and the other where it was not. These regions of acceptance have been obtained using ArcGIS Desktop by ESRI: Release 10.8.

Logically, the higher the density of records in the textural triangle, the more confidence there will be that the graphs of the interpolation functions obtained represent statistical reality.

4. Results

4.1. Descriptive statistics

4.1.1. Least square polynomial regression

Table 1 shows how the value of the coefficient of determination (R^2) increases with the degree of the polynomial (n), although it begins to stabilize at the second degree polynomial.

It is possible to achieve performances of the same order as the complete polynomials by eliminating *a priori* some of the parameters. Specifically, if we impose the logical condition that, if the IS-S sand fraction is zero, then the USDA silt fraction will be equal to the IS silt fraction, we obtain the following second degree polynomial

$$t_U = t_I + 0.4095 \cdot d_I + 0.9843 \cdot d_I \cdot t_I - 0.4232 \cdot {d_I}^2 \quad (13)$$

The coefficient of determination (R^2) of the Eq. (13), with only 4 parameters, is 81.6%, with a negligible difference in relation to the polynomial with 6 parameters (81.8%).

4.1.2. Subdivision of the textural triangle

Table 2 shows how, as the level of subdivision of the textural triangle (r) and the number of parameters (q) increase, the coefficient of determination (R^2) logically increases. The minimization of the AIC allows us to set the level of subdivision of the textural triangle that we are going to use for the subsequent calculations, in this case r equal to 5, which is equivalent to establishing 1,024 equilateral triangles of the same size. The merger of some of these triangles, in order to achieve tiles with 3 or more records each, resulted in 913 tiles, as can be seen in Fig. 4, all of them equilateral triangles but with different sizes, namely, 892 with sides of 3.125%, 17 with sides of 6.25%, and 4 with sides of 12.5%.

Likewise, Fig. 4 shows the relative density of the records in each tile, that is to say, the ratio between the density of the records in each tile and the density of the records across the whole textural triangle. The density of records for each tile is calculated as the number of records in the tile divided by the tile surface area. To aid description, we have superimposed the limits of the textural classes proposed by Tommerup (1934).

The highest relative density values are all located within the loamy sand texture class, with an absolute maximum of 19.37 in the lower left corner tile, whose centroid has coordinates: 98.3% sand, 0.8% silt, and 0.9% clay. From this maximum value the relative density decreases as the clay and silt contents increase. Silty classes are clearly less represented, especially the silty loam class, with most of its tiles far having below average densities. In fact, it is remarkable that the difference between these three silty classes, which represent almost 30% of the textural triangle and contain 5.8% of the records, compared to the loamy sand class, which nearly doubles the amount of records, 11.2%, but only takes up 2.2% of the textural triangle surface. The largest texture class (30.3% of the textural triangle), the heavy clay class, presents, mostly, densities below the average (contains 11.1% of the records). The rest of the texture classes generally have above average densities, with loam textures being the most represented.

4.1.3. Testing for normality

The application of the Shapiro-Wilk test to the data of the USDA silt fraction for each tile led to the rejection of the null hypothesis, with a significance level of 0.05, in 72.6% of the tiles. After applying the transformations of variables most commonly used in these cases (the logarithm transformation, the square root transformation and, after dividing the IS silt fraction by the USDA silt fraction, the arcsine transformation), the results did not improve substantially and yielded a rejection of 71.1% for the logarithmic transformation, of 72.1% for the square root transformation and 76.8% for the arcsine transformation.

4.1.4. Conditional statistical measures

Figs. 5 and 6 show the conditional values of the calculated statistical measures. These values have been generated by means of the inverse distance weighted (IDW) interpolation, taking for the weighting coefficient the function $f(N_i) = 1$ and the parameter $u = 3$ (see Section 3.1.4), both selected for their best performance in terms of least squares.

In Fig. 5a, it can be seen, as it is logical, how the conditional mean of the USDA silt fraction increases as we move away from the axis corresponding to clay (value 0 of the IS silt fraction), and reaches the highest values in the silty loam class. For values of IS silt below 20%, the IS silt fraction is practically the only explanatory variable, as revealed by the parallelism of the contour lines with the axis representing clay content. As we move away from this axis, the influence of clay and sand fractions in the IS-S increases, and when we get to the three silty classes, the influence of the clay fraction is more evident, as indicated by the bending of the contour lines, practically to horizontal.

The IDW interpolation has provided non-sense values of the conditional mean for 2.3% of the area of the textural triangle, coinciding with areas where the density of the records is lower.

The conditional standard deviation of the USDA silt fraction presents a high variability, as can be seen in Fig. 5b, which indicates a clear case of heteroskedasticity. According to the criteria of Bulmer (1979), in Fig. 5c it can be verified that the conditional distribution of the USDA silt fraction is asymmetric in slightly more than half of the textural triangle area (it is only fairly symmetrical in 43.7%). In these circumstances of heteroskedasticity and asymmetry, the use of the median will therefore be more appropriate than the mean to describe the central location of the distribution (Hao and Naiman, 2007, chap. 3).

The conditional first quartile, median, and third quartile of the USDA silt fraction, as can be seen in Fig. 6, follow a trend similar to the conditional mean (Fig. 5a). In each of these charts the "Not valid" region, showing values that do not meet conditions of Eq. (7), has also been

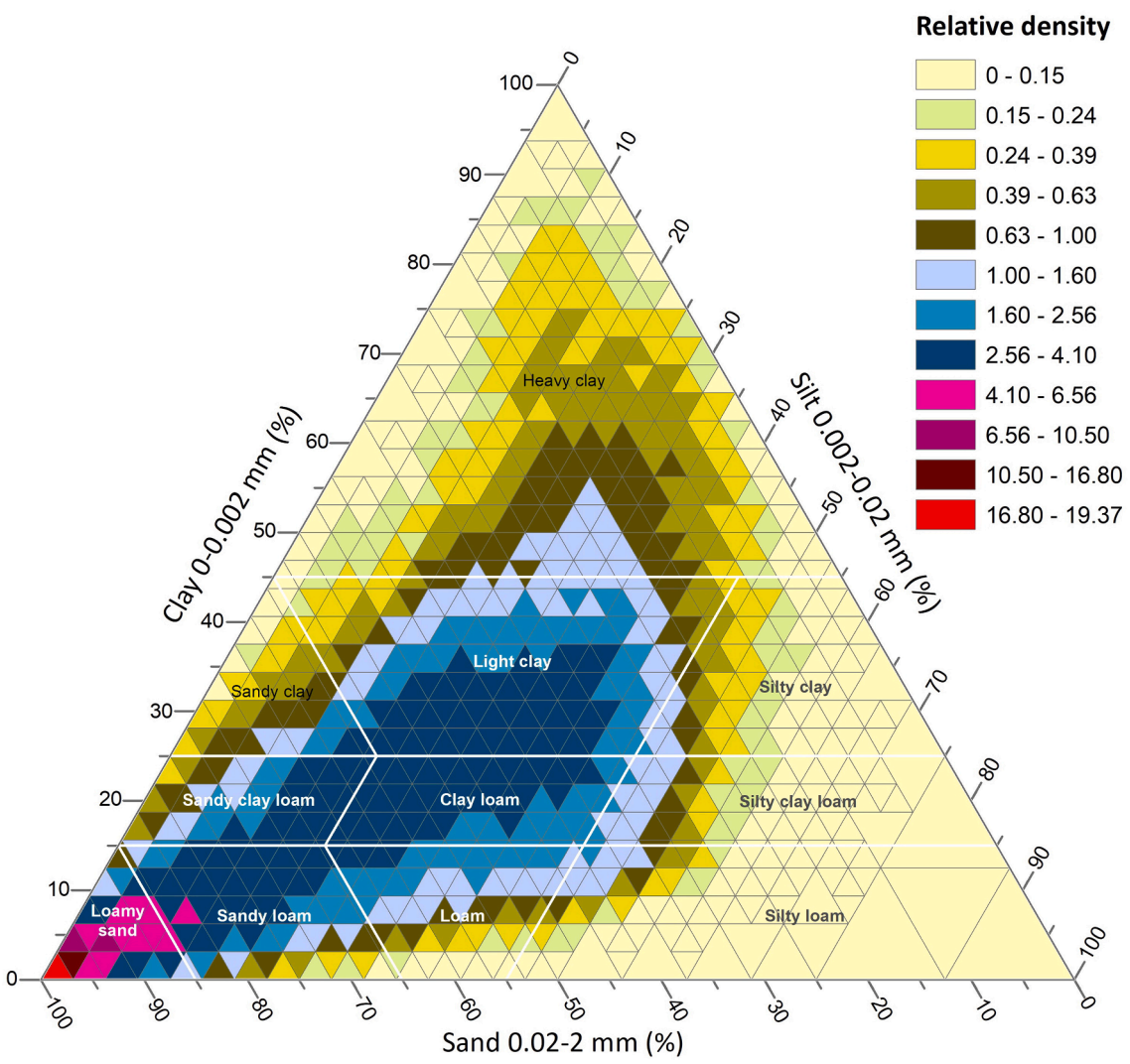


Fig. 4. Relative density of data points, expressed as the ratio between the density of the records in each tile and the density of the records across the whole textural triangle.

indicated, representing 3.6%, 2.3%, and 3.8%, respectively, of the textural triangle.

4.2. Model assessment

The LLI model is acceptable in 15.5% of the area of the textural triangle (Fig. 7a), which contains 18.7% of the records. This region corresponds roughly to values of IS silt fraction between 10% and 20% and with IS-S sand fraction greater than 25%.

The MMB model is acceptable in 26.7% of the textural triangle (Fig. 7b), mainly in the loam, clay loam, and part of the light clay and heavy clay texture classes. This acceptance region includes 39.3% of records.

With respect to the SBJ model, Fig. 7c shows that its region of acceptance covers just over half of the textural triangle (52.6%) and this extends towards the part of the triangle with the most silt, where the number of records is low. The main unacceptable area corresponds approximately to values of less than 20% IS silt fraction and more than 20% IS–S sand fraction, in addition to some scattered areas in the three silty classes. The acceptance region includes 48.9% of the records.

As can be seen in Fig. 7d, the acceptance region of the MM model comes to represent 54.3% of the textural triangle, containing 89.1% of the records. The texture classes in which the model does not work adequately are the three silty classes, part of the heavy clay class, when the IS-S sand fraction is less than 10% or the clay fraction is greater than 70%, and part of the light clay and sandy clay classes, when the IS silt fraction is less than 10%.

The model with the largest region of acceptance is the PPM, which has managed to cover 68.2% of the textural triangle and 95.1% of the records (see Fig. 7e). The main areas of non-acceptance belong to the silty classes when IS–S sand fraction is below 20% or clay fraction is less than 10%, as well as to a small part of the heavy clay class when the clay fraction is greater than 80%.

There are areas in the textural triangle where it has not been possible to perform the evaluation of the models, labeled as “Not evaluable” in Fig. 7. This is because at these points in the triangle the interpolated values of the conditional first quartile and third quartile of the USDA silt fraction do not meet conditions of Eq. (7) (“Not valid” areas in Fig. 6a and c). On the other hand, the areas in which the evaluated model provides non-sense values have been directly included in the “Unacceptable” region.

The acceptance regions of the five models have been superimposed, thus producing unavoidable overlapping regions. At those points in the textural triangle where more than one model is acceptable, we have given preference to the model that provides the value closest to the median. Fig. 7f shows the result of the application of this criterion, where it can be seen that the regions of preference are highly intermixed. Hence, the PMM formula is the model that has the greatest region

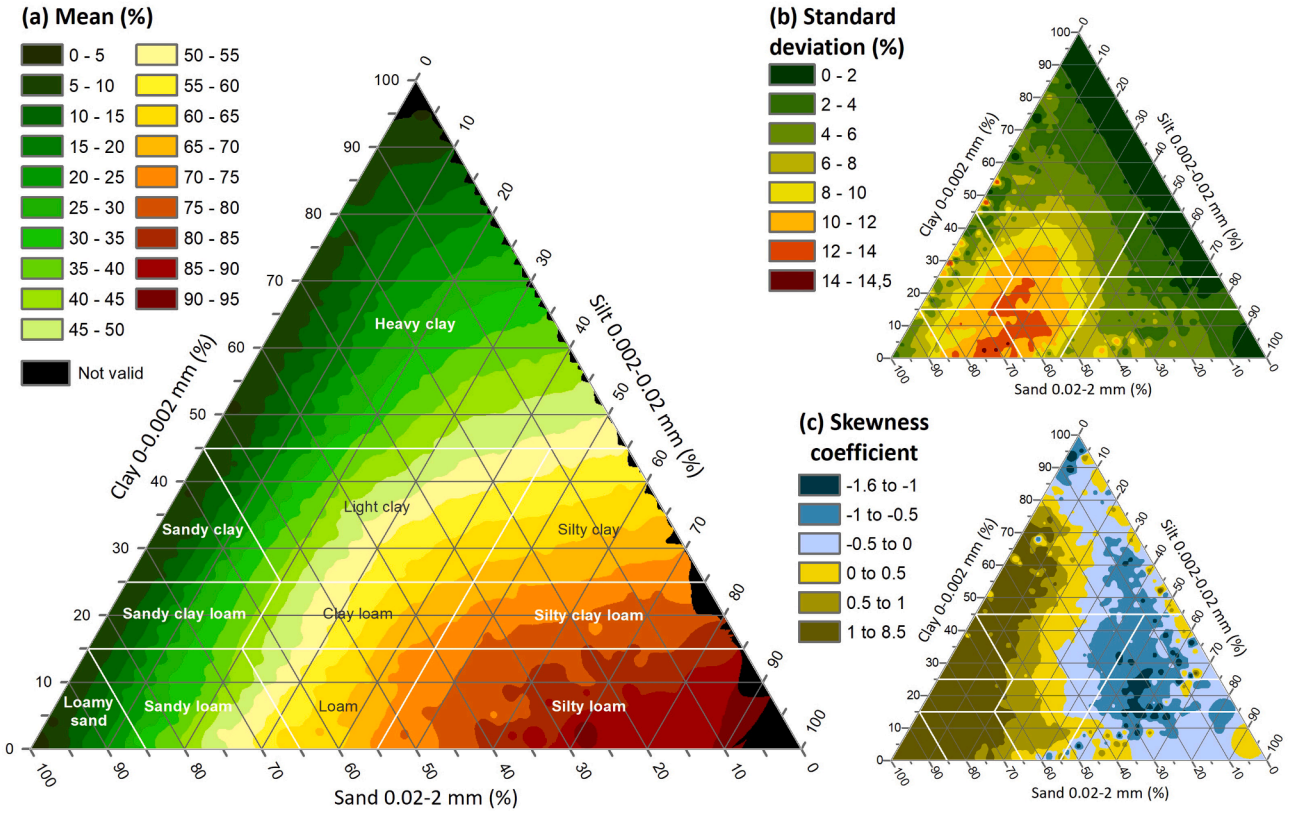


Fig. 5. Interpolated values of the main conditional measures of the USDA silt fraction.

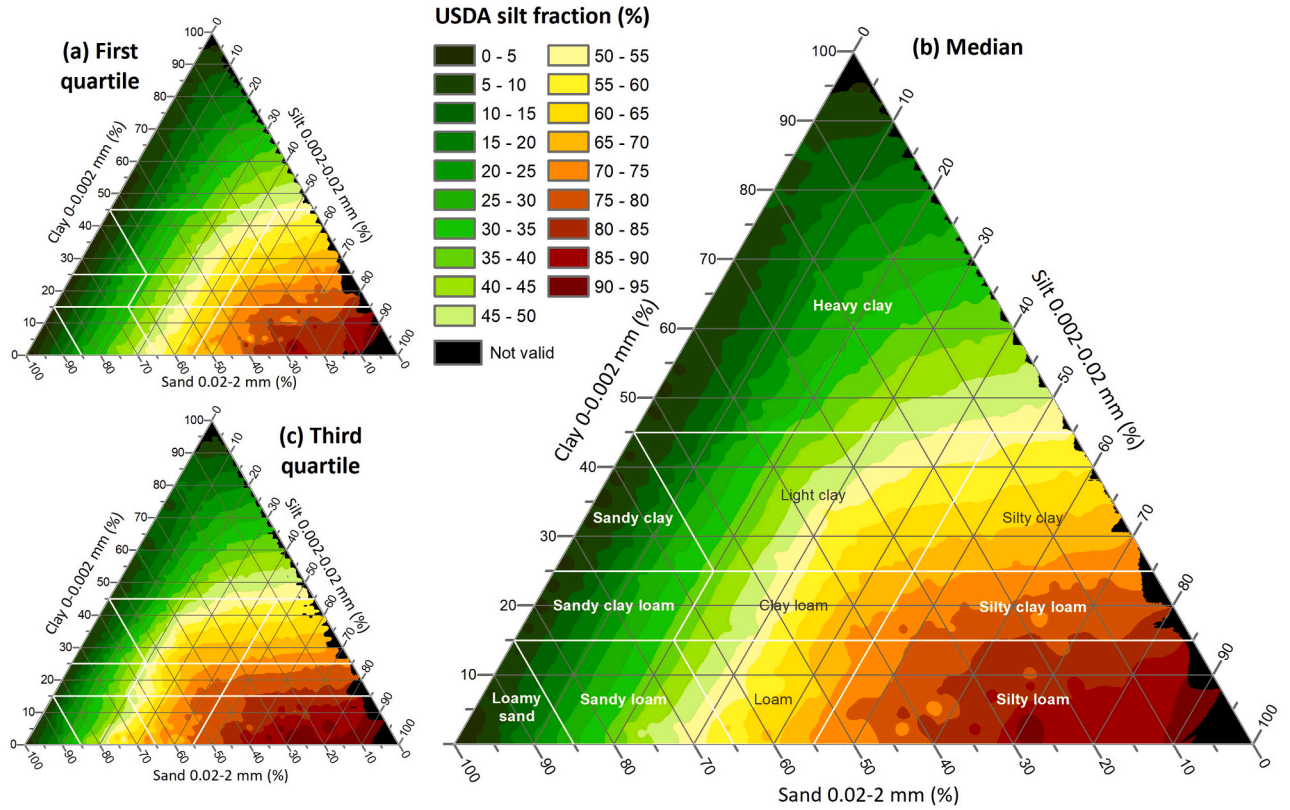


Fig. 6. Interpolated values of the conditional quartiles of the USDA silt fraction.

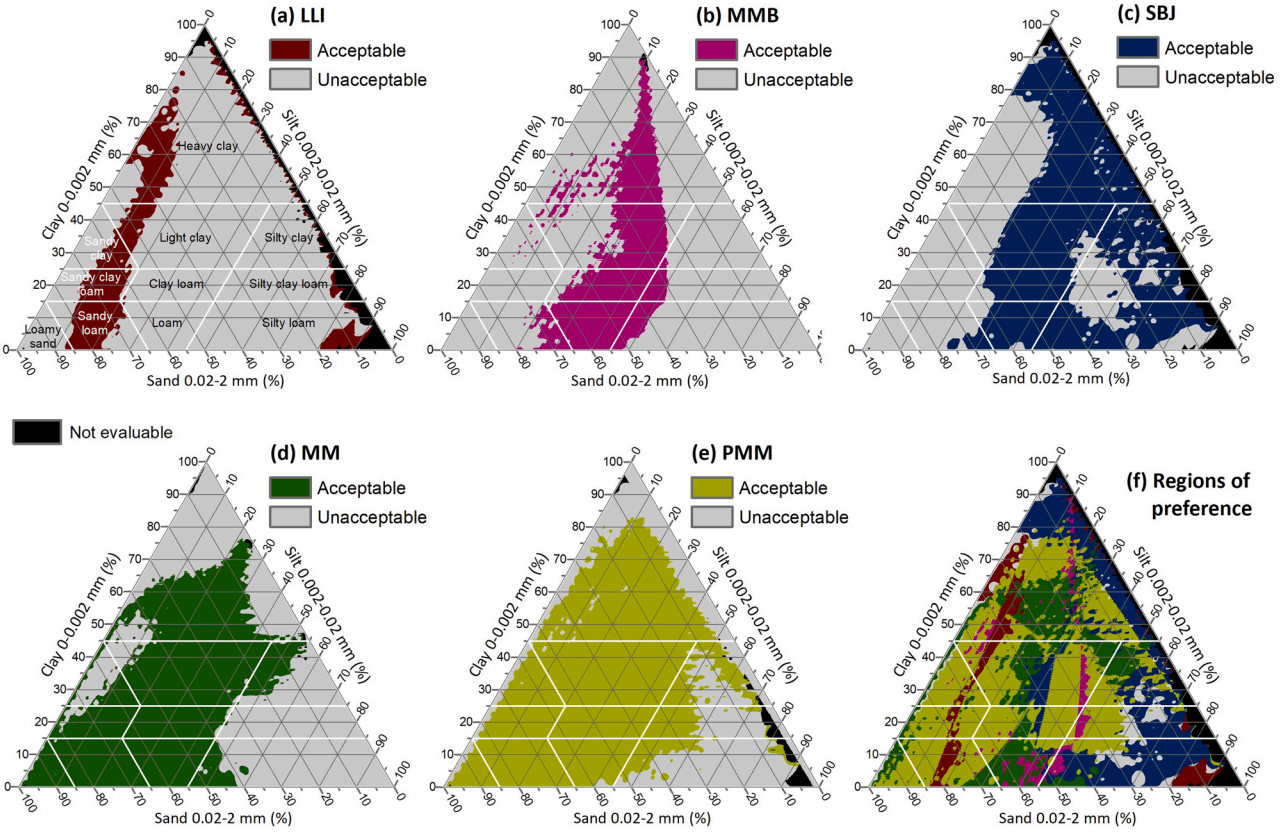


Fig. 7. Acceptance regions of the five models evaluated: (a) the log-linear interpolation method, (b) the Minasny-McBratney-Bristow regression formula, (c) the Shirazi-Boersma-Johnson interpolation method, (d) the Minasny-McBratney regression formula, and (e) the Padarian-Minasny-McBratney regression formula. (f) Preference regions of the models. The colors are the same and taken as legend for each model. It includes also the *Not evaluable* area where IDW interpolation cannot be used.

of preference, 40.3% of the textural triangle, with 62.6% of the records located in it. After that there is the MM model which shows a region of preference in 17.5% of the triangle and includes 19% of the records. The preference region of the SBJ model occupies 21.9% of the triangle and includes 8.6% of the records. The preference region of the LLI model occupies 6.4% of the triangle, where 4.7% of the records are located. The model with the smallest region of preference is MMB, occupying 3.8% of the triangle, which includes 3.4% of the records.

None of the analyzed models are acceptable for 4.9% of the textural triangle and the analysis cannot be performed in the "Not evaluable" area covering 5.2% of the textural triangle. These two areas contain hardly any records and between the two they only include 1.7%.

Fig. 4 and the associated text of section 4.1.2, regarding the density of records in the textural triangle, give us an idea of the degree of reliability of the tool developed for measuring the performance of the models. Logically, the higher the density of records, the more reliable the assessment of the models.

5. Discussion

Our preliminary analysis of the efficiency of the five models (section 3.2), which was carried out in terms of the R^2 coefficient, showed that the models with the best performance were the MM and PMM, practically tied; that the worst performing models were the LLI and SBJ; and that the MMB model had an intermediate performance.

To our knowledge, only two papers have been published that include this kind of comparisons of the performance of different procedures to transform from the IS-S to the USDA classification system: Minasny and McBratney (2001) and Padarian et al. (2012).

Thus, Minasny and McBratney (2001) compared the estimates of their MM model with those of the LLI model in terms of the ME and RMSE. For this, they used the two sets of records mentioned in section 2.2.4. Minasny and McBratney (2001) conclude that the MM model clearly provides better estimates than the LLI model, which coincides with what is deduced from the results of the present study.

Further, Padarian et al. (2012) developed the PMM model in order to improve the estimates provided by the MM model. To this end, those authors introduced two significant improvements (in principle) in the calibration phase of their model. On the one hand, they used a much greater number of records (52,432 vs. 1,210); on the other hand, they used a much more versatile fitting technique, symbolic regression via genetic programming vs. polynomial regression using least squares. The comparison, which was carried out in terms of the R^2 coefficient and the RMSE, was first based on a set of 758 records from Australia and then on a set of 55,282 records from the International Geosphere-Biosphere Programme, Data and Information System (IGBP-DIS) (Global Soil Data Task, 2000). However, the results were not as expected, as the performance of both models proved to be practically identical.

Our preliminary results, discussed above, are consistent with the results of the comparison by Padarian et al. (2012), but they have a broader scope: five models compared instead of two and data from 272,360 records.

However, the results of this preliminary analysis have been enriched and nuanced by the core of the work, which is the comparison based on the values provided by the conditional quartiles. On the one hand, it has allowed us to determine the acceptance regions (i.e., the areas of the textural triangle in which each model provides predictions that we consider acceptable). On the other hand, it has made possible the delimitation of the preference regions (i.e., the areas of the textural triangle in which the acceptable predictions provided by a specific model

would be more accurate than those of the rest of the models). Thus, for example, thanks to the conditional quartiles we can conclude that the PMM model is clearly superior to the MM model.

The LLI and MMB models are, in this order, those that have given the poorest results. Although as far as we know, there is no evidence from published works comparing the performance of the MMB model with that of other models, in fact Minasny and McBratney (2001) developed the MM model with the stated goal of exceeding the accuracy of the MMB model. The correctness of its purpose is in sight, if only for the fact that this latest model provides non-sense values of USDA silt for over more than one third of the surface of the textural triangle.

Segmented log-linear interpolation has been used with some frequency as a general interpolation method in the cumulative particle size distribution of the soil samples. However, after comparing the precision of this interpolation method with that of other general estimation methods, Nemes et al. (1999) and Shang (2013) advise against using it. In light of the results of our analysis, the same conclusion can be drawn in the case at hand: the LLI model has proven to have low accuracy when estimating the USDA silt fraction when only the IS-S clay, silt, and sand fractions are available. However, this circumstance does not have to be identical in other cases. As Moeys (2018) notes, log-linear interpolation is the simplest interpolation procedure that exists and its precision increases as the distance between the involved nodes decreases. Thus, Miao et al. (2008) achieved satisfactory results when estimating the USDA coarse silt fraction (0.02–0.05 mm) from the fine sand fraction in the IS (0.02–0.2 mm), while Shein (2009) obtained acceptable results when estimating the fraction of particles with diameters in the range of 0.001–0.002 mm from the fine silt fraction of the Kachinsky classification (0.001–0.005 mm). Further, Tietje and Hennings (1996) estimated the USDA coarse silt fraction (0.02–0.05 mm) from the coarse silt fraction of the classification used in Germany (0.02–0.063 mm) without questioning the validity, while Padarian et al. (2012) did the same to estimate the fraction in the range of 0.1–0.2 mm diameters from the USDA fine sand fraction (0.1–0.25 mm). Ultimately, the poor performance of the LLI model is mainly due to the fact that the difference between the logarithms of the diameters that define the IS-S sand fraction (0.02–2.00 mm) is too wide in comparison with, for example, the difference between the logarithms of the diameters that define the aforementioned USDA fine sand fraction (0.1–0.25 mm).

To our knowledge, the accuracy of the SBJ model has not previously been compared to that of other models. Our analysis based on quartiles reveals that this model has two virtues. As can be seen in Fig. 7, its acceptance region and the acceptance region of the PMM model complement each other very well, to the point that, if both regions were superimposed, only 14.2% of the surface of the textural triangle would remain uncovered, including 5.2% “Not evaluable”. On the other hand, its region of preference covers a not inconsiderable part of the textural triangle, namely 21.9% of its area, although this only represents 8.6% of the records. However, the model in question also has two drawbacks that make it, in our opinion, not recommendable beyond its preference region. First, the theory of Shirazi et al. (1988) on which the SBJ model is based was severely questioned from a conceptual perspective by Buchan (1989) and Yaalon (1989). Second, its mathematical formulation is awkward to use, given that the standard normal cumulative distribution function and its inverse are not part of the so-called elementary mathematical functions; in fact, Shirazi et al. (1988) expressed their interpolation method using a nomogram.

When comparing the performance of the MM and PMM models with records of the IGBP-DIS, Padarian et al. (2012) found no differences in the values of their respective R^2 coefficients, namely, 81%, in both cases. Our calculations, based on records from the NCSS Soil Characterization Database, have provided a value of the R^2 coefficient of 80.8% for the MM model and 81.1% for the PMM model. This remarkable coincidence between our results and those of Padarian et al. (2012) has a simple explanation: most of the records of the IGBP-DIS come from the NCSS Soil Characterization Database.

The analysis based on the conditional quartiles has shown that the PMM model is clearly preferable to the MM model, both in terms of acceptance region and, above all, in terms of preference region. Thus, we find that the region of preference of the PMM model extends to more than double the size of the MM model preference region: 40.3% of the surface of the textural triangle, representing 62.6% of the records, compared to 17.5% of the surface of the textural triangle, representing 19% of the records. Note that the region of preference of the PMM model represents more than half of the records, thus more records than the rest of the models considered as a whole.

The stabilization of the value of the R^2 coefficient (see Table 1), regardless of the increase in the number of fit parameters, is a sign that any reasonable model with which to estimate the USDA silt fraction from the basic textural fractions in the IS-S will encounter an impassable threshold, namely 82.1%. The fact that the value of the R^2 coefficient of the MM model, which is 80.8% according to our calculations, is very close to that limit implies that the fit improvements that other models could provide would be practically negligible in terms of R^2 and therefore also in terms of RMSE. It is for this reason that Padarian et al. (2012) did not realize the superiority of their PMM model compared to the MM model. The foregoing means that any improvement in fit measured in terms of R^2 and RMSE would entail incorporating new explanatory variables to the modeling process, provided they are available, of course. Padarian et al. (2012, eq. [4]) did this by adding a third explanatory variable: the IS fine sand fraction. Other possible options would be to use the percentage of coarse fragments (diameter greater than 2 mm) or to consider the type of parent material.

Any of the five models evaluated could be used in their respective acceptance regions to estimate the USDA silt fraction, but nothing more. The inferential statistics demonstrate that the point estimates have greater value when they are accompanied by a measurement of their uncertainty (Dagnelie, 2013, chap. 9; Meeker et al., 2017, chap. 1; Peña, 2008, chap. 8). In this sense, if the conditional distribution of the USDA silt was normal and homoskedastic, the models obtained by least squares linear regression, such as in Eq. (13), could provide us with conditional prediction intervals through the Student’s *t*-distribution. On the other hand, if the conditional distribution of the USDA silt was normal, even without being homoskedastic, we would be able to generate conditional prediction intervals by combining the information contained in Fig. 5a (conditional mean of the USDA silt fraction) and Fig. 5b (conditional standard deviation of the USDA silt fraction) with the quantiles of the Student’s *t*-distribution. The problem lies in the fact that, as can be seen in Fig. 5b, the conditional distribution of the USDA silt shows no homoskedasticity, nor approximates a normal distribution, as evidenced in section 4.1.3.

Unlike the standard least squares regression, quantile regression offers the possibility to obtain prediction intervals for the conditional distribution of the dependent variable in case of departure from the distributional assumptions of normality and homoskedasticity. Thus, each prediction interval is provided by two distinct quantile estimates, at any specified value of the explanatory variables (Davino et al., 2014, chap. 2).

The charts in Fig. 6, that is, our tools for the measurement of the performance of the models, have their origin in a local (piecewise) quantile regression, hence they can be used to provide point estimations and prediction intervals. Consequently, when it comes to estimating the USDA silt fraction, we recommend the adoption of the value provided by Fig. 6b (conditional median) along with the 50% prediction interval delimited by the values provided by Fig. 6a (first conditional quartile) and Fig. 6c (third conditional quartile). Thus, for example, if a horizon contains 90% sand, 5% silt, and 5% clay the predicted value for the USDA silt would be 11.1% (Fig. 6b), and its 50% prediction interval would be delimited by 8.9% (Fig. 6a) and 14.7% (Fig. 6c). These nomograms constitute a visual tool of great interest, although, in order to quickly and accurately obtain the USDA silt fraction values, we have created an R language program that is available at the following link: htt

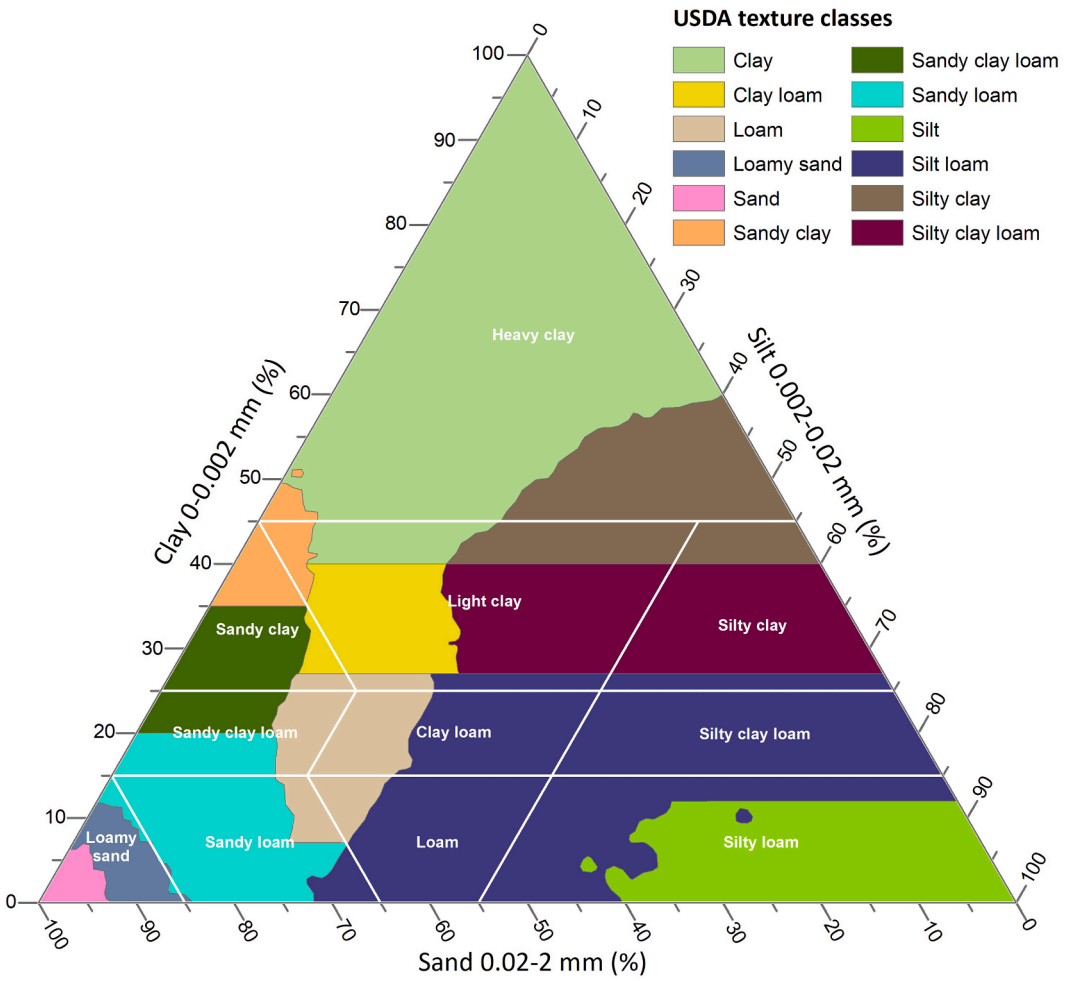


Fig. 8. USDA texture classes represented on the IS-S textural triangle; superimposed, in white color, are the limits of the texture classes proposed by Tommerup (1934).

ps://github.com/evacorral/Estimate-of-siltUSDA-fraction-from-IS-fraction.

An interesting application of the interpolated conditional median of the USDA silt fraction (Fig. 6b) is the estimation of the limits of the texture classes, according to the USDA classification system, on the textural triangle of the IS-S. Fig. 8 shows this estimate, in which the limits of the texture classes proposed by Tommerup (1934) have also been superimposed.

6. Conclusions

For each of the five models evaluated we show, on the textural triangle, both their acceptance regions (i.e., where each model provides predictions that we consider acceptable), and their preference regions (i.e., areas in which the acceptable predictions provided by a specific model would be more accurate than those of the rest of the models). Each of the models could be used for its respective acceptance region.

The above analysis allows us to conclude that the Padarian-Minasny-McBratney regression formula (PMM) is clearly the best of the evaluated models, being acceptable in more than two-thirds of the textural triangle. The performance of the PMM model, measured in terms of R^2, reaches a value close to the maximum limit of 82.1%. Any improvement measured in these terms requires the incorporation into the modeling of new explanatory variables.

The tool developed in this work, which is based on a local (piecewise) quantile regression, allows to measure the performance of any model that estimates the USDA silt fraction using the broad particle-size classes of the simplified International System (IS-S).

Further, this tool allows making point estimates of the USDA silt fraction from the broad particle-size classes in the IS-S, in addition to providing a 50% prediction interval for them. The point estimate would be obtained through the chart of the conditional median and the limits of the prediction interval would be defined by the values provided by the first and third conditional quartiles.

Declaration of Competing Interest

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

Acknowledgements

The authors would like to thank Budiman Minasny, from the University of Sydney, for his clarifications regarding some of the equations used, and to Stephan Peth, Chairman of Commission 1 (Soil physics) of Division 2 (Soil properties and processes) of the IUSS, for his indications on the textural classification system adopted by the IUSS.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Funding for Open Access charges: Universidad de Huelva/CBUA.

Appendix A:. Deduction of the calculation expression for the LLI model

Let F be the cumulative particle size distribution of a soil sample, a function of particle diameter ϕ. And let F_1 and F_2 be, respectively, the values that this distribution function takes for the particle diameters ϕ_1 and ϕ_2, where $\phi_1 \leq \phi \leq \phi_2$. Corral-Pazos-de-Provens et al. (2018, Appendix A) demonstrated that

$$F(\phi) - F_1 = \frac{\log(\phi/\phi_1)}{\log(\phi_2/\phi_1)} \bullet (\mathrm{F_2} - \mathrm{F_1}) \tag{A.1}$$

In the present case, ϕ_1= 0.02 mm, ϕ_2=2 mm, and ϕ= 0.05 mm, so that ϕ_2/ϕ_1 will equal 100, ϕ/ϕ_1 will equal 2.5, $F_2 - F_1$ will be the IS-S sand fraction, d_I, and $F(\phi) - F_1$ will be the USDA coarse silt fraction, t_c. Under these conditions, we can conclude that

$$t_c = \frac{\log 2.5}{\log 100} \bullet \mathrm{d_I} = 0.19897 \bullet \mathrm{d_I} \tag{A.2}$$

Adding the USDA coarse silt fraction to the IS silt fraction, the expression to calculate the USDA silt fraction is as follows

$$t_U = t_I + 0.19897 \bullet \mathrm{d_I}, \tag{A.3}$$

which was to be demonstrated. Note that the result obtained is independent of the logarithmic base used.

Appendix B:. Deduction of the calculation expression for the MMB model

The equation of Minasny et al. (1999), with the values of their fitting parameters transformed to adapt to the scale used in this work, is as follows:

$$t_c = 0.484593 - 0.2225 \bullet \mathrm{d_I} - 0.29 \bullet \mathrm{d_I}^2 - 0.6952 \bullet \mathrm{y} + 0.18 \bullet \mathrm{y}^2, \tag{B.1}$$

where t_c is the USDA coarse silt fraction

Recalling that the sum of the three broad particle-size classes is equal to 1, we can isolate the clay value as function of the other two textural fractions

$$y = 1 - d_I - t_I \tag{B.2}$$

Substituting the clay value in Eq. (B.1) and taking into account that the USDA silt fraction is obtained by adding the values of the IS silt fraction and the USDA coarse silt fraction

$$t_U = t_I + t_c, \tag{B.3}$$

the equation for obtaining the USDA silt fraction will be

$$t_U = -0.030607 + 1.3352 \bullet \mathrm{t_I} + 0.1127 \bullet \mathrm{d_I} + 0.18 \bullet \mathrm{t_I}^2 + 0.36 \bullet \mathrm{t_I} \bullet \mathrm{d_I} - 0.11 \bullet \mathrm{d_I}^2 \tag{B.4}$$

Appendix C:. Deduction of the calculation expression for the SBJ model

Let F be the cumulative particle-size distribution of a soil sample for diameters ϕ between ϕ_1= 0.02 mm and ϕ_2= 2 mm, such that $F_2 = F(\phi_2) = 0.9999$ and $F_1 = F(\phi_1) = 1 - d_I$, where d_I is the IS-S sand fraction of the sample. Corral-Pazos-de-Provens et al. (2018, Appendix B) demonstrate that

$$F(\phi) - F_1 = \Phi\left(\frac{\ln(\phi/\phi_1)}{\ln(\phi_2/\phi_1)}\left[\Phi^{-1}(F_2) - \Phi^{-1}(F_1)\right] + \Phi^{-1}(F_1)\right) - F_1, \tag{C.1}$$

where Φ is the normal standard distribution and Φ^{-1} is the inverse function of Φ.

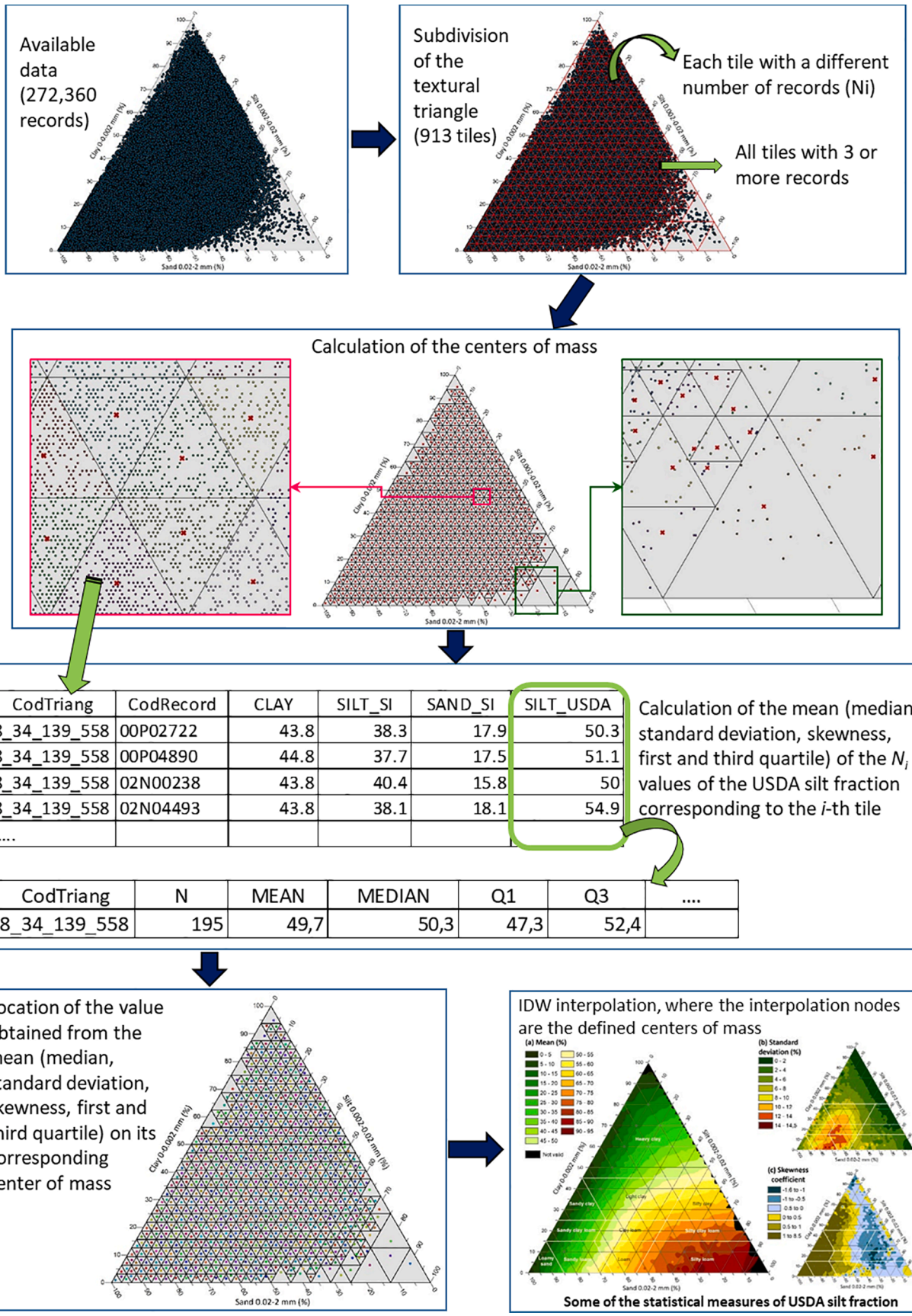
Substituting the variables ϕ_1, ϕ_2, F_1, and F_2 for their respective values and expressions, and considering that, in the present case, $\phi = 0.05$ mm, so that $F(\phi) - F_1$ will be the USDA coarse silt fraction, t_c, we can deduce that

$$t_c = \Phi\left(\frac{\ln 2.5}{\ln 100} \bullet \left[\Phi^{-1}(0.9999) - \Phi^{-1}(1 - \mathrm{d_I})\right] + \Phi^{-1}(1 - \mathrm{d_I})\right) - 1 + \mathrm{d_I} \tag{C.2}$$

Finally, solving the equation (C.2) and adding the USDA coarse silt fraction to the IS silt fraction, the expression for calculating the USDA silt fraction is as follows:

$$t_U = t_I + \Phi\left[0.739973 + 0.801030 \bullet \Phi^{-1}(1 - \mathrm{d_I})\right] - 1 + \mathrm{d_I} \tag{C.3}$$

Appendix D:. Graphical flow chart on the procedure for textural triangle subdivision and calculation of conditional statistical measures



CodTriang	CodRecord	CLAY	SILT_SI	SAND_SI	SILT_USDA
8_34_139_558	00P02722	43.8	38.3	17.9	50.3
8_34_139_558	00P04890	44.8	37.7	17.5	51.1
8_34_139_558	02N00238	43.8	40.4	15.8	50
8_34_139_558	02N04493	43.8	38.1	18.1	54.9
.....					

CodTriang	N	MEAN	MEDIAN	Q1	Q3	....
8_34_139_558	195	49,7	50,3	47,3	52,4	

References

Atterberg, A., 1905. Die rationalle Klassifikation der Sande und Kiese. Chem. Zeitung 29, 195–198.

Blott, S.J., Pye, K., 2012. Particle size scales and classification of sediment types based on particle size distributions: Review and recommended procedures. Sedimentology 59 (7), 2071–2096. https://doi.org/10.1111/j.1365-3091.2012.01335.x.

Buchan, G.D., 1989. Comments on "A Unifying Quantitative Analysis of Soil Texture: Improvement of Precision and Extension of Scale." Soil Sci. Soc. Am. J. 53, 594–594. https://doi.org/10.2136/SSSAJ1989.03615995005300020052X.

Bulmer, M.G., 1979. Principles of Statistics. Dover, New York.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed. Springer, New York.

Corral-Pazos-de-Provens, E., Domingo-Santos, J.M., Rapp-Arrarás, Í., 2018. Estimating the very fine sand fraction for calculating the soil erodibility K -factor. L. Degrad. Dev. 29 (10), 3595–3606. https://doi.org/10.1002/ldr.3121.

Dagnelie, P., 2011. Statistique théorique et appliquée, 2: Inférence statistique à une et à deux dimensions, 3rd ed. De Boeck, Bruxelles.

Dagnelie, P., 2013. Statisque théorique et appliquée, 1: Statistique descriptive et bases de l'inférence statistique, 3rd ed. De Boeck Supérieur, Bruxelles.

Davino, C., Furno, M., Vistocco, D., 2014. Quantile Regression: Theory and Applications. John Wiley & Sons, Chichester, United Kingdom.

Geeves, G.W., Cresswell, H.P., Murphy, B.W., Gessler, P.E., Chartres, C.J., Little, I.P., Bowman, G.M., 1995. The physical, chemical and morphological properties of soils in the wheat-belt of southern NSW and Northern Victoria. CSIRO Division of Soils, Glen Osmond, Australia.

Global Soil Data Task, 2000. Global Soil Data Products CD-ROM (IGBP-DIS), CD-ROM. International Geosphere-Biosphere Programme, Data and Information System, Potsdam, Germany.

Hao, L., Naiman, D.Q., 2007. Quantile Regression. Quantitative Applications in the Social Sciences No. 149. SAGE Publications, Thousand Oaks, California. https://doi.org/10.4135/9781412985550.

Haverkamp, R., Zammit, C., Bouraoui, F., Rajkai, K., Arrue, J.L., Heckman, N., 1997. GRIZZLY, Grenoble Soil Catalogue. Soil survey of field data and description of particle size, soil water retention and hydraulic conductivity functions. Laboratoire d'Ètude des Transfers en Hydrologie et Environnement: Grenoble, France.

Hughes, P., McBratney, A.B., Huang, J., Minasny, B., Micheli, E., Hempel, J., 2017. Comparisons between USDA Soil Taxonomy and the Australian Soil Classification System I: Data harmonization, calculation of taxonomic distance and inter-taxa variation. Geoderma 307, 198–209. https://doi.org/10.1016/j.geoderma.2017.08.009.

ISSS, 1929. The minutes of the First Commission meetings, International Congress of Soil Science, Washington 1927. Proc. Int. Soc. Soil Sci. 4 (3), pp. 215–222.

IUSS Working Group WRB, 2015. 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 No. 106. FAO, Rome.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning: With applications in R, Springer Texts in Statistics. Springer, New York.

Katschinski, N.A., 1956. Die mechanische Bodenanalyse und die Klassifikation der Böden nach ihrer mechanischen Zusammensetzung. Rapports du Sixième Congrés International de la Science du Sol, B. Laboureur & Cie, Paris, pp. 321–327.

Knight, H.G., 1938. New size limits for silt and clay. Soil Sci. Soc. Am. Proc. 2, 592.

Leij, F.J., Alves, W.J., Van Genuchten, M.T., Williams, J.R., 1996. In: The UNSODA Unsaturated Soil Hydraulic Database: User's Manual Version 1.0 (EPA/600/R-96/095). National Risk Management Research Laboratory. Office of Research and Development. U.S. Environmental Protection Agency, Cincinnati, Ohio.

Meeker, W.Q., Hahn, G.J., Escobar, L.A., 2017. Statistical intervals: a guide for practitioners and researchers, 2nd ed. John Wiley & Sons, Hoboken, New Jersey.

Miao, C., Liu, B., Gao, Y., Zhang, T., 2008. Evaluation of different procedures to interpolate particle size distribution in black soils. Int. J. Sustain. Dev. World Ecol. 15 (sup1), 56S–62S. https://doi.org/10.3843/SusDev.15.4:8.

Minasny, B., McBratney, A.B., 2007. Estimating the Water Retention Shape Parameter from Sand and Clay Content. Soil Sci. Soc. Am. J. 71 (4), 1105–1110. https://doi.org/10.2136/sssaj2006.0298N.

Minasny, B., McBratney, A.B., 2001. The Australian soil texture boomerang: a comparison of the Australian and USDA/FAO soil particle-size classification systems. Aust. J. Soil Res. 39, 1443–1451. https://doi.org/10.1071/SR00065.

Minasny, B., McBratney, A.B., 2000. Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. Aust. J. Soil Res. 38, 905–926. https://doi.org/10.1071/SR99110.

Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93 (3-4), 225–253. https://doi.org/10.1016/S0016-7061(99)00061-0.

Moeys, J., 2018. The soil texture wizard: R functions for plotting, classifying, transforming and exploring soil texture data. http://cran.ms.unimelb.edu.au/web/packages/soiltexture/vignettes/soiltexture_vignette.pdf.

NCSS, 2017. National Cooperative Soil Survey Characterization Database. http://ncsslabdatamart.sc.egov.usda.gov/ (accessed 25 September 2017).

Nemes, A., Rawls, W.J., 2006. Evaluation of different representations of the particle-size distribution to predict soil water retention. Geoderma 132 (1-2), 47–58. https://doi.org/10.1016/j.geoderma.2005.04.018.

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 (3-4), 187–202. https://doi.org/10.1016/S0016-7061(99)00014-2.

Osborne, T.B., 1887. The methods of mechanical analysis. In: State of Connecticut, Annual Report of the Connecticut Agricultural Experiment Station for 1886. Tuttle, Morehouse & Taylor, New Haven, Connecticut, pp. 141–158.

Padarian, J., Minasny, B., McBratney, A., 2012. Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. Soil Res. 50, 443–446. https://doi.org/10.1071/SR12139.

Peña, D., 2008. Fundamentos de estadística, 2nd ed. Alianza Editorial, Madrid.

Porta Casanellas, J., López-Acevedo Reguerin, M., Roquero de Laburu, C., 2003. Edafología para la agricultura y el medio ambiente, 3rd ed. Mundi-Prensa, Madrid.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richer-de-Forges, A., Feller, C., Jamagne, M., Arrouays, D., 2008. Perdus dans le triangle des textures. Étude et Gestion des Sols 15 (2), 97–111.

Royston, P., 1992. Approximating the Shapiro-Wilk W-test for non-normality. Stat. Comput. 2 (3), 117–119. https://doi.org/10.1007/BF01891203.

Sakaguchi, A., Eguchi, S., Kasuya, M., 2014. Examination of the water balance of irrigated paddy fields in SWAT 2009 using the curve number procedure and the pothole module. Soil Sci. Plant Nutr. 60 (4), 551–564. https://doi.org/10.1080/00380768.2014.919834.

Shang, S., 2013. Log-Cubic Method for Generation of Soil Particle Size Distribution Curve. Sci. World J. 2013, 1–7. https://doi.org/10.1155/2013/579460.

Shein, E.V., 2009. The particle-size distribution in soils: Problems of the methods of study, interpretation of the results, and classification. Eurasian Soil Sci. 42 (3), 284–291. https://doi.org/10.1134/S1064229309030053.

Shirazi, M.A., Boersma, L., Hart, J.W., 1988. A Unifying Quantitative Analysis of Soil Texture: Improvement of Precision and Extension of Scale. Soil Sci. Soc. Am. J. 52 (1), 181–190. https://doi.org/10.2136/sssaj1988.03615995005200010032x.

Shirazi, M.A., Boersma, L., Johnson, C.B., 2001. Particle-Size Distributions: Comparing texture systems, adding rock and predicting soil properties. Soil Sci. Soc. Am. J. 65 (2), 300–310. https://doi.org/10.2136/sssaj2001.652300x.

Smettem, K., Gregory, P., 1996. The relation between soil water retention and particle size distribution parameters for some predominantly sandy Western Australian soils. Aust. J. Soil Res. 34, 695–708. https://doi.org/10.1071/SR9960695.

Soil Survey Staff, 1995. Soil Characterization and Profile Description Data. CD-ROM. Soil Survey Laboratory, Natural Resources Conservation Service, USDA, Lincoln, Nebraska.

Soil Survey Staff, 1951. Soil Survey Manual. Agriculture Handbook 18. Agricultural Research Administration, USDA, Washington, D.C.

Tietje, O., Hennings, V., 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69 (1-2), 71–84. https://doi.org/10.1016/0016-7061(95)00050-X.

Tommerup, E.C., 1934. The field description of the physical properties of soils. In: Comptes Rendus de la Conférence de la Première Commission (Physique du Sol) tenue au Centre National de Recherches Agronomiques de Versailles, (2-5 juillet 1934). Imprimerie Nationale, Paris, France, pp. 155–158.

USDA-NRCS, 1997. National Soil Characterization Database. USDA-NRCS Soil Survey Division.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Statistics and Computing, 4th ed. Springer, New York https://doi.org/10.1007/978-0-387-21706-2.

Yaalon, D.H., 1989. Comments on "A unifying quantitative analysis of soil texture." Soil Sci. Soc. Am. J. 53, 595.