Geoderma Regional 12 (2018) 18–27

Contents lists available at ScienceDirect

Geoderma Regional

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

A spatial dataset of topsoil texture for the southern Argentine Pampas

Mauricio Castro-Franco[a,b,*], Marisa Beatriz Domenech[b], Marta Renée Borda[b], José Luis Costa[c]

[a] Consejo Nacional de Investigaciones Científicas y Técnicas CONICET, Av. Rivadavia 1917 (C1033AAJ), Buenos Aires, Argentina
[b] Instituto Nacional de Tecnología Agropecuaria INTA, CEI Barrow, Ruta 3 Km 488 (7500), Tres Arroyos, Argentina
[c] Instituto Nacional de Tecnología Agropecuaria INTA, EEA Balcarce, Ruta 226, Km 73.5, Balcarce, Argentina

1. Introduction

Soil texture governs essential physical and chemical soil processes such as soil water storage, nutrient retention, soil structure and stability, and soil erosion (Arya and Paris, 1981; Minasny and McBratney, 2007; Wang and Shi, 2017). Hence, accurate information about the spatial distribution of soil texture at regional, sub-regional and local scales allows a better understanding of the importance of the soil as a fundamental resource for agriculture and the environment and is required for climate, hydrological, ecological and crop models (Arya et al., 1999; Saxton et al., 1986; Shangguan et al., 2012). However, in many agricultural regions in the world, particularly in the Argentine Pampas, this information is scarce (Berhongaray et al., 2013; McBratney et al., 2003). Hence, it is crucial to optimize and update soil cartography especially that required for modeling.

In Argentina, the soil cartography generated by the National Institute of Agricultural Technology (INTA) at a scale of 1:50,000 is the main source of soil spatial information for the Argentine (INTA, 2013b). However, this soil dataset is based on limited profile data, which, in terms of accuracy, scale, coverage and digital format, is not appropriate to apply in the type of models mentioned above. The cartography generated by the INTA maps soil types by using polygons of classification, without considering the spatial-temporal variability of soil properties. Thus, there can be a large degree of uncertainty in the accuracy of soil properties labeled in each polygon (Castro Franco et al., 2015). To overcome this issue, it is necessary to generate a spatial dataset of soil texture by using existing information and information that allows developing continuous and accurate maps of soil texture in digital format.

Digital soil mapping (DSM) techniques offer a promising approach to generate a spatial dataset of soil properties based on data such as digital elevation models, remotely sensed data, physical and chemical attributes obtained through laboratory analysis of soil samples, and information from existing soil maps (Boettinger, 2010; Hartemink et al., 2008; McBratney et al., 2003). DSM is defined as the use of quantitative methods to establish a spatial relationship between soil ancillary variables (represented by soil-forming factors) and soil properties, according to the scale (Cook et al., 2008). These models allow considering two essential aspects needed to produce soil digital cartography: i) that spatial predictions of soil properties can be made at unobserved locations at a defined scale, and ii) that the uncertainty of prediction can be quantified (Ließ et al., 2012).

Several DSM techniques have been proposed to generate spatial datasets of soil texture at different scales and world zones (Adhikari et al., 2013; Akpa et al., 2014; Buchanan et al., 2012; Odeh et al., 2003; Shangguan et al., 2012). Among these, the machine learning (ML) algorithm stands out because of its accurate prediction of soil texture by using a limited number of soil samples (Akpa et al., 2014). Specifically, Cubist model is a ML algorithm frequently used in pedometry due to its efficiency, classification of important variables, and versatility with a large number of predictors (Malone et al., 2016). So far, there are no studies on the use of RF to predict texture and other properties in soils of the Argentine Pampas at a regional scale, where the demand for soil information is increasing (Angelini et al., 2016; Domenech et al., 2017).

The aim of this study was to produce a 1-km spatial resolution dataset of soil texture in digital format for the southern Argentine Pampas. To this end, DSM techniques and soil existing information were coupled to predict soil particle-size fraction and USDA texture classification. We expect that our soil texture digital maps are suitable for climate, hydrological, ecological and crop models. The relevance of DSM techniques offers a novel approach to optimize the predictions of soil properties in the conditions of the Argentine Pampas.

2. Materials and methods

2.1. Study area

The study area covered 119,155 km^2 of the southern Pampean region, province of Buenos Aires, from 36°–39° S to 56°–63° W (Fig. 1). The land surface is basically plain, interrupted by Olavarría, Azul and Juárez hills in the northwest, which correspond to fractions of the Tandilia hilly system, and by Ventania and Pillahuinco systems towards the center-west. These hilly systems provide an elevation in the extended plain and are called Sierras Australes. The climate is mesothermal humid-sub-humid. The mean temperature is in the range of 10.2–15.8 °C yr^{-1}, while seasonal rainfall distribution varies from 500 to 950 mm yr^{-1}, with a maximum

* Corresponding author at: Ruta 3 Km 488 CC 50 (7500), Tres Arroyos, Argentina.
E-mail address: castrof.mauricio@inta.gob.ar (M. Castro-Franco).

https://doi.org/10.1016/j.geodrs.2017.11.003
Received 17 June 2017; Received in revised form 2 November 2017; Accepted 8 November 2017
Available online 10 November 2017

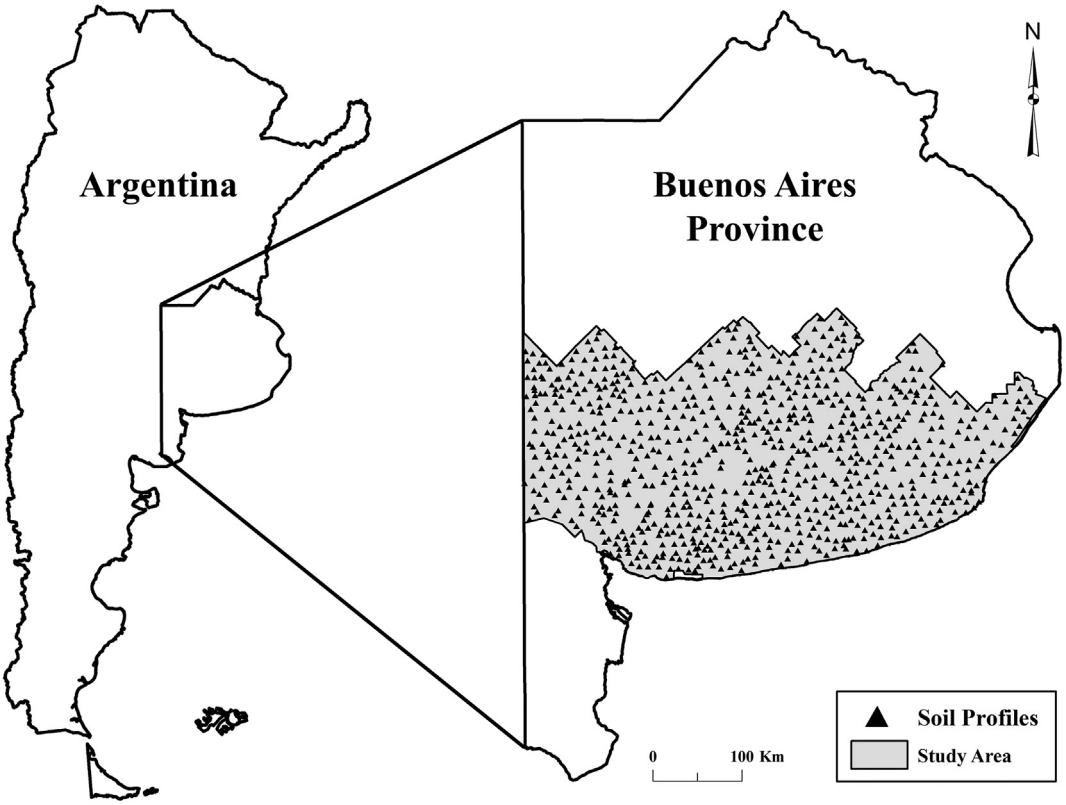

Fig. 1. Spatial distribution of soil profiles (n = 813) over the southern Argentine Pampas.

pattern (> 60%) during spring and summer. Rainfall is distributed from east to west, which leads to a more extreme climatic condition further west in the region (Amiotti et al., 2001; Blanco and Stoops, 2007). The main soils are classified as Argiudoll, Haplustoll, Hapludoll and Natraquoll (Soil Survey Staff, 2014). The soils of the study region are dominated by Argiudolls and are developed over loess sediments with grassland, organic matter content between 4% and 5.5% in the upper horizon and an argillic horizon whose clay content decreases with depth (Matteucci, 2012; Pazos, 1984). Below these horizons, there is usually a continuous caliche layer, locally termed *tosca*, which has great variability in $CaCO_3$ content, depth, structure and degree of induration (Pazos and Mestelan, 2002). The main land uses in the study area are agriculture and livestock. The main crops are cereals and oilseeds such as maize, sunflower, soybean, wheat and barley.

2.2. Soil data

Soil profile data were obtained from the Atlas of soil conditions of Buenos Aires province (MAA, 1965) and INTA conventional cartography at a 1:50,000 scale (INTA, 2013a), which were generated during 1960 and 1970, respectively. In total, the soil textural data from 813 soil profiles were converted to digital format. Specifically, the clay (< 0.002 mm), silt (0.002–0.02 mm) and sand (0.2–2 mm) particle-size fractions at 0–20 cm and 20–40 cm depth were used and expressed in percentage mass (g 100 g^{-1}). Original geographic coordinates of soil profiles were transformed from the International 1924 datum to the UTM coordinate system datum WGS84, using QGIS v2.16.1 (QGIS v2.16.1 Nodebo, 2016). The USDA was used as the texture classification system (Soil Survey Staff, 2014).

2.3. Predictors

A total of thirty-seven environmental variables were used as predictors of particle-size fractions at 0–20 cm and 20–40 cm depth. These predictors were divided in three types: those obtained from a digital elevation model (DEM), bioclimatic variables, and spectral indices from satellite images (Table 1).

A 90-m spatial-resolution DEM obtained from NASA Shuttle Radar Topography Mission data was used. The DEM was then hydrologically corrected by identifying and filling the sinks with 0.01° minimum slope and using Wang and Liu (2006) algorithm. This post-process is important when working with water flow-related calculation within the study area (Adhikari et al., 2013). Then, fourteen terrain and hydrological indices were calculated based on the corrected DEM. All procedures of pre-process and indices calculation were carried out using the Terrain Analysis module of SAGA-GIS v4.0.1 (SAGA Development Team, 2016).

The eighteen bioclimatic variables were obtained based on the WorldClim database (Hijmans et al., 2005a, 2005b). Original raster layers of these variables were used without modification (Table 1). Several studies have demonstrated that these bioclimatic variables have a high accuracy to predict soil properties at regional scale (Cambule et al., 2013; Hartemink et al., 2008). Hence, they have been useful as inputs to generate models of DSM, especially where geospatial information of ancillary soil variables is limited (Vaysse and Lagacherie, 2015).

The spectral vegetative indices used as predictors were satellite-derived Normalized Difference Vegetation Index (NDVI) from Landsat 5 MSS-TM and Landsat 8 OLI-TIRS images with atmospheric correction over a time period of four years (Table 1). These indices were calculated using QGIS v2.16.1 (QGIS v2.16.1 Nodebo, 2016). Various studies have reported that NDVI properly predicts land cover and has been useful as ancillary information to predict many soil properties at different scales (Boettinger et al., 2008; McBratney et al., 2003).

All predictors were adjusted to the UTM coordinate system datum WGS84. Table 1 shows the mean and range for each predictor within the study area.

Table 1
Description of environmental variables used to predict particle-size fractions over the southern Argentine Pampas.

Predictor	Brief description	Data source	Spatial resolution	Mean (Range)
Digital elevation model (DEM)	Elevation in meters	DEM-SRTM	90 m	155 (0–1186)
Slope	Slope gradient in percentage	DEM-SRTM	90 m	0.01 (1.6×10^{-8}-71)
Aspect	Grid cells of aspect or facing direction in radians	DEM-SRTM	90 m	3.2 (7.5×10^{-6}-6.3)
Flow Acummulation (Flow_accum)	Number of upslope cells	DEM-SRTM	90 m	5.7×10^5 (7350.8–8.9×10^8)
Hillshading	Angle between the surface and the incoming light beams, measured in radians	DEM-SRTM	90 m	0.8 (0.05–2.1)
Catchment area (Catch_area)	Discharge contributing upslope area of each grid cells	DEM-SRTM	90 m	2.4×10^5 (7350–2.9×10^8)
Channel network distance	Calculates the vertical distance to a channel network base level	DEM-SRTM	90 m	151.1 (0–554.6)
LS_Factor	Index of soil erosion using the modified LS Factor for universal soil loss equation	DEM-SRTM	90 m	0.1 (0–14.2)
Relative Slope Position (RSP)	Height of a given location above the nearest downslope channel relative to the height of the nearest upslope crest above the same point	DEM-SRTM	90 m	0.1 (0–1)
Topographic Wetness Index (TWI)	TWI = ln (As/tan b); where *As* is upslope catchment area and b is the slope gradient.	DEM-SRTM	90 m	11.4 (0–25.7)
Terrain Ruggedness Index (TRI)	Indicates terrain heterogeneity	DEM-SRTM	90 m	1 (0–53.4)
SAGA TWI	Same as TWI but uses modified catchment area instead	DEM-SRTM	90 m	11 (3.5–14.5)
Valley depth	Relative position of the valley	DEM-SRTM	90 m	78.9 (0–412.7)
Vertical distance to a channel network base level (Vertical_dist)	Calculates vertical distance to channel network for each cell	DEM-SRTM	90 m	4.8 (0–741.7)
Wind exposition index (Wind_exp)	Indicate areas exposed to wind	DEM-SRTM	90 m	1 (0.8–1.32)
BioClim1	Annual Mean Temperature in °C	WorldClim-Bioclimatic variables	1 km	14.1 (10.2–15.8)
BioClim2	Mean diurnal range (Mean of monthly (max temp - min temp)) in °C	WorldClim-Bioclimatic variables	1 km	12.7 (9.4–14.7)
BioClim4	Temperature Seasonality (SD *100)	WorldClim-Bioclimatic variables	1 km	50.5 (40.9–57.7)
BioClim5	Maximum temperature of warmest month in °C	WorldClim-Bioclimatic variables	1 km	29.7 (25.3–32.5)
BioClim6	Minimum temperature of coldest month in °C	WorldClim-Bioclimatic variables	1 km	2.7 (− 1.5–4.8)
BioClim7	Temperature annual range in °C	WorldClim-Bioclimatic variables	1 km	27.2 (21.9–31.4)
BioClim9	Mean temperature of driest quarter in °C	WorldClim-Bioclimatic variables	1 km	7.9 (3.9–17.2)
BioClim10	Mean temperature of warmest quarter in °C	WorldClim-Bioclimatic variables	1 km	20.6 (17–23)
BioClim11	Mean temperature of coldest quarter in °C	WorldClim-Bioclimatic variables	1 km	7.8 (3.9–9.2)
BioClim12	Annual precipitation	WorldClim-Bioclimatic variables	1 km	780 (500–950)
BioClim13	Precipitation of wettest month	WorldClim-Bioclimatic variables	1 km	98.7 (65–126)
BioClim14	Precipitation of driest month	WorldClim-Bioclimatic variables	1 km	34 (16–58)
BioClim15	Precipitation Seasonality (CV)	WorldClim-Bioclimatic variables	1 km	28.4 (12–42)
BioClim16	Precipitation of wettest quarter	WorldClim-Bioclimatic variables	1 km	251.3 (164–323)
BioClim17	Precipitation of driest quarter	WorldClim-Bioclimatic variables	1 km	121.3 (58–187)
BioClim18	Precipitation of warmest quarter	WorldClim-Bioclimatic variables	1 km	230.9 (148–299)
BioClim19	Precipitation of coldest quarter	WorldClim-Bioclimatic variables	1 km	121.8 (58–211)

[†]NDVI, Normalized Difference Vegetation Index; CV, coefficient of variation; SD, standard deviation.

Table 2
Descriptive statistics of particle-size fractions from total, calibration and validation dataset of soil profiles over the southern Argentine Pampas.

	Minimum	Maximum	Mean	CV (%)	Skewness	Kurtosis
0–20 cm soil depth						
Clay (%)						
Total dataset	4.00	71.00	20.50	29.32	1.18	− 0.70
Calibration dataset	11.30	33.60	20.82	26.61	0.23	− 0.93
Validation dataset	11.62	33.01	19.67	26.76	0.64	− 0.58
Silt (%)						
Total dataset	5.00	57.00	26.77	35.84	0.13	− 0.07
Calibration dataset	5.00	57.00	26.52	37.01	0.17	− 0.04
Validation dataset	7.00	53.00	27.36	33.20	0.06	− 0.20
Sand (%)						
Total dataset	1.00	82.00	52.75	21.81	0.15	0.29
Calibration dataset	1.00	82.00	52.97	22.21	0.01	0.41
Validation dataset	31.00	82.00	52.24	20.80	0.51	− 0.09
20–40 cm soil depth						
Clay (%)						
Total dataset	9.65	46.10	24.77	27.39	0.02	− 0.48
Calibration dataset	9.65	46.10	24.48	27.16	0.06	− 0.42
Validation dataset	9.88	43.57	25.48	26.62	0.04	− 0.54
Silt (%)						
Total dataset	2.62	80.80	27.05	38.40	0.29	0.46
Calibration dataset	2.62	56.22	26.95	37.89	0.16	− 0.26
Validation dataset	4.80	80.80	27.29	39.81	0.54	1.67
Sand (%)						
Total dataset	14.18	87.71	48.22	26.31	0.66	0.05
Calibration dataset	21.27	87.71	48.68	26.01	0.68	0.05
Validation dataset	14.18	82.51	47.15	26.98	0.61	0.02

†. SD, standard deviation; CV, coefficient of variation.

Table 3
Fitting performance for the particle-size fractions predictive models.

Evaluation		Clay	Silt	Sand
		0–20 cm soil depth		
Calibration dataset	RMSE (%)	4.94	7.85	9.19
	R^2	0.19	0.35	0.39
	r_c	0.39	0.56	0.59
	Bias	− 0.23	− 0.34	− 0.03
Validation dataset	RMSE (%)	5.78	8.95	10.86
	R^2	0.07	0.18	0.18
	r_c	0.23	0.38	0.40
	Bias	− 0.89	− 0.32	− 0.36
		20–40 cm soil depth		
Calibration dataset	RMSE (%)	6.20	8.71	10.33
	R^2	0.17	0.32	0.36
	r_c	0.36	0.53	0.56
	Bias	0.23	− 0.35	− 0.51
Validation dataset	RMSE (%)	6.67	9.51	11.63
	R^2	0.11	0.15	0.24
	r_c	0.28	0.36	0.46
	Bias	− 0.29	− 0.76	− 1.42

2.4. Prediction of particle-size fractions

The Cubist model was used as a regression tool to determine particle-size fractions based on predictors. This model is a rule-based model, similar to a regression tree model, which the data are partitioned into smaller subsets based on the target variable (Quinlan, 1992). Cubist produces a set of "if-then-else" rules, where each condition is based on a threshold for one or more predictors (Malone et al., 2014). The result consists of a number of rules each associated with a multiple regression that only applies when the conditions specified on the rule are met (Adhikari et al., 2014). Cubist model has been used in several fields and has been effective for various purposes of soil mapping over large areas (Malone et al., 2014; Viscarra Rossel et al., 2015). For more details on Cubist model, see Quinlan (1992) and Henderson et al. (2005).

2.4.1. Prediction performance

The Cubist model perform was evaluated on two data sets: calibration (internal evaluation) and validation (external evaluation) data sets. Calibration data set was generated using five-hundred-sixty-eight soil profiles (~70%), which were randomly selected from the total data set. Validation set was generated using two-hundred-forty soil profiles (~30%), which were randomly excluded from the total data set, as an independent data set. Internal and external evaluations were calculated using the following error metrics: Root Mean Square Error (RMSE) (Eq. (1)), Coefficient of determination (R^2) (Eq. (2)) and Lins's Concordance Correlation Coefficient (r_c) (Eq. (3)). These error metrics were calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (1)$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})} \quad (2)$$

$$\rho_c = \frac{2S_{xy}}{S_x^2 + S_y^2 + (\bar{x} - \bar{y})^2} \quad (3)$$

where, $S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, $S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$. For all error metrics equations, $\bar{x}$ and $\bar{y}$ are the mean value for samples x (measured particle-size fraction) and y (predicted particle-size fraction) of size of number of samples n, x_i and y_i are paired ith values from the sample x and y.

RMSE measures is a measurement of the average error (or accuracy) of the prediction. R^2 measures the agreement between measured and predicted data. Generally, a good model will have a RMSE close to 0 and an R^2 close to 1. The r_c assess covariation and correspondence between measured and predicted data. This metric combines measures of both precision and bias to determine how fat the measured data deviate from the line of perfect concordance, which is the 1:1. The value of this metric is scaled between 1 and − 1. A value < 0.60 denoted poor agreement, value between 0.60 and 0.75 moderate agreement, between 0.75 and 0.90 substantial agreement and value of + 1 perfect agreement (Viscarra Rossel et al., 2015).

2.4.2. Estimates of uncertainty

The prediction uncertainties was estimated using the method of non-parametric bootstrap (Efron and Tibshirani, 1993). This method has been used in the spatial modeling for three purposes (Viscarra Rossel et al., 2015). First, provide sets of residuals on which model the random component. Second, bootstrapping involves repeated random sampler with replacement of the available data. Then, the spatial modeling on each bootstrap sample is perform, obtains probability distributions of the out comes from the modeling at each pixel. The result is robust estimates by averaging the predictions made on the different bootstrap samples. And third, with the probability distributions is possible quantify the uncertainty of the modeling by computing a prediction interval given a specified level of confidence (Malone et al.,

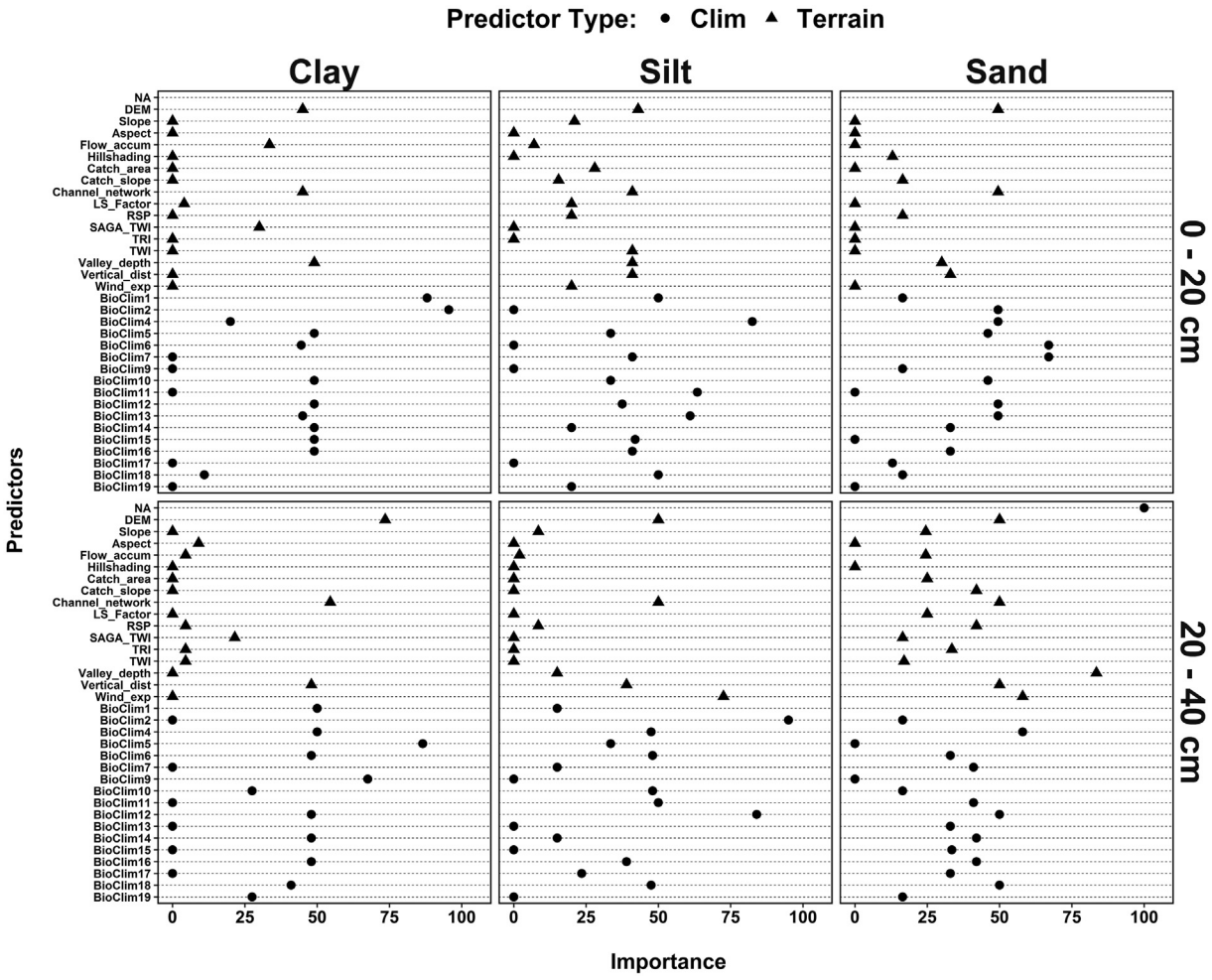

Fig. 2. Predictor importance plots based on Cubist regression tree of particle-size fractions.

2016).

In this study, 50 bootstrap samples of particle-size fractions and associated predictors were used. For each bootstrap sample, the Cubist model was used to predict the values of particle-size fractions. Two third of the total data set were used to fitting models. The remaining third was used to model validation.

2.4.3. Uncertainty model validation

The prediction interval coverage probability (PICP) was calculated to validate the quantifications of uncertainty (Malone et al., 2014, 2016). The PICP is the probability that all observed values fit within their estimated prediction interval (PI) (Malone et al., 2011; Solomatine and Shrestha, 2009). The PICP was estimated accordingly for a 90% PI based on the count of observed values that lie within the PI for each site at each particle-size fractions. PIs were constructed for various confidence levels ranging from 5% to 99%, to assess the sensibility of the model by means of reducing the confidence limit sequentially. Ideally, it is expect the PICP values to be close to the corresponding 100(1-a)% confidence level.

2.4.4. Mapping of particle-size fractions and their uncertainties

The maps of the predictions of particle-size fractions were estimated as the product of the spatial modeling. Generally, these predictions have a level of uncertainty that can be quantified, which is important for guiding decision-making processes (Akpa et al., 2016). In this study, the uncertainty of the prediction of particle-size fractions were estimated by calculating 90% prediction intervals from the individual bootstrap predictions for each pixel. The 90% PI for each particle-size fraction were then used to calculate uncertainty of the final prediction. This generated three maps for each particle-size fraction. The first map indicating the lower prediction limit; the second map, the mean prediction and the third map, the upper prediction limit.

3. Results and discussion

3.1. Particle-size fractions

Table 2 shows the comparison of descriptive statistics of the original, calibration and validation data set of particle-size fractions, at the two depths studied. At 0–20 cm soil depth, the range between minimum and maximum clay content, was considerably higher for calibration and validation dataset that to the original dataset, whereas for silt content was similar to all datasets. The range to sand content was considerably lower for validation dataset that to the original and calibration datasets. At 20–40 cm soil depth, the range between minimum and maximum clay content were similar for all datasets, whereas for silt and sand content, were lower to the original and validation dataset than to the calibration dataset. The maximum values for sand content were similar to all datasets, at the two depths. In general, the mean and coefficient of variation (CV) particle-size fractions were similar between depths, suggesting a similar sediment genesis (Amiotti et al., 2001). Sand is the particle-size fraction predominant at both depths. The CV of silt were

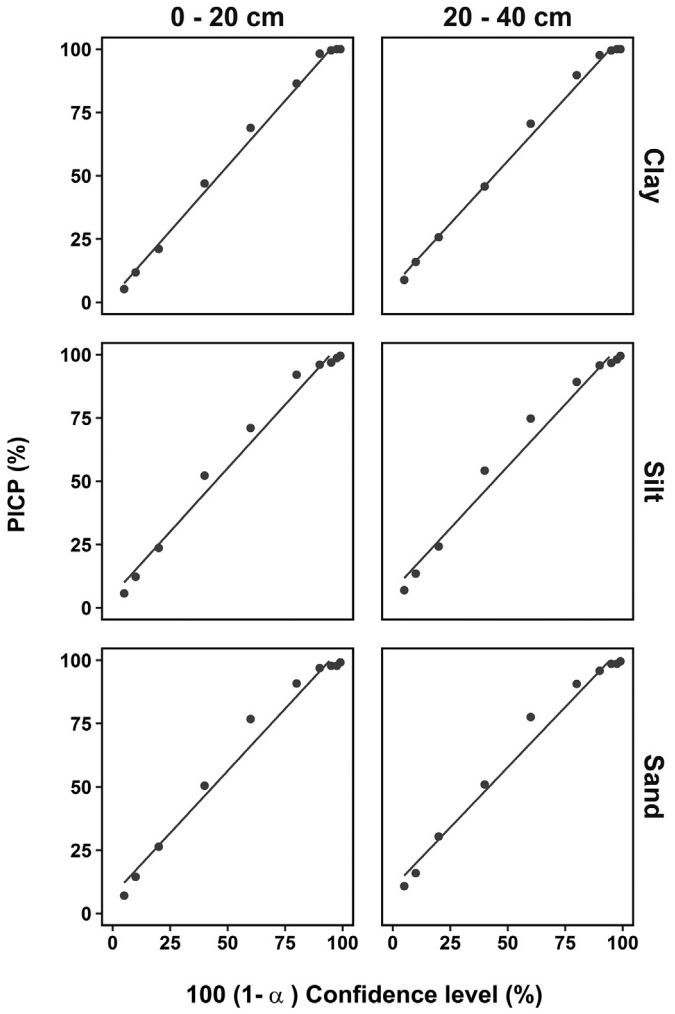

Fig. 3. Prediction interval coverage probability (PICP) for particle-size fractions predictions. For a given α, the PICP is the proportion of predicted values that are inside the 100 (1-α) confidence interval.

higher at both depths, while sand were lower.

Actually, the impact of the intense and long-term volcanic activity from Andes Mountains at different ages, on the particle-size topsoil distribution is subject to discussion. Amiotti et al. (2001) argued that the volcanic materials that were deposited on study area had > 40% of silt, in detriment of the clay and sand factions (Blanco and Stoops, 2007; Buschiazzo et al., 1990). This volcanic materials were deposited overlying a petrocalcic horizon, with a gradual transition from east to west, along with a higher sand content (Gonzalez Uriarte et al., 1990).

3.2. Performance of predictions

The performance of Cubits model in predicting particle-size fractions based on error metrics (Eqs. (1)–(3)) are shown in Table 3. In general, Cubist predictions were slightly better for 0–20 cm than for 20–40 cm soil depth, corroborating the influence of the depth. At both depths, R^2 for clay content to calibration (internal validation) and validation (external validation) dataset was lower, whereas for silt and sand content were higher and similar. In terms of r_c, Cubist performance was better for silt and sand content that to clay content, for calibration and validation datasets. RMSE were lower for particle-size fractions in the calibration dataset and were higher for sand content.

The results indicate that there is a clear deterioration in the Cubist performance to the validation results. However, numerous DSM studies also reported similar validation results for particle-size particles and other soil properties (Bishop et al., 2015; Lacoste et al., 2014; Malone et al., 2009; Viscarra Rossel et al., 2015). In terms of R^2, validation results were between 7 and 18% and 11–24% for particle-size particles at 0–20 cm and 20–40 cm soil depth, respectively, indicated that the predictions were not so good. Alike, in terms of r_c, validation results were between 0.23 and 0.40 and 0.28–0.46% to particle-size fractions at 0–20 cm and 20–40 cm soil depth, respectively, indicated a moderate agreement between the predicted and measured particle-size fraction.

There are three possible reasons for these results. The first is that in the original dataset were collected with ~30 years apart with respect to predictors. Although it is usual in DSM studies to use legacy soil data to map contemporaneous conditions (Bishop et al., 2015; Grunwald, 2009), in the study area, it could have led to biases due to changes in the land use in the last decades and it effect on particle-size fractions in the topsoil. The second is that the predictors, although they are important sources of soil variability, failed to capture a significant proportion of the variability of particle-size fractions at the topsoil. Therefore, it is necessary to explore new predictors that will also explain their soil texture variability in the topsoil. The third is the scale of prediction. In the central and southeastern of the study area, the soils are underlain by an almost continuous caliche, locally termed "*tosca*", which has great variability in depth, and structure (Castro-Franco et al., 2016; Pazos and Mestelan, 2002). Several studies have reported that the variability of the tosca had considerable effect on the spatial pattern of particle-size fractions at field scale (scale > 1:2.000) (Castro-Franco et al., 2015, 2017; Domenech et al., 2017). Therefore, it is possible that this effect reduces the predictive capacity of the Cubist models.

3.3. Predictors importance

Fig. 2 shows a comparison of top importance predictors for particle-size fractions at both depths. In general, these results revealed that climate indices were more important that terrain indices. Specifically, predictors related to temperature (BioClim1, BioClim2, BioClim4, BioClim7 and BioClim9) were the most important to predict particle-size fractions at 0–20 cm soil depth, while the predictors related to temperature (BioClim1, BioClim2, BioClim5 and BioClim9), rainfall (BioClim12) and terrain indices (Valley depth and Wind_exp) were the most important to predict particle-size fractions at 20–40 cm soil depth.

Previous studies have reported different results, where in general, predictors associated with terrain indices have been reported as the most importer to predict particle-size fractions at regional scale (Akpa et al., 2014; Ließ et al., 2012). Is possible that in zones where soil pedogenesis was determined by erosion or deposition, it is likely that terrain features have a crucial role and thus terrain attributes have higher prediction performance for soil texture. However, in zones where soil pedogenesis was determined by aeolian depositions, as in this study area, predictors related to climate have higher prediction performance for soil texture.

3.4. Uncertainly model validation

Fig. 3 shows the comparison of the results of the PICP analysis, which was used to evaluate the efficacy of the 90% prediction intervals. In general, the uncertainties quantified from the bootstrapping approach were acceptably higher (PICP between 78 and 85%) to particle-size fractions at both soil depths. However, value close 90% were expected. Despite this, the results of PICP analysis are comparable to other studies of particle-size fractions predictive mapping. In this respect, Odgers et al. (2015) reported PICP values < 90% to clay content in a DSM study across Western and South Australia. These authors concluded that PICP values varied considerably across depth intervals and

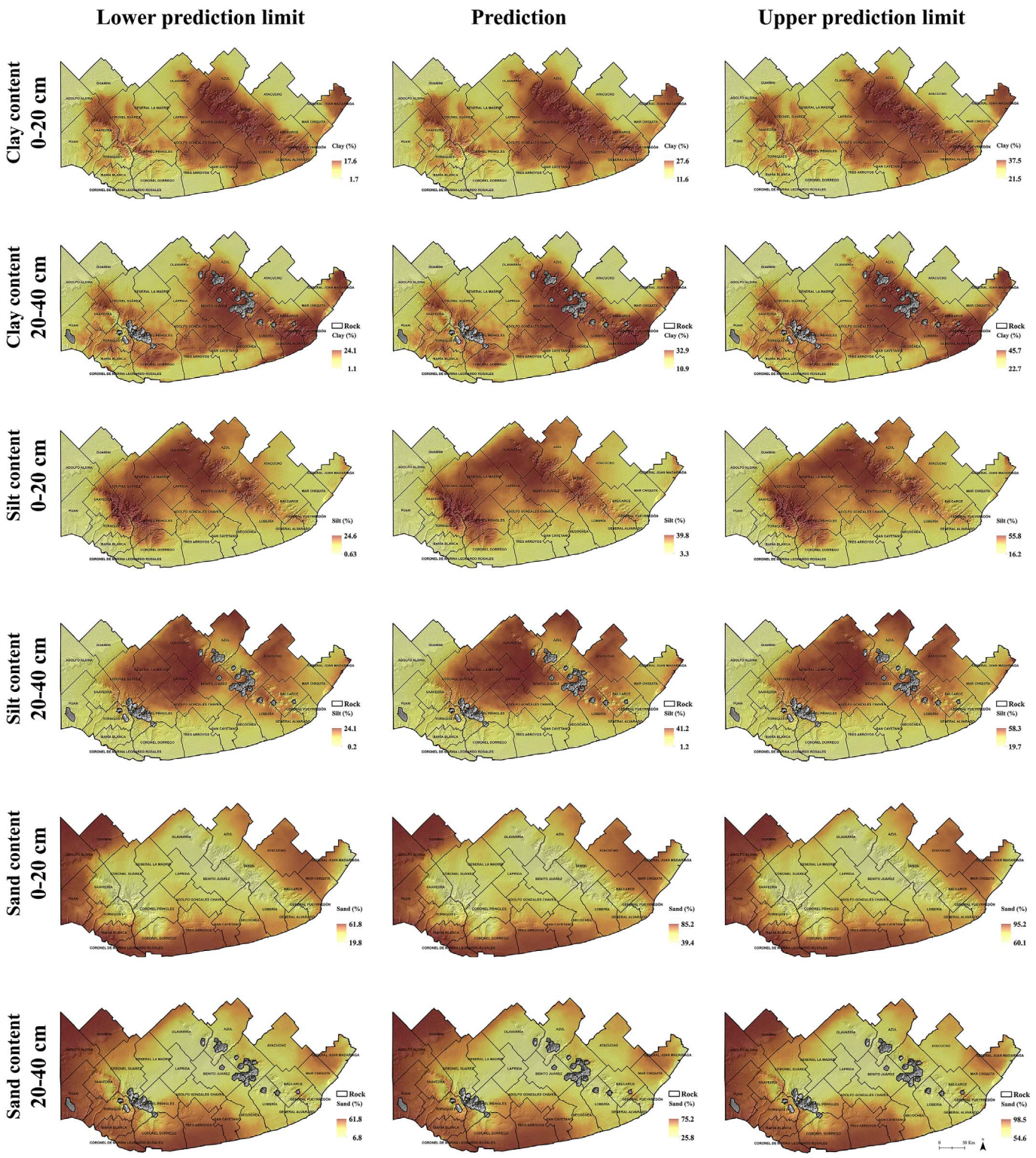

Fig. 4. Maps of particle-size fractions across the southern Argentine Pampas showing, from left to right, the lower, final prediction and upper prediction interval limit.

texture grades. Tranter et al. (2010) also reported PICP values < 90% to clay content in a DSM study across New South Wales, Australia.

3.5. Spatial distribution of particle-size fractions and their uncertainties

Maps of particle-size fractions at both depths for the southern Argentine Pampas are shown in Fig. 3. In general, the spatial distribution of the particle-size fractions was similar at both depths, supporting the vertical homogeneity of the profiles in the whole study area at a depth lower than 40 cm. As expected, variations in sand, in detriment of clay and silt, from the east to west, caused an increase in the sand content further west. Clay and silt contents were higher around

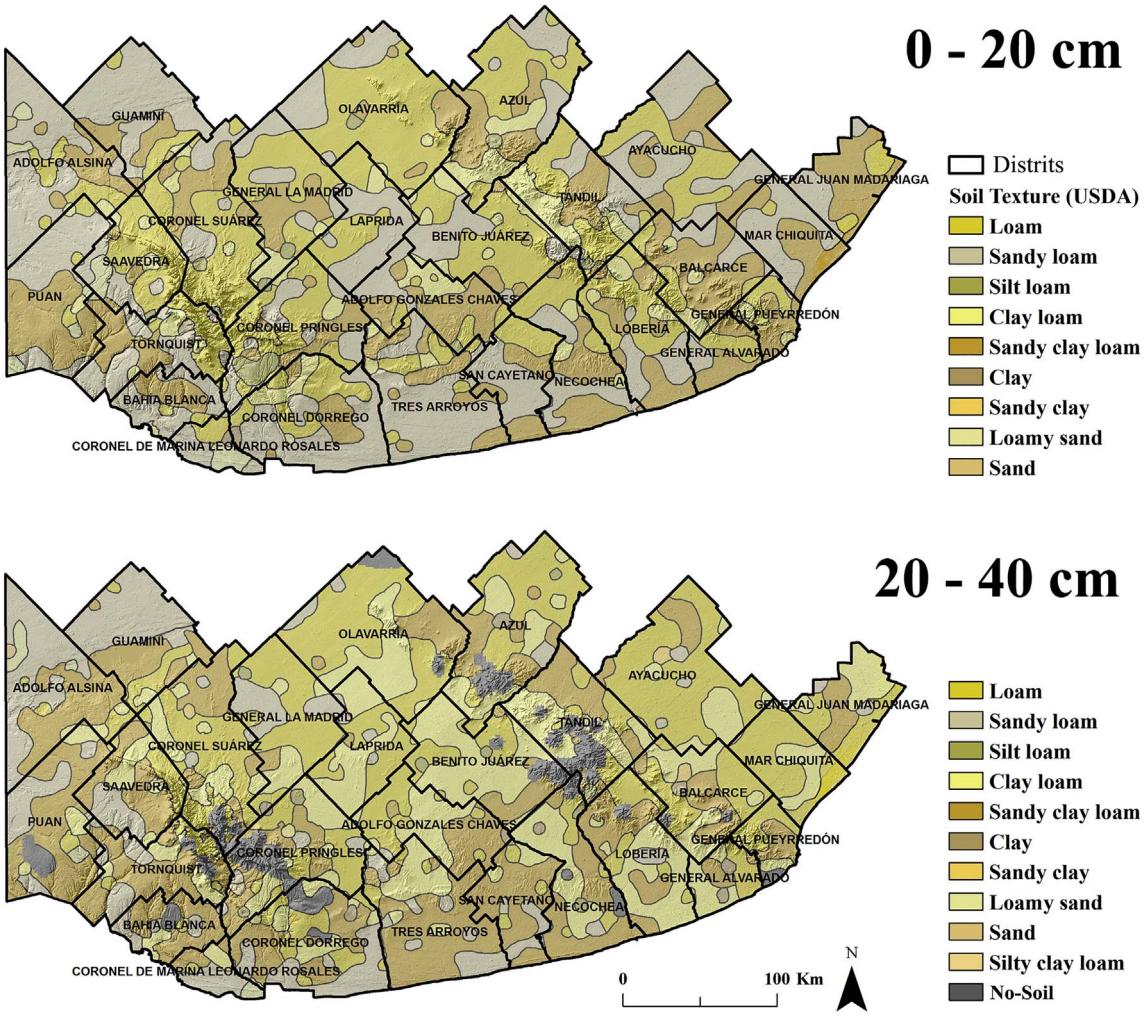

Fig. 5. Maps of textural classes based on the USDA (United States Department of Agriculture) classification across the southern Argentine Pampas.

and in the middle of hilly zones. (See Fig. 4.)

The spatial pattern of the particle-size fractions may have been explained for two geomorphic processes (Amiotti et al., 2001). The first was the deposition, when loess sediments in the Holocene were transported from the west to the east from the Andes Range to the extra-Andean Pampean plain, which contributed to concentrating them in the southeast and proximal to the hills, during several sedimentary pulses (Blanco and Stoops, 2007). Sediments were deposited over the undulated paleotopography of the *tosca* layer. The second was the colluvial–alluvial processes, which contributed to the distribution of the particle-size fractions within the study area, especially in unstable topographic positions located near the hilly systems (Blanco et al., 2003).

3.6. Spatial distribution of soil texture

The spatial distribution of textural classes based on the USDA classification over the study area is shown in Fig. 5. The transition from coarse to fine textures from west to east at both depths is consistent with the soil pedogenesis in the study area. In general, soils at 0–20 cm depth are Sandy loam (~4.1 million ha), Loam (~3.7 million ha), and Sandy clay loam (~3.1 million ha), while those at 20–40 cm are Sandy clay loam (~3.7 million ha), Loam (~3.6 million ha), Clay loam (~2.9 million ha) and Sandy loam (~1.5 million ha). At 20–40 cm depth, basset with *tosca* might be common because of the vertical variability of the petrocalcic horizon at depths lower than 40 cm. Thus, this condition was classified as no-soil. Percentages of the representation of texture classes at both depths are presented in Fig. 6. This is a spatial pattern commonly found in the main Argiudolls of the study area, as reported by Blanco and Stoops (2007) and Cabria and Culot (1994).

The percentage comparison of each texture class in each District studied is shown in Fig. 7. At 0–20 cm depths, Sandy loam soils occupied two zones: the south-center, specifically Tres Arroyos and Coronel Dorrego Districts and the northwest, Guaminí and Adolfo Alsina Districts. Loam soils were mainly in the north of the area, Benito Juarez, Coronel Suarez, Olavarría and Azul Districts. At 20–40 cm depths, Sandy clay loam soils were gathered in two zones: the south-center, Tres Arroyos and San Cayetano, and the west, Puan and Adolfo Alsina Districts. Loam soils were mainly found in the northern Districts, Clay loam soils in Benito Juarez, Necochea and Lobería, and Sandy loam soils in Coronel Suarez. Most of the no-soil zones were found in Bahía Blanca, Coronel Dorrego, Coronel Pringles and Tandil Districts.

4. Conclusions

A DSM technique was used to provide a spatial dataset of soil texture for the southern Argentine Pampas. This includes a map of particle-size fractions at 1-km spatial resolution and a map of the soil texture

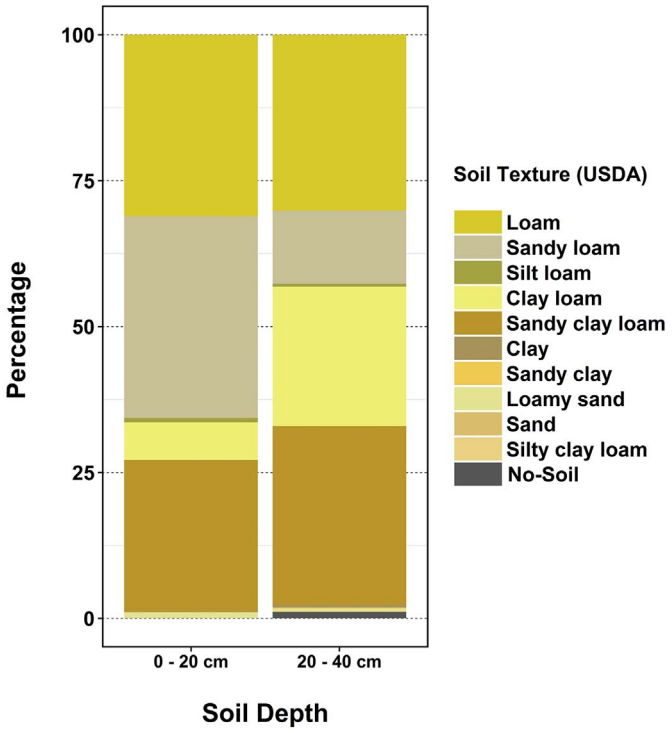

Fig. 6. Percentage of total area of predicted USDA (United States Department of Agriculture) soil textural classes.

based on the USDA classification at 0–20 cm and 20–40 cm depths. It also includes a relationship among textural classes for each District studied.

The new spatial dataset could be useful for climate, hydrological, ecological and crop models. Also, maps of particle-size fractions may be used for soil compaction studies, assessing risks of pesticide pollution or soil erosion models.

The Cubist model was moderately robust to predict particle-size fractions. Although some important properties such as soil effective depth was not included, this model is an acceptable approach of DSM of particle-size fractions for the southern Argentine Pampas. Further works exploring new DSM techniques could possibly improve our results. The salient findings of this study are: (*i*) the main soils in the southern Argentine Pampas are: sandy loam, loam, sandy clay loam and clay loam; (*ii*) at a depth lower than 40 cm, the soil texture has similar spatial patterns; (*iii*) the soil texture varies gradually becoming finer westward; (iv) predictors related to climate were the most important to predict particle-size fractions; and (*v*) Cubist model was able to determine complex spatial relationships between predictors and particle-size fractions in the southern Argentine Pampas conditions.

Conflict of interest

The authors confirm and sign that there is no conflict of interests with networks, organizations and data centers referred in the paper.

Acknowledgements

This study was supported by INTA (PNSUELO-1134023), Argentina.

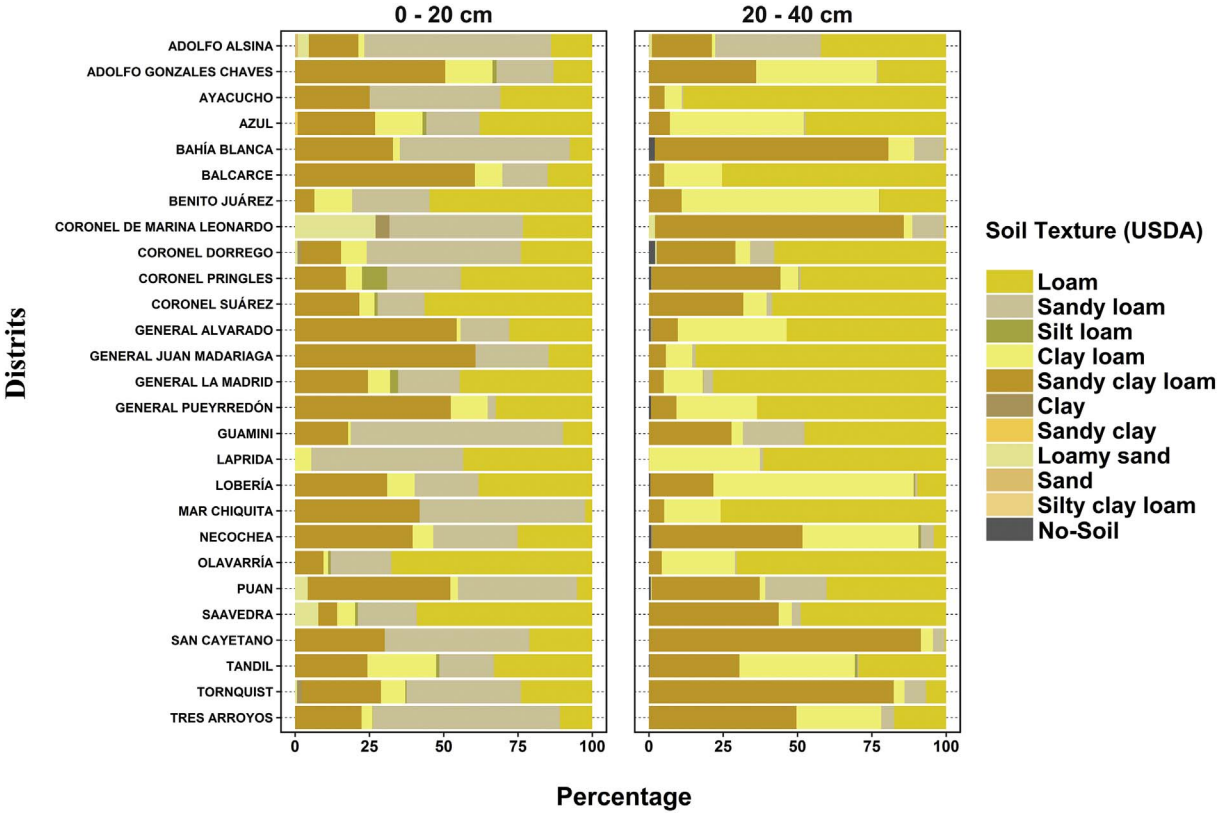

Fig. 7. Percentage of total area by District of predicted USDA soil textural classes.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.geodrs.2017.11.003. These data include Google map of the most important areas described in this article.

References

Adhikari, K., Kheir, R.B., Greve, M.B., Bøcher, P.K., Malone, B.P., Minasny, B., McBratney, A.B., Greve, M.H., 2013. High-resolution 3-D mapping of soil texture in Denmark. Soil Sci Soc Am J 77 (3), 860–876.

Adhikari, K., Kheir, R.B., Greve, M., Greve, M.H., Malone, M., Minasny, B., McBratney, A., 2014. Mapping soil pH and bulk density at multiple soil depths in Denmark. In: GlobalSoilMap: Basis of the Global Spatial Soil Information System. Taylor & Francis, London, pp. 155–160.

Akpa, S.I., Odeh, I.O., Bishop, T.F., Hartemink, A.E., 2014. Digital mapping of soil particle-size fractions for Nigeria. Soil Sci Soc Am J 78 (6), 1953–1966.

Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E., Amapu, I.Y., 2016. Total soil organic carbon and carbon sequestration potential in Nigeria. Geoderma 271 (Supplement C), 202–215.

Amiotti, N., del C. Blanco, M., Sanchez, L.F., 2001. Complex pedogenesis related to differential aeolian sedimentation in microenvironments of the southern part of the semiarid region of Argentina. Catena 43 (2), 137–156.

Angelini, M.E., Heuvelink, G.B., Kempen, B., Morrás, H.J., 2016. Mapping the soils of an Argentine Pampas region using structural equation modelling. Geoderma 281, 102–118.

Arya, L.M., Paris, J.F., 1981. A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. Soil Sci Soc Am J 45 (6), 1023–1030.

Arya, L.M., Leij, F.J., van Genuchten, M.T., Shouse, P.J., 1999. Scaling parameter to predict the soil water characteristic from particle-size distribution data. Soil Sci Soc Am J 63 (3), 510–519.

Berhongaray, G., Alvarez, R., De Paepe, J., Caride, C., Cantet, R., 2013. Land use effects on soil carbon in the Argentine Pampas. Geoderma 192, 97–110.

Bishop, T., Horta, A., Karunaratne, S., 2015. Validation of digital soil maps at different spatial supports. Geoderma 241, 238–249.

Blanco, M.d.C., Stoops, G., 2007. Genesis of pedons with discontinuous argillic horizons in the Holocene loess mantle of the southern Pampean landscape, Argentina. J South Am Earth Sci 23 (1), 30–45.

Blanco, M.d.C., Amiotti, N., Ruiz, J.A., 2003. Reconstrucción de la evolución geopedogenética en una toposecuencia del sudoeste pampeano. Cienc. Suelo 21, 59–70.

Boettinger, J.L., 2010. Digital Soil Mapping: Bridging Research, Production, and Environmental Application. Springer Science & Business Media.

Boettinger, J., Ramsey, R., Bodily, J., Cole, N., Kienast-Brown, S., Nield, S., Saunders, A., Stum, A., 2008. Landsat Spectral Data for Digital Soil Mapping, Digital Soil Mapping with Limited Data. Springer, pp. 193–202.

Buchanan, S., Triantafilis, J., Odeh, I., Subansinghe, R., 2012. Digital soil mapping of compositional particle-size fractions using proximal and remotely sensed ancillary data. Geophysics 77 (4), WB201–WB211.

Buschiazzo, D.E., Hevia, G., Hepper, E., 1990. Variaciones de la adsorción de fósforo y parámetros edáficos en una climosecuencia de los suelos de la región semiárida pampeana central. Cienc. Suelo 8, 119–126.

Cabria, F., Culot, J.P., 1994. Selección y utilización de características edáficas para discriminar series de Argiudoles en el sudeste bonaerense. Cienc. Suelo 12, 41–45.

Cambule, A., Rossiter, D., Stoorvogel, J., 2013. A methodology for digital soil mapping in poorly-accessible areas. Geoderma 192, 341–353.

Castro Franco, M., Costa, J.L., Peralta, N., Aparicio, V., 2015. Prediction of soil properties at farm scale using a model-based soil sampling scheme and random forest. Soil Sci 180, 1–12.

Castro-Franco, M., Domenech, M., Costa, J.L., Aparicio, V.C., 2016. Modelling effective soil depth at field scale from soil sensors and geomorphometric indices. Acta Agron 66 (2).

Castro-Franco, M., Díaz, H.J., Londoño Quiroz, M., Cicore, P., Costa, J.L., 2017. Predicción del contenido de arcilla superficial mediante conductividad eléctrica aparente y esquema de muestreo basados en modelos. Cienc Suelo 35 (1), 1–18.

Cook, S., Jarvis, A., Gonzalez, J., 2008. A New Global Demand for Digital Soil Information, Digital Soil Mapping with Limited Data. Springer, pp. 31–41.

Domenech, M.B., Castro-Franco, M., Costa, J.L., Amiotti, N.M., 2017. Sampling scheme optimization to map soil depth to petrocalcic horizon at field scale. Geoderma 290, 75–82.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Taylor & Francis.

Gonzalez Uriarte, M., Navarro, M., Aldacour, H., 1990. Formaciones loessicas superficiales en el sur bonaerense (Argentina). In: International Symposium on Loess. INQUA, Mar del Plata, Argentina, pp. 55–57.

Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152 (3–4), 195–207.

Hartemink, A.E., McBratney, A.B., de Lourdes Mendonça-Santos, M., 2008. Digital Soil Mapping with Limited Data. Springer Science + Business Media B.V.

Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124 (3), 383–398.

Hijmans, R., Cameron, S., Parra, J., Jones, P., Jarvis, A., Richardson, K., 2005a. WorldClim, Version 1.3. University of California, Berkeley.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005b. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25 (15), 1965–1978.

INTA, 2013a. Sistema de geoinformación para consulta y procesamiento de datos georreferenciados de Argentina GeoINTA. www.geointa.inta.gov.ar ((Ed.), Buenos Aires Argentina).

INTA, 2013b. Sistema de geoinformación para la consulta y procesamiento de datos gerreferenciados de Argentina GEOINTA. http://geointa.inta.gov.ar/visor/ ((Ed.), Buenos Aires Argentina).

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311.

Ließ, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture: comparison of regression tree and random forest models. Geoderma 170, 70–79.

MAA, 1965. Atlas de condiciones de los suelos. Estudio de desarrollo agropecuario de la Provincia de Buenos Aires. Ministerio de Asuntos Agrarios de la Provincia de Buenos Aires, Buenos Aires, Argentina.

Malone, B.P., McBratney, A., Minasny, B., Laslett, G., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154 (1), 138–152.

Malone, B., McBratney, A., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160 (3), 614–626.

Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232, 34–44.

Malone, B.P., Minasny, B., McBratney, A.B., 2016. Using R for Digital Soil Mapping. Springer International Publishing.

Matteucci, S., 2012. Ecorregión Pampa. In: Morello, J., Matteucci, S., Rodríguez, A., Silva, M. (Eds.), Ecorregiones y complejos ecosistémicos argentinos. Ed. Orientación Gráfia, Buenos Aires, pp. 391–446.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.

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.

Odeh, I.O., Todd, A.J., Triantafilis, J., 2003. Spatial prediction of soil particle-size fractions as compositional data. Soil Sci 168 (7), 501–515.

Odgers, N.P., Holmes, K.W., Griffin, T., Liddicoat, C., 2015. Derivation of soil-attribute estimations from legacy soil maps. Soil Res 53 (8), 881–894.

Pazos, M., 1984. Relación arcilla Iluvial/arcilla total en Molisoles del sudeste de la provincia de Buenos Aires. Cienc Suelo 2, 131–136.

Pazos, M.S., Mestelan, S.A., 2002. Variability of Depth to Tosca in Udolls and Soil Classification, Buenos Aires Province, Argentina. Soil Sci Soc Am J 66 (4), 1256–1264.

QGIS v2.16.1 Nodebo, 2016. Development Team. Quantum GIS geographic information system, Essen, Germany.

Quinlan, J.R., 1992. Learning with Continuous Classes. In: 5th Australian Joint Conference on Artificial Intelligence. Singapore, pp. 343–348.

SAGA Development Team, 2016. System for Automated Geoscientific Analysis (SAGA).

Saxton, K., Rawls, W.J., Romberger, J., Papendick, R., 1986. Estimating generalized soil-water characteristics from texture. Soil Sci Soc Am J 50 (4), 1031–1036.

Shangguan, W., Dai, Y., Liu, B., Ye, A., Yuan, H., 2012. A soil particle-size distribution dataset for regional land and climate modelling in China. Geoderma 171, 85–91.

Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed. United States Department of Agriculture - Natural Resources Conservation Service, Washington, DC.

Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resour Res 45 (12).

Tranter, G., Minasny, B., McBratney, A., 2010. Estimating pedotransfer function prediction limits using fuzzy-means with extragrades. Soil Sci Soc Am J 74 (6), 1967–1975.

Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Reg 4, 20–30.

Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res 53 (8), 845–864.

Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. Int J Geogr Inf Sci 20 (2), 193–213.

Wang, Z., Shi, W., 2017. Mapping soil particle-size fractions: a comparison of compositional kriging and log-ratio kriging. J Hydrol 546, 526–541.