Geoderma 322 (2018) 101–111

Contents lists available at ScienceDirect

Geoderma

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

A pedometric technique to delimitate soil-specific zones at field scale

Mauricio Castro-Franco[a,b,*], Mariano A. Córdoba[a,c], Mónica G. Balzarini[a,c], José Luis Costa[d]

[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] Catedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ing Agr. Felix Aldo Marrone 746, Ciudad Universitaria, C.C. 509, 5000 Cordoba, Argentina
[d] Instituto Nacional de Tecnología Agropecuaria INTA, EEA Balcarce, Ruta 226 km 73.5, Balcarce, Argentina

A R T I C L E I N F O

Handling Editor: A.B. McBratney

Keywords:
Argentina
Digital soil mapping
Random forest
Spatial principal components
Precision agriculture

A B S T R A C T

Delimitation of soil types within a farm field is key for site-specific crop management. An alternative to this, is to develop pedometric techniques that allow an efficient combination of soil survey information and high-resolution terrain attribute data. The aim of this study was to present and evaluate a pedometric technique to delimit soil-specific zones at field scale by coupled Random forest, fuzzy k-means clustering and spatial principal components algorithms (RF-KM-sPCA) and by using information from soil surveys and terrain attributes derived from a digital elevation model. The protocol involves three-steps: 1) automatic classification of small (20x20m) spatial units (SU) using the knowledge of the soil map units present in the farm landscape, 2) aggregation of SUM at farm scale and 3) validation of soil-specific zones. For the first step, we used the random forest algorithm with 10 terrain attributes. For the second step, KM-sPCA algorithms were used to cluster within field SU accounting for autocorrelation. For the third step, apparent soil electrical conductivity and yield maps was used to validate the delimitation of soil-specific zones. This technique produced more contiguous zones than other cluster methods which do not use spatiality. Six farm fields with highly differences in soils were partitioned by the proposed pedometric strategy. Apparent soil electrical conductivity and yield maps present significant differences among zones in all experimental fields. This analytic strategy, based in easy-to-obtain data, could be used to improve precision agricultural managements.

1. Introduction

Soil properties that limit crop yield within agricultural fields often vary considerably over space and time (Castro Franco et al., 2015; Gebbers and Adamchuk, 2010). Usually, this variability is intentionally ignored in soil sampling schemes, laboratory analyses and agronomic strategies for crop management. Hence, it appears that applying strategies for soil-specific conditions in the context of precision agriculture would have the potential to improve the way in which soils are currently managed. To achieve this, the method used to delimit the complexity of soils of agricultural fields in subareas according to the soil type should be simplified, so that these subareas could be individually controlled with respect to management decisions (Fraisse et al., 2001; Johnson et al., 2001).

Pedometrics techniques is the application of mathematical and statistical models to study the distribution and genesis of soils (Rossiter, 2012), which within the context of digital soil mapping (McBratney et al., 2003), could be useful to define soil-specific zones in agricultural fields. Generally, there are three pedometric approaches. The first one, known as disaggregation of soil map units (DgSMU), allows delimiting soil-specific zones at different scales by combining information obtained from conventional soil surveys with information obtained from digital soil mapping (Bui and Moran, 2001). The second one estimates the spatial distribution of soil properties by using geostatistical methods (Hempel et al., 2008). The main disadvantage of this approach, is that a huge number of soil samples have to be collected and analyzed to adequately represent the soil spatial variability (Fraisse et al., 2001). The third approach estimates soil spatial patterns through machine learning algorithms by using ancillary data such as apparent soil electrical conductivity, remote sensing, and digital elevation models (DEM) (Ahmad et al., 2010; Castro Franco et al., 2015; Nitze et al., 2012; Scull et al., 2003).

In South America, the first and third approaches have had a great potential to generate useful cartography to be implemented in soil-specific management strategies. This is because in this region several soil surveys are available, which can be disaggregated using digital soil mapping tecniques (Pennock et al., 2015; Sanchez et al., 2009); Spatially, this conventional survey is formed by polygons or Soil Map Units

* 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.geoderma.2018.02.034
Received 20 February 2017; Received in revised form 30 January 2018; Accepted 21 February 2018
Available online 24 February 2018
0016-7061/

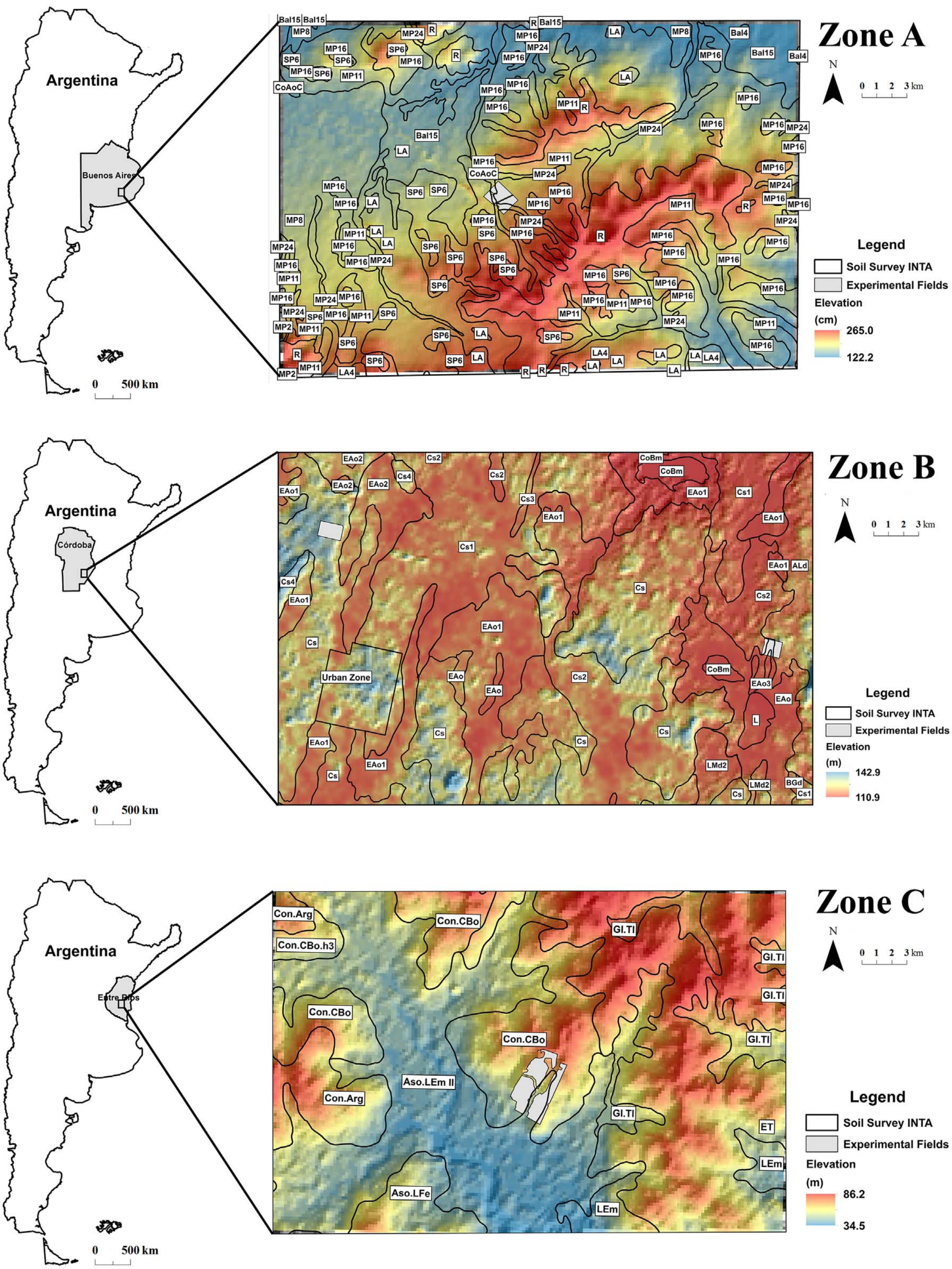

Fig. 1. Mapping of Soil Map-units (SMUs) according to INTA soil survey at scale 1:50,000 for each agricultural zone, Argentina. Spatial distribution of elevation from digital elevation model (MDE-Ar) (back)

(SMU) according to their soil-landscape relationships (Jenny, 1941; McBratney et al., 2003). Each SMU represents the "*aggregation*" of a number of soil series which are identified by their spatial correspondence; hence, each SMU is considered as a spatial generalization that can be disaggregated (Nauman and Thompson, 2014). Also, multiple ancillary information is available, which can be used to classify SMUs from machine learning algorithms (Brungard et al., 2015; Heung et al., 2016; Massawe et al., n.d.). Generally, these algorithms are used to determine the spatial correlation among SMUs and ancillary information of environmental data derived from DEM, remote sensing, and soil sensing, in order to develop a training dataset (McBratney et al., 2003). The learning relationships between SMU and environmental data are adjusted in a model which is then applied in the validation procedure. Within machine learning, Random Forest (RF) is an outstanding algorithm (Gambill et al., 2016). The RF technique is an ensemble learning technique which generates many classification trees that are aggregated

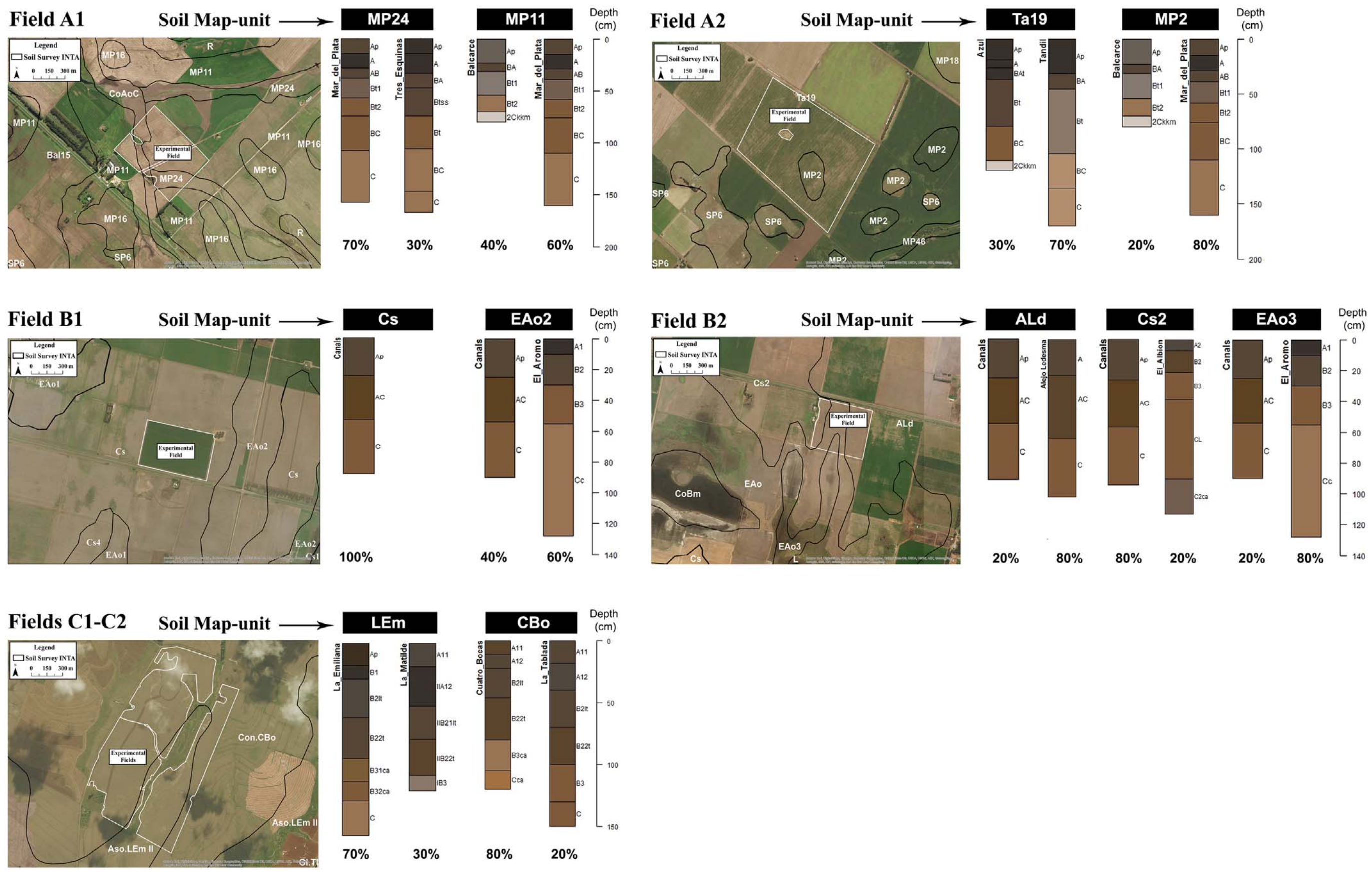

Fig. 2. Spatial distribution of Soil Map-unit (SMU) nearby six experimental fields (left), soil series description plot and percentage within SMU (right).

to compute a classification (Breiman, 2001; Brungard et al., 2015). RF usage has had a remarkable growth in pedometrics due to its efficiency to predict soil properties and classify SMU (Heung et al., 2016).

The couple of fuzzy k-means clustering and spatial principal component analysis (KM-sPCA) consists of a spatial multivariate clustering algorithm which has been demonstrated to be efficient to delineate soil-specific zones at field scale using ancillary information derived from precision agriculture technologies (Córdoba et al., 2013). KM-sPCA is based on principal component analysis but incorporates a constraint due to spatial correlation or dependence among variables (Córdoba et al., 2016). To our knowledge, no studies have evaluated coupled RF-KM-sPCA as a pedometric technique for DgSMU to define soil-specific zones at field scale.

The aim of this study was to present and evaluate a pedometric technique for DgSMU to delimit soil-specific zones at field scale by coupled RF-KM-sPCA and by using information from soil surveys and terrain attributes derived from DEM. RF was used as an automatic classification algorithm for SMU. Once the classification was computed, KM-sPCA was used as a spatial clustering algorithm within each SMU and soil-specific zones were defined. This strategy was tested in farm fields of three subregions in the Argentine Pampas region. Soil-specific zones were validated using information obtained by apparent electrical conductivity and yield maps. The application of these algorithms will offer a novel approach to optimize the automatic delimitation of soil-specific zones at field scale that could be worth for implementation of precision agriculture management and hydrological models.

2. Materials and methods

2.1. Study area

The study area covered three agricultural zones of the Pampean region, Argentina, from 32S – 37S to 58W – 62W (Fig. 1), with roughly different pedological conditions: the southeast of Buenos Aires province (Zone A), the southeast of Córdoba province (Zone B) and the southeast of Entre Ríos province (Zone C). Zone A is characterized by the presence of Petrocalcic Argiudolls with limited soil depth, Zone B is characterized by sandy soils, and Zone C is characterized by Vertic Argiudolls.

Two agricultural fields were selected in each zone. In Zone A, fields A1 and A2 had 42 and 114 ha respectively and were located in Balcarce (−37.6080 S, −58.6340 W; Datum WGS84) and Tandil (−37.2070 S, −59.4190 W), respectively. In Zone B, fields B1 and B2 had 42 and 44 ha, respectively, and were located in Canals (field B1: −33.5223 S, −62.8955 W; field B2: −33.5558 S, −62.7623 W). In Zone C, the fields selected were adjacent (−32.5144 S, −59.4717 W) with 84.4 and 44.9 ha, respectively.

2.2. Conventional soil survey

In these zones, a conventional soil survey at a scale of 1:50,000 was developed by the project called “Soil Map Project of the Pampean Region” (Plan Mapa de Suelos de la Región Pampeana), and was carried out by the National Institute of Agricultural Technology (INTA) in the sixties and early nineties (Moscatelli and Pazos, 2000), adopting the US Department of Agriculture (USDA) Soil Taxonomy as the soil classification system (Soil Survey Staff, 2014) (Figs. 1 and 2).

According to that, soils in field A1 correspond to SMUs called M11 and M16, placed in capability classes II_e and II_es, and productivity indices 79.4_B and 70.4_B, respectively. The SMU M11 is a consociation of Mar del Plata and Balcarce series. Mar del Plata series is a deep Typic Argiudoll soil profile, while Balcarce is a Petrocalcic Argiudoll soil (Fig. 2). A2 soils belong to SMUs Ta19 and MP2, placed in capability classes III_es and II_es, and productivity indices 67_B and 77.2_B, respectively. Ta19 is a consociation of Tandil and Azul series. MP2 is a

Table 1
Area and fluctuation of true/false predictions for SMU in each experimental field.

Field	Soil map-unit	Area represented (%)	Spatial units	Prediction (%)		Field	Soil map-unit	Area represented (%)	Spatial units	Prediction (%)	
				True	False					True	False
A1	MP8	16.86	32,742	97.30	2.70	A2	MP16	0.57	1178	92.02	7.98
	MP11	47.37	92,020	94.70	5.30		SP6	3.73	7689	87.92	12.08
	CoAoC	1.74	3371	86.68	13.32		MP18	20.41	42,046	97.85	2.15
	Bal15	7.26	14,112	96.87	3.13		Tdf1	0.20	408	82.35	17.65
	SP6	1.81	3522	89.98	10.02		MP44	5.00	10,297	93.07	6.93
	MP2	0.06	109	77.06	22.94		Eg	1.29	2649	75.80	24.20
	LA	0.92	1791	87.27	12.73		MP45	1.71	3527	82.76	17.24
	LA4	2.55	4955	95.88	4.12		Bal29	1.39	2853	84.30	15.70
	MP24	7.22	14,025	87.24	12.76		Az46	8.09	16,660	97.88	2.12
	MP16	0.82	1597	5.26	94.74		CoAoC	0.52	1061	75.02	24.98
	R	13.31	25,856	94.81	5.19		Ta19	51.98	107,066	98.00	2.00
	Bal4	0.08	153	91.50	8.50		CC9	0.02	50	58.00	42.00
							MP46	2.60	5356	79.48	20.52
							MP2	2.49	5136	79.77	20.23
B1	Cs2	8.15	20,433	96.77	3.23	B2	Cs2	12.88	32,267	97.31	2.69
	EAo2	5.91	14,811	96.03	3.97		Eao	7.56	18,940	95.55	4.45
	Cs4	1.95	4882	93.42	6.58		Cs4	0.62	1542	94.62	5.38
	EAo2	1.98	4966	94.76	5.24		CoBm	0.66	1660	95.24	4.76
	Cs3	0.99	2481	93.99	6.01		EAo3	0.22	546	86.08	13.92
	EAo3	0.11	281	91.81	8.19		L	0.72	1793	96.77	3.23
	Cs	45.71	114,626	98.29	1.71		Cs1	9.43	23,620	97.81	2.19
	EAo1	14.20	35,608	96.04	3.96		LMd2	3.09	7748	96.59	3.41
	Cs1	21.00	52,653	98.06	1.94		BGd	0.12	302	93.38	6.62
							Cs	43.83	109,761	98.19	1.81
							Cs3	0.12	294	94.90	5.10
							Ald	2.36	5912	97.53	2.47
							EAo2	1.22	3064	95.27	4.73
							EAo1	17.17	43,002	96.89	3.11
C1-C2	Lfe	3.29	5982	98.38	1.62						
	ET	30.29	54,995	98.28	1.72						
	LEm_II	34.61	62,846	98.58	1.42						
	GI.TI	6.44	11,700	93.26	6.74						
	Cbo	15.92	28,913	98.09	1.91						
	Arg	5.54	10,058	97.27	2.73						
	LEm	2.60	4713	96.24	3.76						
	Cbo.h3	1.30	2362	94.83	5.17						

consociation of Mar del Plata (80%) and Balcarce (20%) series. Both fields are in a hillslope landscape, developed on well-drained, non-saline, and non-alkaline loessic silt loam sediments, and slopes with gradients from 1 to 3% (INTA, 2010).

Field B1 soils correspond to SMUs called Cs and EAo2. The first is completely formed by Canals series (100%), while the second is a consociation of El Aromo and Canals series. B2 soils correspond to SMUs ALd, Cs2 and EAo3, all of which are consociations. The first one is between Alejo Ledesma and Canals series, the second one is between Canals and El Albión series, and the third one is between Canals and El Aromo series (Fig. 2).

Finally, C1 and C2 soils belong to the SMUs called LEm_II and CBo. The first SMU is a consociation of La Emiliana and La Matilde series, which are related to rolling gentle plain surface with a thick loess layer. The second one is a consociation of Cuatro Bocas and La Tablada series, which are located in rolling peniplains with a thin loess layer. Cuatro Bocas and La Tablada soil series are classified as Vertic Argiudolls, while La Emiliana is classified as Typic Argiudoll (Fig. 2).

2.3. Terrain attributes

A 30-m spatial resolution DEM was used to calculate terrain attributes. This DEM is a revised version of DEM SRTM 30 m, developed at the Instituto Geográfico Nacional of Argentina (MDE-Ar), available on line at *www.ign.gob.ar*. In this DEM, (*i*) the water bodies were identificated, delineated and determinated; (*ii*) the voids were filled; (*iii*) the pixels values were spatially filtered with respect to their neighboring values; and (*iv*) the limits were masked (IGN, 2016).

The terrain attributes calculated from this DEM were: relative slope position, aspect, curvature, flow accumulated, catchment slope, catchment area, modified catchment area, topographic wetness indices, convergence index, channel network, valley depth, vertical distance and topographic position index. All of them were calculated using SAGA GIS v2.3.1 (SAGA Development Team, 2016).

2.4. Procedure to delimit soil-specific zones

In each study zone, the different SMUs were gathered in an area of 600 ha approximately around each experimental field (Fig. 1). All terrain attributes were delimited according to these areas and a 20-m regular grid was built. Then, the projected coordinate system POSGAR-07 Datum WGS84 and the values for each point were added to the grid attributes table. This procedure was carried out using QGIS v2.16.2 (QGIS v2.16.1 Nodebo, 2016). This grid was used as input for the DgSUM process, which was done following three steps:

2.4.1. Step 1: SMU classification using RF

A classification was performed with RF based on a 20-m regular grid, which had the information of all terrain attributes, and the SMU to which it spatially belongs (Breiman, 2001). The propose of this classification was to imitate the point of view of soil surveyors, during the delimitation of each SMUs in the soil survey and to define a topographic fingerprint for each one (Bui and Moran, 2001). RF used a specific number of decision trees, n_{tree}, by random selection from the 20-m regular grid. At each node, a bootstrap sample of predictors was taken (*mtry* argument). Two thirds of the bootstrap sample was used, by

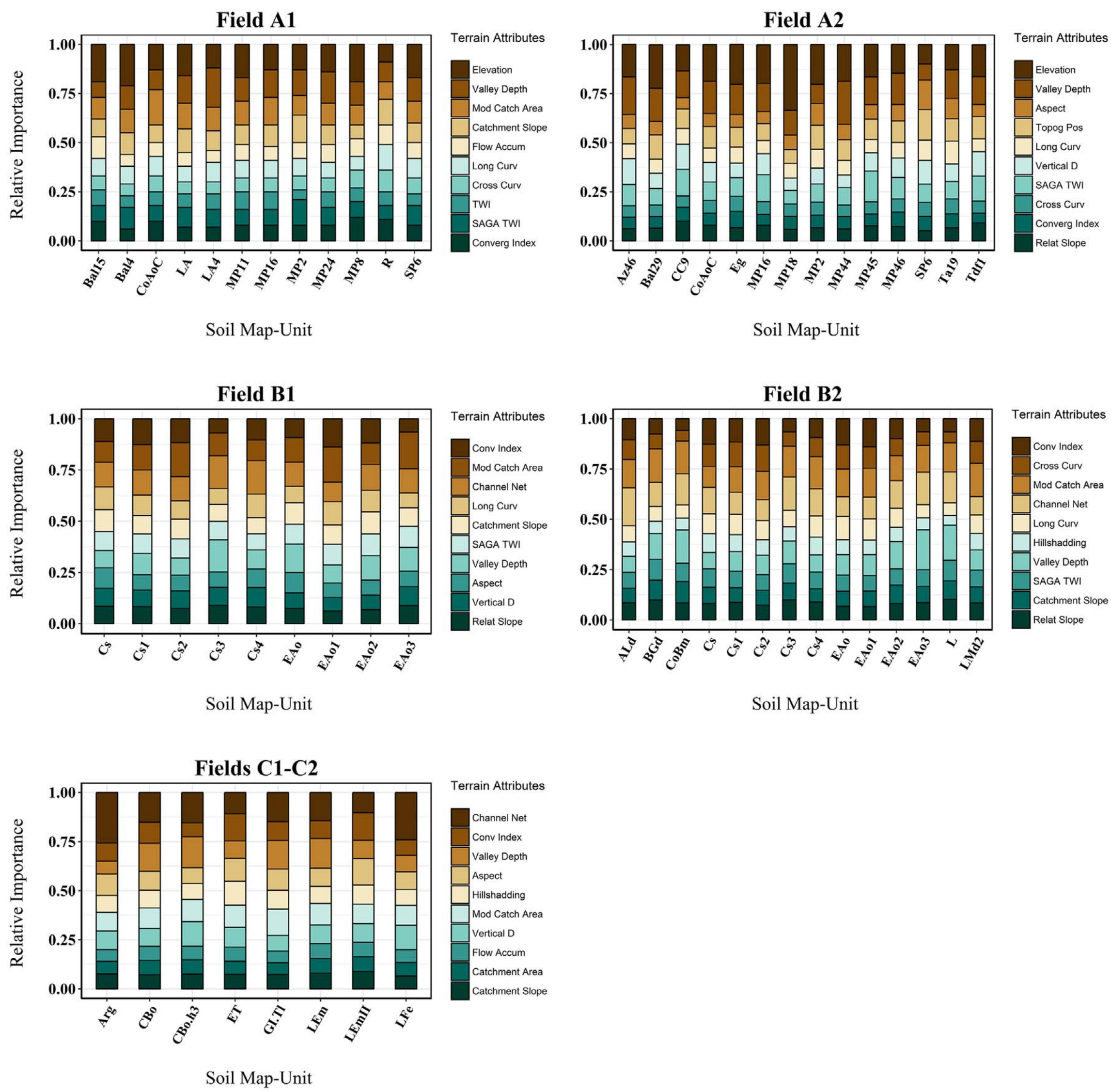

Fig. 3. Relative importance of terrain attributes in the classification of each Soil Map-unit (SMU) in each field, based on RF importance classification.

default, to build classification trees. The rest bootstrap (*out-of-bag)* was used to estimate the error rate called *out-of-bag* error (*OOB-error*), which was calculated to predict the SMU class of the remaining data and their comparison with respect to known SMU class. RF was run using randomForest R package (Liaw and Wiener, 2002; R Core Team, 2016).

RF allowed to detect the dependences between terrain attributes and SMU class, and to select the relevant terrain attributes and also calculate variable importance measure (Domenech et al., 2017; Vogels et al., 2017). Mean-Decrease-in-Accuracy was used as a measure of variable importance, which is defined as the loss of accuracy measured by the *OOB-error* when leaving out a variable (Breiman, 2001). The global importance for each terrain attribute were compared and analyzed for all SMUs, by field. Finally, to compare the importance of terrain attributes between SMUs, a metric of relative importance was calculated as the ratio between importance of each terrain attribute and the highest importance at each SMU.

2.4.2. Step 2: spatial multivariate clustering within each SMU

The couple KM-sPCA algoritms was implemented as a methodology to define the topographic fingerprint in each soil series associated with different SMUs (Córdoba et al., 2013; Córdoba et al., 2016). The sources of information were terrain attributes. sPCA is an extension of principal component analysis, which incorporates spatial dependence between original variables before they are computed in principal components. The constraint imposed by spatial data is incorporated by Moran's index (Córdoba et al., 2016). The goals of clustering analysis by sPCA are (*i*) to include many sources of information derived from rising technologies such as terrain attributes, because this methodology is focused on spatial multivariate analysis; and (*ii*) to specify an optimal number of zones defined by performance indices. KM clustering is a classical non-hierarchical clustering algorithm that can be implemented by using statistical software. This algorithm has three primary matrices involved. The first contains the data matrix $\boldsymbol{X}$ involves n observations with a $< p$ sPCA each; the second matrix is the cluster centroid matrix $\boldsymbol{V}$, consisting of k cluster centroids located in the attribute space defined by the retained spatial principal components. The third is the fuzzy membership matrix U, consisting of membership values to every cluster in V for each sample point in X, bounded by the constraint that the sum of membership values for each observation should be equal to 1. An optimal fuzzy k partition is defined as a minimization of the weighted measure of squared distance between data points as class centroids:

$$j_m(U, v) = \sum_{j=1}^{n} \sum_{i=1}^{k} (\boldsymbol{u}_{ij})^m (\boldsymbol{d}_{ij})^2$$

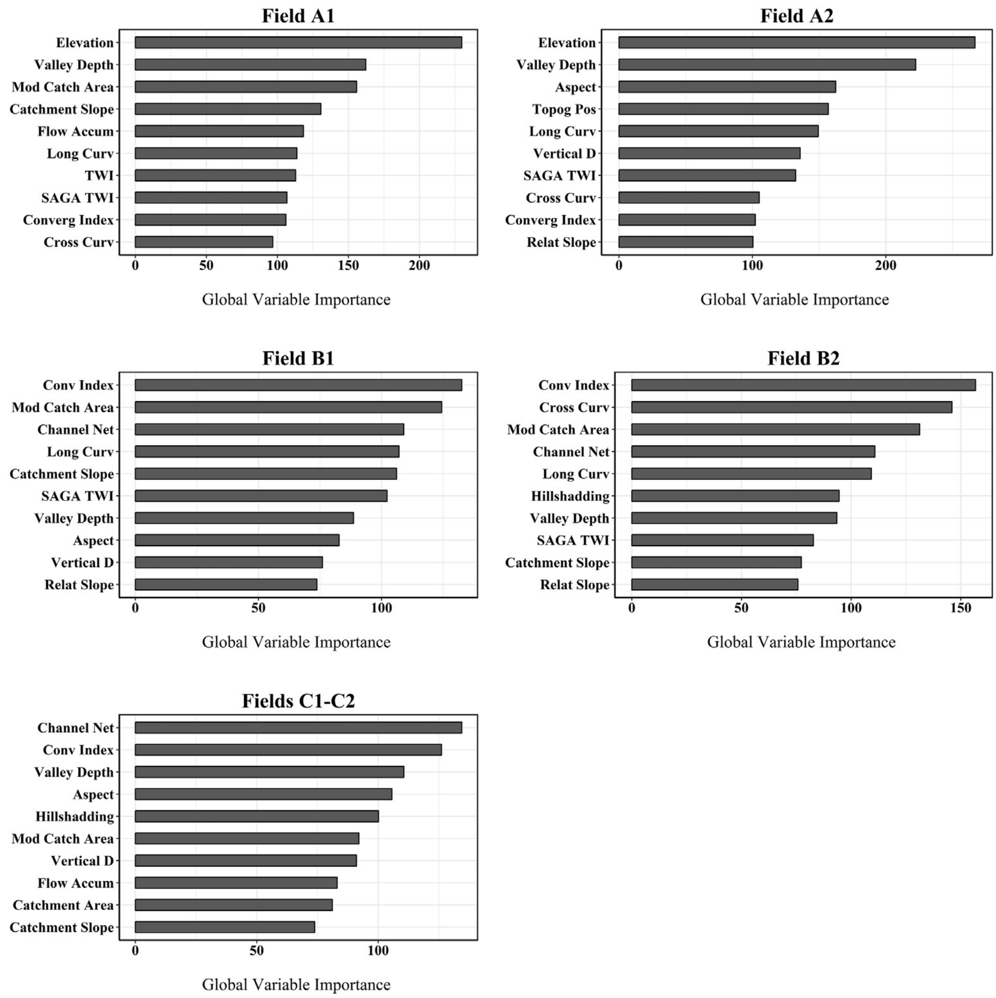

Fig. 4. Relative importance of terrain attributes in each Soil Map-unit (SMU) classification for experimental field (left) and global variable importance (right) based on Random Forest algorithm results.

where m is the fuzziness weighting coefficient ($1 \leq m < \infty$) and $(\boldsymbol{d}_{ij})^2$ is the squared distance in the attribute space between point j and class centroid i. The fully description of the couple fuzzy k-means clustering and sPCA algoritms used in this work, can be consulted in Córdoba et al. (2013).

The performance indices of the couple KM-sPCA were coefficient of partition (fuzziness performance index (FPI)), classification entropy (normalized classification entropy (NCE)) (Odeh et al., 1992), Xie-Beni index (Pal and Bezdek, 1995), Fukuyama index (Fukuyama and Sugeno, 1989) and exponent of proportion (Windham, 1981). These indices were combined in a summary index, which was calculated using the Euclidian distance of all indices values, previously normalized by their maximum values obtained through different classification methods (Córdoba et al., 2013). The sPCA procedure was run using the InfoStat v2015p Spatial Statistical module (Di Rienzo et al., 2015).

2.4.3. Step 3: validation of soil-specific zones

To validate soil-specific zone delimitation, apparent soil electrical conductivity (ECa) measured by sensor Veris 3100® (VERIS Technologies, Salina, KS, USA) and soybean yield maps for 2010 and 2011 were used. Both are tools of precision agriculture which have been implemented to define site-specific management zones in farm fields (Bobryk et al., 2016). The operational characteristics and protocol measurement for Veris 3100® are described in Corwin and Lesch (2005). The procedure for correction and removal of erroneous yield monitor data was carried out following the protocol proposed by Blackmore and Moore (1999).

Considering that ECa and yield may not be independent variables (i.e. spatial correlation), a random sampling scheme without replacement was carried out in each field. To do this, 100 samples whose size represented 15% of data were chosen. For each sample, an ANOVA was adjusted to investigate the effect of soil-specific zones on ECa and yield. The model applied was:

$$y_{ij} = \mu + S_i + Z(S)_{j(i)} + e_{ij}$$

where y_{ij} represents the value of ECa or yield within SMU i, in the soil-specific zone j; μ represents the general mean; S_i is the effect of SMU; and $Z(S)_{j(i)}$ is the effect of the soil-specific zone within each SMU. This procedure was carried out using InfoStat v2015p (Di Rienzo et al., 2015). The Fischer's LSD test was used to determine differences between means of ECa and yield of soil-specific zones. Differences were considered significant at $P < 0.05$. Finally, a visual analysis of the differences between soil-specific zones in all experimental fields, was carried out by a graphical method.

3. Results and discussion

3.1. SMU classification based on RF

In Zone A, the classification using RF was estimated for 14 SMUs of A1 and for 12 SMUs of A2. In Zone B, it was estimated for 14 SMUs for both fields. In Zone C, it was estimated for 8 SMUs. RF was computed

Table 2
Optimum number of zones within each experimental field based on summary index. Bold letters means lowest summary index value.

Field	Soil map-unit	Number of zones	Summary index*
A1	CoAoC	2	**2.00**
	CoAoC	3	2.61
	CoAoC	4	2.75
	MP11	2	2.35
	MP11	3	**1.95**
	MP11	4	2.87
	MP24	2	**1.70**
	MP24	3	1.76
	MP24	4	2.22
A2	MP2	2	**1.52**
	MP2	3	1.80
	MP2	4	2.12
	Ta19	2	**1.65**
	Ta19	3	2.04
	Ta19	4	2.27
B1	Cs	2	1.97
	Cs	3	**1.60**
	Cs	4	2.25
B2	ALd	2	**1.46**
	ALd	3	1.89
	ALd	4	1.93
	EAo3	2	**1.73**
	EAo3	3	1.93
	EAo3	4	1.99
	Cs2	2	**2.00**
	Cs2	3	2.11
	Cs2	4	2.51
C1-C2	LEm	2	**1.97**
	LEm	3	2.19
	LEm	4	2.46
	CBo	2	2.02
	CBo	3	**1.55**
	CBo	4	2.08

* Summary index based on fuzziness performance index (FPI), normalized classification entropy (NCE), Xie-Beni index, Fukuyama index and a proportion exponent.

using 1000 trees in all zones ("*ntree*" argument), because using a higher number of trees did not improve the classification performance in any zone (results not shown). Three terrain attributes were randomly selected for each node in the tree (*mtry* argument), according to James et al. (2013).

Table 1 shows the fluctuations of the classification results for each experimental field. In all zones, the classification had more true predictions than false ones. Based on these results, we confirmed that the classification and spatial delimitation of SMUs from available soil survey for all experimental zones can be achieved using RF and terrain attributes, in agreement with that reported by Häring et al. (2012). The proportion of false predictions was related to (i) poor representation of the SMU area; hence, small proportion of data for classification with RF; and (ii) strong differences among SMUs at short distance due to changes in effective soil depth (fields A1 and A2) and soil texture (fields B1, B2 and C1-C2).

3.2. Importance of terrain attributes

RF allowed calculating the most important terrain attributes to classify each SMU and a global importance scale in each zone (Figs. 3 and 4). The importance of terrain attributes was not homogeneous for SMUs or experimental zones. Therefore, this process was able to characterize and identify the topographic fingerprint in each SMU.

In Zone A, elevation and valley depth were the most important to classify SMUs. Although aspect was the third most important for A1, it did not have the same importance for A2. In Zone B, convergence index, modified catchment area and channel network were globally the most important variables. Although curvature was the second most important variable in B1, it did not have the same importance for B2. In Zone C, channel network and convergence index were globally the most important terrain attributes.

These results confirmed those reported by Häring et al. (2012), who reported that variations of relative importance of terrain attributes provide essential information to classify SMUs. These variations define both small scale variations as well as landscape scale patterns. For example, the relative importance for elevation differed between MP11 and MP24 in field A1. MP11 has 40% of Balcarce soil series, which is classified as Petrocalcic Argiudolls, while MP24 presents soil series classified as Typic Argiudolls. It is well known that elevation patterns and petrocalcic horizon are spatially correlated, due to the soil genesis process in Zone A (Amiotti et al., 2001; Pazos and Mestelan, 2002). Thus, elevation is a remarkable terrain attribute to classify SMUs that include Petrocalcic Argiudoll soil series, as shown in SMUs MP11 and MP24.

3.3. Soil-specific zones

Table 2 shows summary index values for the SMUs in each experimental field. The optimal number of zones to delimit within each SMU is determined when the summary index value is minimized (Córdoba et al., 2013). MP11 in field A1, Cs in field B1 and CBo in field C1-C2 reached a minimum value when three zones were delimited, while the rest of SMUs reached this value when only two zones were delimited. Similar results have been reported by Peralta et al. (2015), who, by using ECa in the southeastern of the Argentine pampas, determined that elevation, effective soil depth, yield, FPI and NCE were minimized when delimiting two or three zones within agricultural fields. Results of zones delimited within each SMU for each field are shown in Fig. 5.

3.4. Effect of soil-specific zones on ECa

ECa was significantly different among zones delimited in every sets of samples, for each soil type in all experimental fields (Fig. 6). These results suggest that the proposed pedometric technique was efficient in delimiting zones in all experimental fields. The lack of differences for ECa_30 cm and ECa_90 cm was due to the particular features of each SMU. For example, ECa_30 cm showed a uniform pattern in SMU MP24, which was probably due to a heterogeneous soil profile at a depth further than 60 cm, as is shown for ECa_90 cm. A similar interpretation might be applied to SMU Ao3 in field B2. In TA19, the uniform pattern for ECa_90 cm could be explained by similitudes in the thickness and depth of the argillic horizon (Bt) between Azul and Tandil soil series (Myers et al., 2010).

ECa_30 cm and ECa_90 cm had greater statistical differences in zones delimited by soil type in fields C1-C2 than in the rest of SMUs (Fig. 6). This could be explained by the presence and thickness of horizons whose clay content was higher than 40% in fields C1-C2. It is widely known that clay content is a soil property which determines ECa spatial patterns (Vertic Argiudolls). According to Kitchen et al. (2005), ECa is a key source of information to identify soil types, when soils have high clay content. On the other hand, the significant differences for ECa_90 cm in field B2 might be because of the high salt content at a depth further than 30 cm, in zones related to SMU EAo3 in field B2 (Corwin et al., 2006).

In fields in Zone A, ECa_30 cm and ECa_90 cm had the least difference among soil-specific zones, and their values were similar to those reported by Peralta et al. (2015). Pazos and Mestelan (2002) determined that in soils whose SMUs have presence of Petrocalcic Argiudolls as SMUs MP11 in field A1 or TA19 and MP2 in field A2, the depth of the petrocalcic horizon varies at short distances (Domenech et al., 2017). Thus, complex patterns are found because of the vertical changes in clay content (Amiotti et al., 2001). On the other hand, high values for ECa_30 cm and ECa_90 cm can be due to zones where the depth of the petrocalcic horizon is smaller than 50 cm and a shallow argillic horizon exists, as proposed by Boettinger et al. (1997). In

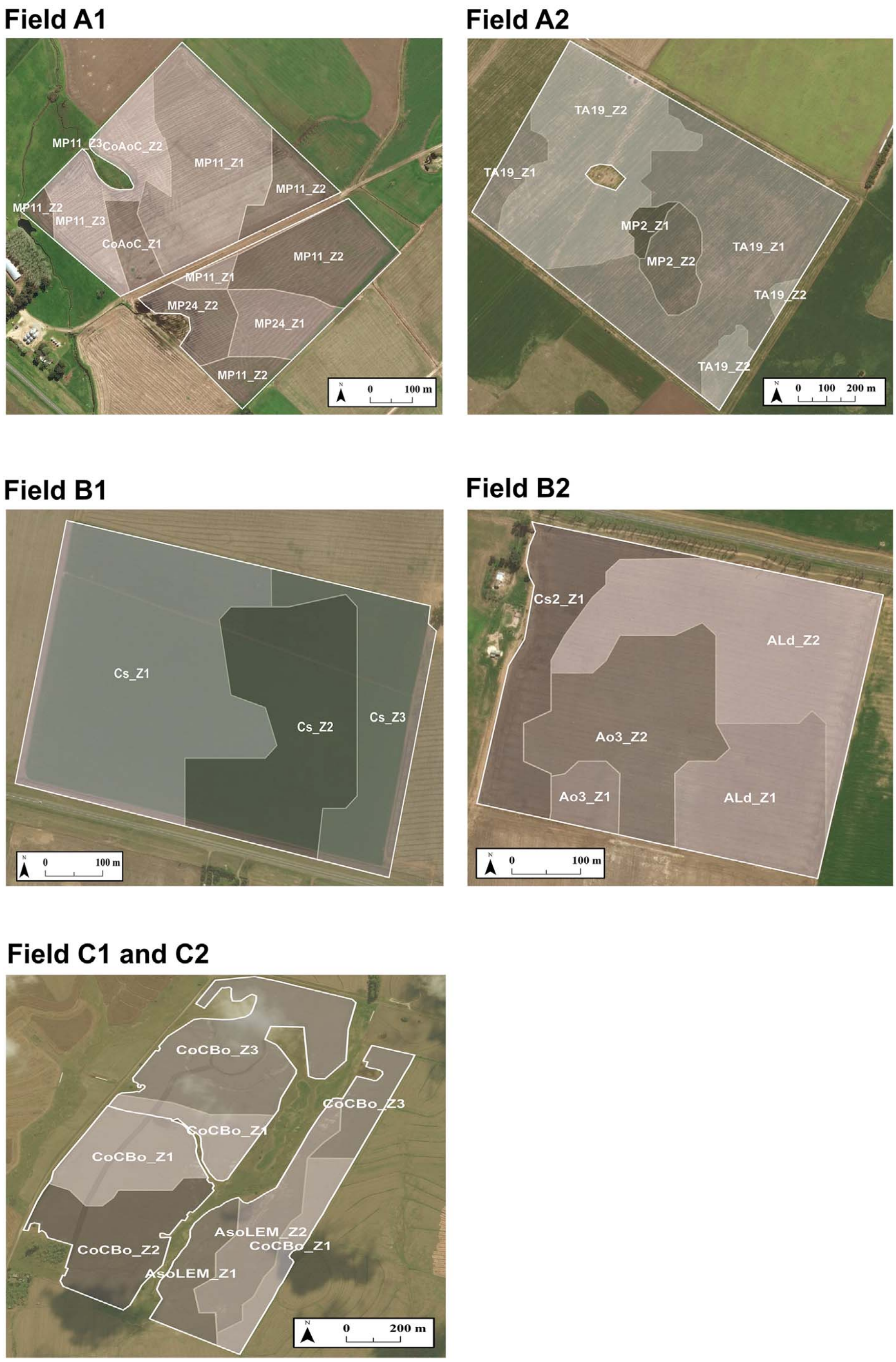

Fig. 5. Spatial distribution of soil-specific zones delimited in each experimental farm field.

contrast, low values for ECa_30 cm and ECa_90 cm can be due to the presence of Typic Argiudolls, which have an effective soil depth nearer than 80 cm, and to the presence of thicker and deeper argillic horizons (Doolittle et al., 1994).

3.5. Effect of soil-specific zones on yield

Soybean yield was significantly different among zones delimited in every sets of samples, for each zone in all experimental fields (Fig. 7). Based on these results (*i*), zone delimitation could be used to apply site-specific management and (*ii*) yield maps may allow the validation of zones defined by soil type at field scale. A few studies have reported the use of yield maps to validate pedometric techniques at field scale (Bobryk et al., 2016).

Soybean yield had the least difference in fields C1-C2 and the largest in fields A1 and B1. When comparing ECa and soybean yield among zones in fields C1-C2, their values did not change in the same magnitude. Those differences of ECa and yield among zones reflect that the magnitude of the effect of soil variability depends on the crop and its management practices. For example, Sadras and Calviño (2001) reported that in zones with presence of Petrocalcic Argiudolls as SMUs MP11 in field A1 or TA19 and MP2 in field A2, the response of crops to physical restriction of root elongation were strongly influenced by the interaction between crop phenology and seasonal pattern rain. So, these authors demonstrated a marked difference in crop yield response to soil depth among crop species and reported a ranking of tolerance to

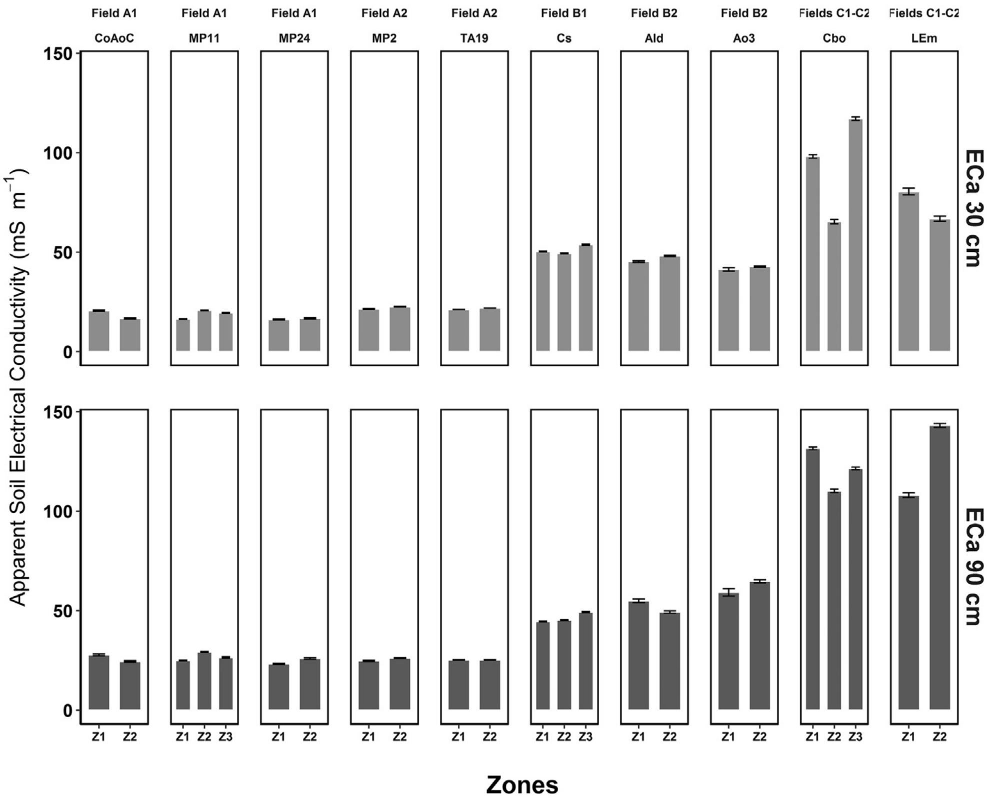

Fig. 6. Comparison between apparent electrical conductivity 0 to 30 cm (ECa_30 cm) and 0 to 90 cm (ECa_90 cm) for soil-specific zone in each experimental field.

shallow soil as: wheat > soybean > sunflower > maize. Thus, is possible that the magnitude of differences of ECa among zones reflect only the soil variability, whereas the magnitude of differences of crop yield among zones reflect the interaction soil variability, crop species and agronomic management practices.

Two aspects of our results should be highlighted. Firstly, the pedometric technique developed here allowed applying zone delimitation by soil type in agricultural fields from soil classification using conventional soil survey (1:50,000) and from zone delimitation for site-specific management. Numerous techniques of zone delimitation for site-specific purposes have been reported (Córdoba et al., 2013; Córdoba et al., 2016; Gavioli et al., 2016). However, a few of them have proposed using both soil classification and zone delimitation for site-specific management. Secondly, the simultaneous use of both ECa and crop yield allowed determining the usefulness of the delimitation. Therefore, the implementation of this pedometric technique could bring useful and worthy information, enabling farmers an improvement in the management of their resources and thus contributing to mitigate the environmental risk in their agricultural production system.

4. Conclusions

Here, we presented a robust pedometric technique to delimit soil-specific zones in farm fields, which coupled two other widely applied ones. The first was based on an automatic SMU classification using the *Random Forest* algorithm, information included in conventional soil survey and terrain attributes obtained from a digital elevation model. The automatic SMU classification allowed us to determine association rules among terrain attributes which determined the spatial distribution of each SMU. The second technique was based on a spatial disaggregation for each SMU using a cluster analysis of spatial principal components for terrain attributes in each SMU. This procedure was applied in six farm fields located in three Argentine agricultural regions, resulting in a delimitation of zones in each field. Zone validation for decision support of soil-specific management was carried out through ECa and crop yield maps, which are commonly used as sources of ancillary information for precision agriculture. Statistical differences for ECa and soybean yield for each soil-specific zone in all fields demonstrated that the technique developed was efficient in identifying variability and in classifying soil types at field scale. The lack in some cases of statistical differences was due to particular aspects related to the depth of the petrocalcic horizon and the depth and thickness of the argillic one. The validation of soil-specific zones using ECa and yield maps simultaneously demonstrated to be a useful method to quantify differences among zones and to determine their potential as a source of ancillary information for implementing precision agriculture practices. We conclude that the application of pedometric techniques involving soil-landscape relationships, information available in conventional soil survey, sources of topographic digital information and tools of precision agriculture, might be appropriate to delimit soil types in agricultural fields. This strategy may allow optimizing the use of inputs (seeds, fertilizers, and pesticides) and the practices of soil and crop management, mitigating the environmental risk on the current agricultural production system and increasing the profit of emergent technologies.

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. We would like to thank Virginia Aparicio for providing Veris® 3100 equipment and Luis Alonso for his assistance with field measurements.

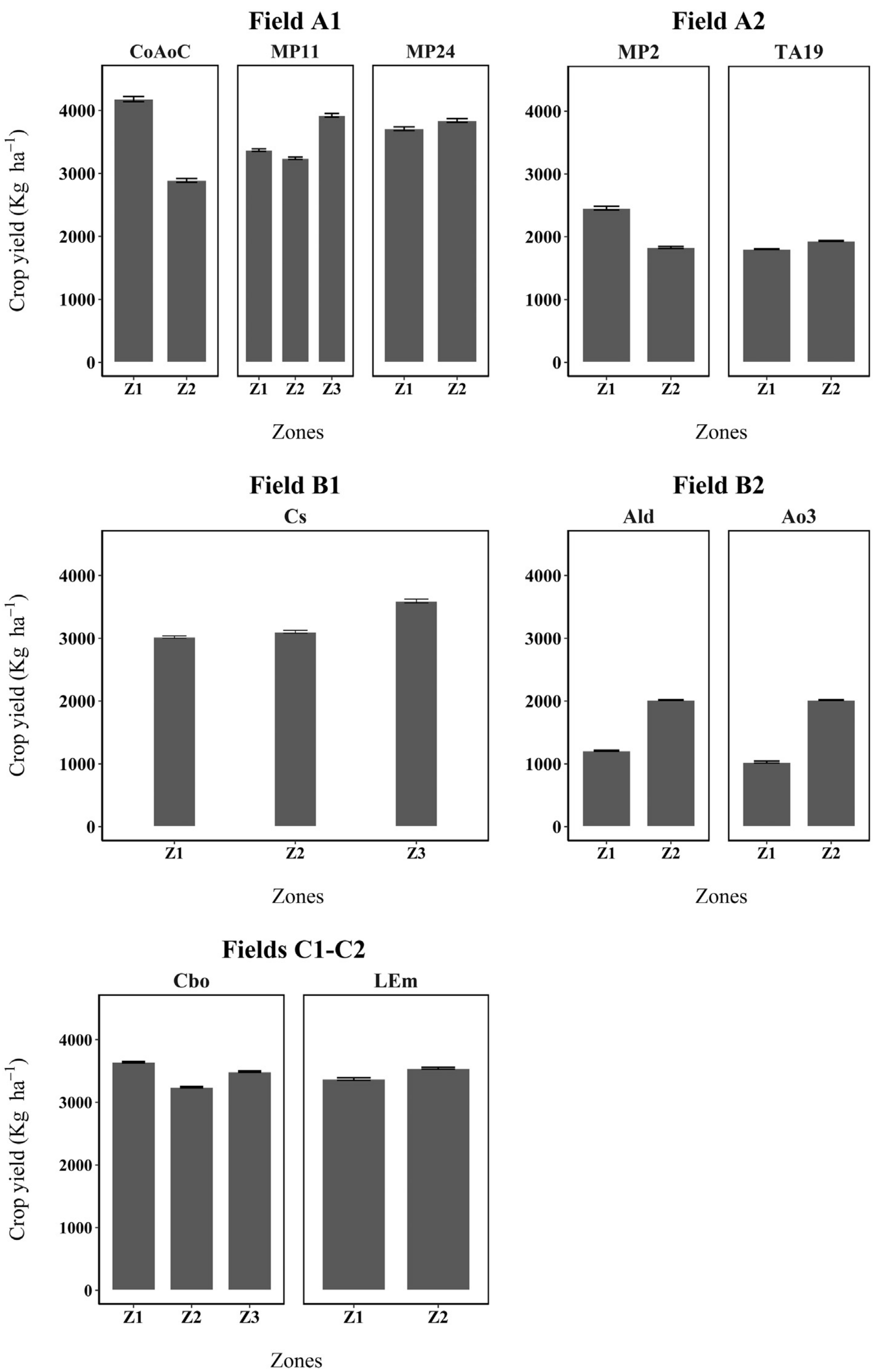

Fig. 7. Comparison of soybean yields for soil-specific zone in each experimental field.

References

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: a machine learning approach. Adv. Water Resour. 33 (1), 69–80.

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.

Blackmore, S., Moore, M., 1999. Remedial correction of yield map data. Precis. Agric. 1 (1), 53–66.

Bobryk, C.W., Myers, D.B., Kitchen, N.R., Shanahan, J.F., Sudduth, K.A., Drummond, S.T., Gunzenhauser, B., Gomez Raboteaux, N.N., 2016. Validating a digital soil map with corn yield data for precision agriculture decision support. Agron. J. 108 (3), 957–965.

Boettinger, J.L., Doolittle, J.A., West, N.E., Bork, E.W., Schupp, E.W., 1997. Nondestructive assessment of rangeland soil depth to petrocalcic horizon using electromagnetic induction. Arid Soil Res. Rehabil. 11 (4), 375–390.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240 (0), 68–83.

Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103 (1–2), 79–94.

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.

Córdoba, M., Bruno, C., Costa, J., Balzarini, M., 2013. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. Comput. Electron. Agric. 97 (0), 6–14.

Córdoba, M.A., Bruno, C.I., Costa, J.L., Peralta, N.R., Balzarini, M.G., 2016. Protocol for multivariate homogeneous zone delineation in precision agriculture. Biosyst. Eng. 143, 95–107.

Corwin, D.L., Lesch, S.M., 2005. Characterizing soil spatial variability with apparent soil electrical conductivity: I. Survey protocols. Comput. Electron. Agric. 46 (1–3), 103–133.
Corwin, D.L., Lesch, S.M., Oster, J.D., Kaffka, S.R., 2006. Monitoring management-induced spatio–temporal changes in soil quality through soil sampling directed by apparent electrical conductivity. Geoderma 131 (3–4), 369–387.
Di Rienzo, J.A., Casanoves, F., Balzarini, M.G., Gonzalez, L., Tablada, M., Robledo, C.W., 2015. InfoStat.
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.
Doolittle, J.A., Sudduth, K.A., Kitchen, N.R., Indorante, S.J., 1994. Estimating depths to claypans using electromagnetic induction methods. J. Soil Water Conserv. 49 (6), 572–575.
Fraisse, C., Sudduth, K., Kitchen, N., 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Trans. ASAE 44 (1), 155–166.
Fukuyama, Y., Sugeno, M., 1989. A new method of choosing the number of clusters for the fuzzy c-means method. In: Proc. 5th Fuzzy Syst. Symp, pp. 247–250.
Gambill, D.R., Wall, W.A., Fulton, A.J., Howard, H.R., 2016. Predicting USCS soil classification from soil property variables using random Forest. J. Terrramech. 65, 85–92.
Gavioli, A., de Souza, E.G., Bazzi, C.L., Guedes, L.P.C., Schenatto, K., 2016. Optimization of management zone delineation by using spatial principal components. Comput. Electron. Agric. 127, 302–310.
Gebbers, R., Adamchuk, V.I., 2010. Precision agriculture and food security. Science 327 (5967), 828–831.
Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: a decision-tree based approach in Bavarian forest soils. Geoderma 185–186, 37–47.
Hempel, J., Hammer, R., Moore, A., Bell, J., Thompson, J., Golden, M., 2008. Challenges to digital soil mapping, Digital Soil Mapping with Limited Data. Springer, pp. 81–90.
Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77.
IGN, 2016. Modelo digital de elevaciones de la República Argentina. Instituto Geográfico Nacional de la República Argentina, Buenos Aires Argetina, pp. 83.
INTA, 2010. Carta de Suelos de la Provincia de Buenos Aires. Instituto Nacional de Tecnología Agropecuaria, Buenos Aires - Argentina.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer, New York.
Jenny, H., 1941. Factors of soil formation. In: A System of Quantitative Pedology. McGraw-Hill Book Company, New York, London.
Johnson, C.K., Doran, J.W., Duke, H.R., Wienhold, B.J., Eskridge, K.M., Shanahan, J.F., 2001. Field-scale electrical conductivity mapping for delineating soil condition. Soil Sci. Soc. Am. J. 65 (6), 1829–1837.
Kitchen, N.R., Sudduth, K.A., Myers, D.B., Drummond, S.T., Hong, S.Y., 2005. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. Comput. Electron. Agric. 46 (1–3), 285–308.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2 (3), 18–22.
Massawe, B.H.J., Subburayalu, S.K., Kaaya, A.K., Winowiecki, L., Slater, B.K., (n.d.) Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. Geoderma.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52.
Moscatelli, G., Pazos, M.S., 2000. Soils of Argentina: Nature and Use (April). pp. 17–22.
Myers, D.B., Kitchen, N.R., Sudduth, K.A., Grunwald, S., Miles, R.J., Sadler, E.J., Udawatta, R.P., 2010. Combining proximal and penetrating soil electrical conductivity sensors for high-resolution digital soil mapping. In: Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.), Proximal Soil Sensing. Progress in Soil Science. Springer, Netherlands, pp. 233–243.
Nauman, T.W., Thompson, J.A., 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213 (0), 385–399.
Nitze, I., Schulthess, U., Asche, H., 2012. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. In: Proc. of the 4th GEOBIA, pp. 7–9.
Odeh, I.O.A., Chittleborough, D.J., McBratney, A.B., 1992. Soil pattern recognition with fuzzy-c-means: application to classification and soil-landform interrelationships. Soil Sci. Soc. Am. J. 56 (2), 505–516.
Pal, N.R., Bezdek, J.C., 1995. On cluster validity for the fuzzy c-means model. IEEE Trans. Fuzzy Syst. 3 (3), 370–379.
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.
Pennock, D., McKenzie, N.J., Montanarrella, L., 2015. Status of the World's Soil Resources. Technical Summary FAO, Rome, Italy.
Peralta, N.R., Costa, J.L., Balzarini, M., Castro Franco, M., Córdoba, M., Bullock, D., 2015. Delineation of management zones to improve nitrogen management of wheat. Comput. Electron. Agric. 110 (0), 103–113.
QGIS v2.16.1 Nodebo, 2016. Development Team. Quantum GIS geographic information system, Essen, Germany.
R Core Team, 2016. R: A Language and Enviroment for Stadistical Computing. R Fundation for Stadistical Computing, Viena, Austria.
Rossiter, D., 2012. A pedometric approach to valuing the soil resource. In: Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia. CRC Press, pp. 25.
Sadras, V.O., Calviño, P.A., 2001. Quantification of grain yield response to soil depth in soybean, maize, sunflower, and wheat. Agron. J. 93 (3), 577–583.
SAGA Development Team, 2016. System for Automated Geoscientific Analysis (SAGA).
Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., de Lourdes Mendonça-Santos, M., 2009. Digital soil map of the world. Science 325 (5941), 680–681.
Scull, P., Franklin, J., Chadwick, O., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27 (2), 171–197.
Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed. United States Department of Agriculture - Natural Resources Conservation Service. Washington, DC.
Vogels, M.F.A., de Jong, S.M., Sterk, G., Addink, E.A., 2017. Agricultural cropland mapping using black-and-white aerial photography, object-based image analysis and random forests. Int. J. Appl. Earth Obs. Geoinf. 54, 114–123.
Windham, M.P., 1981. Cluster validity for fuzzy clustering algorithms. Fuzzy Sets Syst. 5 (2), 177–185.