Catena xxx (xxxx) xxx-xxx

Contents lists available at ScienceDirect

Catena

journal homepage: http://ees.elsevier.com

Effectiveness of a soil mapping geomatic approach to predict the spatial distribution of soil types and their properties

Costanza Calzolari[a], Fabrizio Ungaro[a], Andrea Vacca[b]

[a] CNR Institute for BioEconomy, via Madonna del Piano 10, Sesto Fiorentino 50019, Italy
[b] Dipartimento di Scienze Chimiche e Geologiche, Università degli Studi di Cagliari, Cittadella Universitaria (Blocco D), Monserrato 09042, Italy

ARTICLE INFO

Keywords
Soil mapping
GIS
Landforms and parent materials
Stepwise multiple linear regressions
General linear models
WRB Reference Soil Groups

ABSTRACT

A soil map (1:50,000 scale) was recently produced in Sardinia (Italy) using a cost-effective GIS approach. In this study we aimed to verify, in two pilot areas and by means of statistical analysis, the effectiveness of the adopted methodology in representing and predicting the spatial distribution of soil types and properties. We focused on evaluation of 1) the influence of landforms and parent materials on soil types (WRB Reference Soil Groups) and selected soil properties and 2) the suitability of the adopted methodology for calibrating a model to predict land unit composition in terms of different soil types. Leptosols, Regosols and Cambisols were prevalent on slopes, with Leptosols being more frequent on convex slopes and Regosols and Cambisols on concave slopes. In flat areas, soil types mainly depended on the type and age of parent material, with Regosols and Cambisols prevailing on Holocene deposits and highly developed soils (mainly Luvisols) largely prevailing on Pleistocene deposits. On hard rock, Leptosols were very frequent on terrigenous metamorphic rock and frequent on granite. Besides Leptosols, Regosols occurred more frequently than Cambisols on both parent materials. Landforms strongly influenced soil depth and available water capacity. Soils on plains were deeper than those on slopes, where convex forms had shallower soils than concave forms. A similar trend applied to the available water capacity. The parent material had a significant effect on topsoil properties (thickness, texture, pH and organic carbon content) of soils belonging to the same WRB Reference Soil Group (analysis done on the most relevant WRB Reference Soil Groups, i.e. Leptosols, Regosols and Cambisols). We calibrated and tested stepwise multiple linear regressions (MLR) and general linear models (GLM) to predict the composition of map units in terms of different WRB Reference Soil Groups. The two models gave very similar results, with distinct distribution patterns that were coherent with the relationships observed between soil groups and specific combination of terrain attributes and parent materials. Results showed that both models were more reliable in predicting the absence rather than presence of a given soil type.

1. Introduction

There is an increasing demand for soil information, since the knowledge and understanding of soil and how it is distributed across the landscape is considered essential for its effective use, management and conservation (Grealish et al., 2015). Consequently, soil information is also essential to help decision makers in land planning and in drafting environmental management policy (van Delden et al., 2011; Brungard et al., 2015), and may even be required by law (Vacca et al., 2014).

Soil information can be provided by soil maps, which are graphic representations for transmitting information about the spatial distribution of soil attributes (Yaalon, 1989). In general terms, soil maps can be produced in a conventional or in a digital way. Conceptually, conventional soil maps and digital soil maps (DSM) are very similar (Kempen et al., 2012), since both approaches use a soil-landscape model to predict soil at unobserved locations (Hudson, 1992). The main difference is that in conventional soil maps the soil-landscape model is a qualitative model based on soil surveyors' expert knowledge, while in DSM the soil-landscape model is quantitative. Comprehensive overviews of DSM were provided by McBratney et al. (2003), Grunwald (2006), Minasny and McBratney (2016), and Arrouays et al. (2020). Among the first conceptualizations of DSM, McBratney et al. (2003) formalized the so-called scorpan model as "empirical quantitative descriptions of relationships between soil and other spatially referenced factors with a view to using these as soil spatial prediction functions". The possibility of producing DSM strongly depends on the availability of ancillary data (Zeraatpisheh et al., 2017), including existing soil data (e.g. polygon-based soil maps and soil profile databases), which can serve as both training and validation datasets (Zhang et al., 2017). Consequently, in areas with limited existing soil data, producing an accurate DSM can be challenging (Stoorvogel et al., 2009), so this method has rarely been used for routine production mapping or addressing land management questions (Grealish et al., 2015). In these areas, pragmatic and easy-to-apply relationships for predicting soil properties under different environmental conditions, and assist in soil data collection, are needed to provide answers for the current issues that require a fast delivery of information (Gray et

https://doi.org/10.1016/j.catena.2020.104818
Received 19 November 2019; Received in revised form 20 July 2020; Accepted 28 July 2020
Available online xxx
0341-8162/

al., 2009; Grealish et al., 2015). Several different statistical approaches have been tested to generate quantitative predictions of categorical soil variables from limited samples with the general aim of producing soil maps for unsampled or sparsely sampled areas at different scales, from national to sub-regional (Minasny and McBratney, 2016). Grunwald (2009) provided a multi-criteria characterization of digital soil mapping and modelling approaches, classifying DSM techniques in three wide categories based upon predictor variables and modelling approaches, which can then integrate data driven statistical approaches with pedotransfer functions and dynamic mechanistic modelling of soil properties.

In Sardinia (Italy), the use of soil information and maps in land use planning is specifically required by law (RAS, 2006, 2008). Because the scale of the three available soil maps covering the island (Arangino et al., 1986; Aru et al., 1990; Madrau et al., 2006) was considered not adequate for local land planning strategies, a new project was recently initiated for the production of a new soil map, at a scale of 1:50,000. The general structure of the project and the methodology used were described in Vacca et al. (2014). The existing soil dataset (point data and maps) was considered insufficient and inappropriate to produce a DSM without resorting to the support of ancillary variables. Adopting a cost-effective approach, existing digital environmental data, along with soil data, were therefore used to delineate homogeneous spatial areas in terms of soil, geological substrate, landform, and land cover in a GIS environment.

This paper aimed to verify, in two of the pilot areas and by means of statistical analysis, the effectiveness of the adopted methodology (Vacca et al., 2014) in representing and predicting the spatial distribution of soil types and soil properties. This is considered crucial, as it affects the reproducibility of the model. There appears to be a need for clarification of the quantitative relationships between soil properties and environmental covariates in order to reduce the uncertainty of the model and allow better prediction.

The specific objectives of this paper were to (1) evaluate the influence of landforms and parent material on soil types; (2) evaluate their influence on soil properties; and (3) evaluate if the adopted methodology is suitable for calibrating a model to predict land units composition in terms of principal soil types.

2. Materials and methods

2.1. Study area

This study was conducted in two pilot areas of Pula and Muravera, located in southern Sardinia (Italy), as shown in Fig. 1. The two areas have similar geology, topography, climate and land use. There are two distinct physiographic regions within each area: a hilly part and a coastal plain. Geology of the hilly sectors consists mainly of Paleozoic metamorphic rocks, which were deformed and affected by low-grade metamorphism during the Hercynian orogenesis, and granitoids of the Carboniferous (RAS, 2010). Quaternary slope deposits are also present. In a small sector of the eastern study area, the Paleozoic rocks are unconformably overlain by Eocene sediments made up of conglomerate and sandstone that pass to marlstone. In the western study area, there are extensive outcrops of Oligo-Miocene andesite. Geology of the coastal plains mainly consists of Quaternary fluvial and alluvial deposits. In the hilly sector, the reliefs are characterized by deep valleys with steep slopes that were cut by the main rivers and streams. These features were mainly driven by the Plio-Pleistocene tectonic evolution, which was responsible for the uplift of the areas and the present dissected morphology (Carmignani et al., 2001; Barca et al., 2009). The valleys follow these lineaments, the axis of schistosity, and tectonic contacts. On the coastal plains, depositional glacis, alluvial fans and fluvial terraces are the main depositional landforms. Surface slope dynamics (erosion/deposition) are the main geomorphic processes in the hilly sectors. Streams are characterized by a torrential regime, with alternating periods of drought, or minimum flow, and floods.

The study areas have a warm temperate Mediterranean climate (Mediterranean-subcontinental, *sensu* Finke et al., 2001). The pedoclimate is characterized by a thermic soil temperature regime associated with a xeric soil moisture regime (*sensu* Soil Survey Staff, 2014).

The Mediterranean maquis, in its different stages of development/degradation, and holm oak (*Quercus ilex*) and cork oak (*Quercus suber*) woods are the most widespread land cover in the hilly sectors. Land use in these areas is mainly related to recreation, but there is extensive pasturing of goats and pigs and coppicing for firewood production. On the plains, the main land use is agriculture but, as a result of land abandonment, there are also significant areas of natural vegetation.

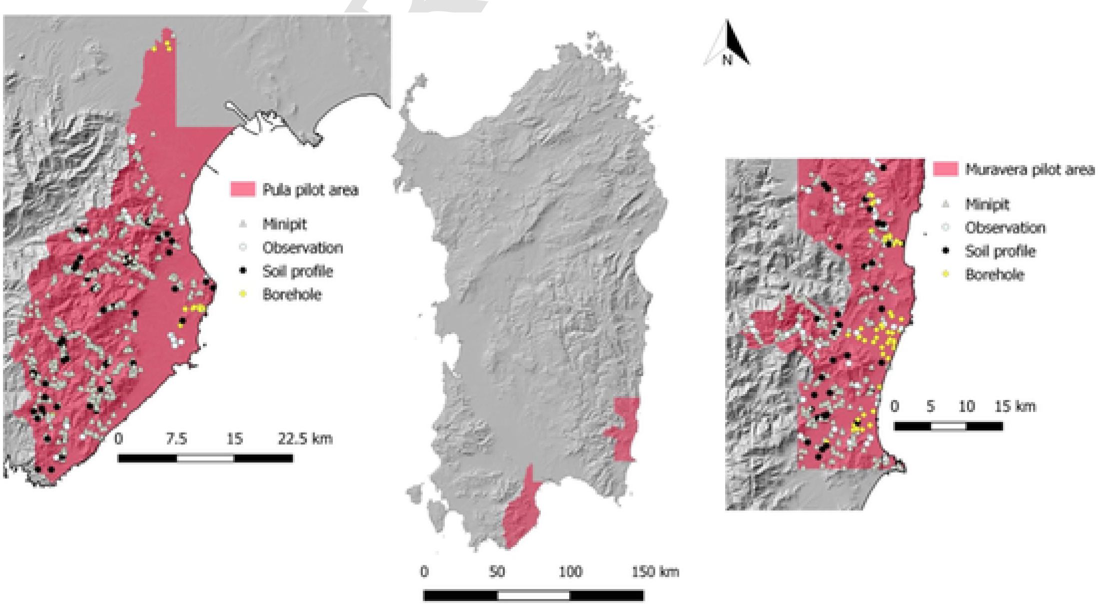

Fig. 1. Study areas. Left, Pula pilot area; right, Muravera pilot area.

2.2. GIS based soil mapping

A full description of the methodology used to produce a new soil map, at a scale of 1:50,000, to support land use planning in Sardinia was given in Vacca et al. (2014). The map was produced using a GIS-based approach and applying the soil-landscape paradigm (Hudson, 1992). The available digital data on soil-forming factors (land cover, geology and topography) were processed and classified according to their influence on weathering and soil properties, to allow prediction of the soil types and their spatial distribution or prediction of their properties. The methods used in the interpretation were based on consolidated and generalized knowledge about the influence of geology, topography and land cover on soil properties. Validated and standardized existing soil data were input into a database specifically created for the project, which included soil properties and soil classification according to the World Reference Base (IUSS Working Group WRB, 2015). Using expert interpretation and applying GIS functions, all digital data were merged to produce a first draft map, which was implemented with the existing soil data and new soil data collected during a field survey. The survey density unit was of 1 observation per 50 ha, with at least 5% of the observations being soil profiles. Following soil data implementation and field validation, the final map was produced.

In this study we used two of the scorpan factors (McBratney et al., 2003), i.e. relief and parent material, to derive spatial soil prediction functions in terms of occurrence of WRB RSGs.

For the landform classification, a DEM, year 2011, distributed in ESRI GRID format, with 10-m pixel pitch and a vertical and horizontal accuracy of 2.5 m was used. Landform was classified on the basis of slope and curvature (*sensu* Zevenbergen and Thorne, 1987). These two parameters can indicate, particularly the curvature, whether a cell is prone to accumulate water (Shary et al., 2002; MacMilland and Shary, 2009) and, consequently, can delineate the areas with potential sediment accumulation and soil development and those prone to soil erosion. Four classes of slope were used, namely > 35%, 15–35%, 2.5–15%, and < 2.5%, based on the regional soil distribution model (Arangino et al., 1986; Aru et al., 1990; Madrau et al., 2006; Marrone et al., 2008). A description of the GIS based landform units (LFU) is given in Table 1.

Using the 1:25,000 scale geological map of Sardinia (http://www.sardegnageoportale.it/webgis2/sardegnamappe/?map=mappetematiche), a map of parent material with 58 new units was derived according to their influence on weathering processes and soil properties. Lithology was the main criteria used for the grouping, but genetic character, texture, structure, composition and age were also used (Birkeland, 1999; Brady and Weil, 2008; Sierra et al., 2009; Buol et al., 2011). Following the soil data implementation and field validation, a final map with 48 soil mapping units was produced. Of these, 26 were present in the two pilot areas of this paper.

Table 1
The seven landform units used to compile the landform map.

Curvature	Slope	Landform unit	Description
Concave	> 35%	−3	Concave areas with slope > 35%. Depending on lithology, this includes concave areas on mountain slopes and upper parts of hillslopes.
Concave	15–35%	−2	Concave areas with slope from 15 to 35%. Depending on lithology, this includes concave areas on mountain middle slopes, hillslopes and upper parts of fans.
Concave	2.5–15%	−1	Concave areas with slope from 2.5 to 15%. Depending on lithology, this includes concave areas on lower parts of hillslopes, floodplains, fans and erosional surfaces.
	< 2.5%	0	Concave and convex areas with slope < 2.5%. Depending on lithology, this includes alluvial plains, terraced surfaces, planation surfaces, morphostructural features, erosional and sedimentation plains.
Convex	2.5–15%	1	Convex areas with slope from 2.5 to 15%. Depending on lithology, this includes convex areas on ridges, fans, erosional and sedimentation surfaces located between hillslopes and plains.
Convex	15–35%	2	Convex areas with slope from 15 to 35%. Depending on lithology, this includes convex areas on upper parts of ridges, upper parts of fans, morphostructural features.
Convex	> 35%	3	Convex areas with slope > 35%. Depending on lithology, this includes convex areas on mountain ridges and edges of morphostructural features.

2.3. Soil data: landforms, parent material and soil types

The dataset from the two pilot areas consisted of a total of 1461 georeferenced observations (minipits 75%, boreholes 7%, soil profiles 10%, roadcuts and other exposures 8%), 782 from the Pula pilot area and 679 from the Muravera pilot area (Fig. 1). A table reporting the occurrence of the observations in the different landform units for each area is provided in the supplementary material. The following data were available: site description (all observations), morphological description (soil profiles, minipits and boreholes) and a set of laboratory data for soil profiles, including sand, silt and clay fractions, pH, organic carbon content, and water content at field capacity (−33 kPa) and at wilting point (−1500 kPa). Moreover, field estimated texture and an estimate of available water capacity calculated by pedotransfer functions (Saxton et al., 1986) were available for each soil horizon of the minipits. To study the relationships between parent material and soil properties of functional relevance for the most widespread soil types, the following available data for the topsoil horizon were used: thickness (cm, N = 1396), sand, silt and clay fractions (%, N = 532), pH (-, N = 512) and organic carbon content (%, N = 315).

Based on the main lithological characteristics, the 26 parent materials originally recognized in the two pilot areas were grouped in the following twelve groups (the abbreviations are given in brackets): terrigenous metamorphic rock (M); limestone and marble (K); marl, clay, and lacustrine deposits (F); acidic vulcanite (P); basic vulcanite (A); granite (Y); sandstone (D); colluvium deposits (DC); slope deposits (DV), which were then classified according to age in Holocene slope deposits (DVO), and Pleistocene slope deposits (DVP); Pleistocene deposits (DP); alluvial deposits (AL).

The numerical consistence of the dataset in terms of parent material is summarized in Table 2.

According to WRB (IUSS Working Group WRB, 2015) soils were mainly classified as Regosols, Leptosols, and Cambisols (N = 406, 315 and 240 respectively, 75% of observations). Well developed soils (Luvisols, Alisols and Lixisols) represented 13% of observations, with most being classified as Luvisols (N = 148). Other Reference Soil Groups (RSG) were sporadic or linked to specific parent materials and landforms: Fluvisols (3% of observations) were present on alluvial deposits of the main rivers; Phaeozems (4.1% of observations) and Umbrisols (2.5% of observations) occurred on > 15% slopes, where dominant land cover was forest and/or pasture. For 185 observations a classification was not provided. A table reporting the occurrence of the different RSGs is provided in the supplementary material.

Table 2
Number of observations for the different parent materials in the two pilot areas.

Parent material	Label	Number of observations	%
Basic vulcanite	A	18	1.2
Alluvial deposits (Holocene)	AL	318	21.8
Sandstone	D	27	1.8
Colluvial deposits	DC	52	3.6
Pleistocene deposits	DP	116	7.9
Slope deposits (Holocene and Pleistocene)	DV	210	14.4
Holocene	DVO	155	
Pleistocene	DVP	55	
Marl, clay, lacustrine deposits	F	36	2.5
Limestone, marble	K	20	1.4
Terrigenous metamorphic rock	M	252	17.3
Acid vulcanite	P	2	0.1
Granite	Y	410	28.0
Total		1461	

2.4. Statistical analysis

We used statistical significance tests (Tukey Kramer HSD for unequal sample size) to highlight whether and to what extent landforms and parent material affect soil properties, and to evaluate differences among soil types in terms of their properties, and their occurrence in the observed combinations of landforms and parent material. With the aim of assessing whether available data could support predictive soil mapping, stepwise multiple linear regressions (MLR) and general linear models (GLM, Nelder and Wedderburn, 1972) were calibrated to assess the occurrence of soil groups as a function of parent material and landform. In the case of MLR, we dummy-coded the parent material classes and the landform units' classes. The GLM, the popularity of which is increasing in digital soil mapping applications due to its flexibility (Lane, 2002; Evans and Hartemink, 2014; Mosleh et al., 2016), is a linear predictor that relates the mean of the response variable to a linear combination of the explanatory variables (Edward et al., 2012). The GLM is a generalization of the linear regression model, as effects can be tested for categorical predictor variables, as well as for continuous predictor variables. Furthermore, designs can have a single or multiple dependent variables. The relationship between response and predictors is not linear, and a link function provides a transformation of the response so that the transformed response is linearly related to the predictors. A GLM with binomial data, such as the presence/absence of a given soil order (or other taxon), is called logistic regression. In this case, the link function is a logit function, which is the log of the odds ratio (probability of presence/probability of absence).

Root mean square error (RMSE) and mean error (ME) were considered to assess the performance of the models. The statistical analyses were performed using the Statistica 64 v.12 software (StatSoft Inc., 2013).

3. Results and discussion

3.1. Relationships between morphometric parameters and parent material

Considering only the most common parent materials among those described in Table 2 (AL, DP, DC, DV, M, Y; 1358 observations), some relationships with morphometric parameters were found (Fig. 2). Metamorphic rocks, M, were significantly ($p < 0.05$) associated to >15% sloping convex slopes (35% on landform unit 3, 28% on landform unit 2). The same trend was observed on granite even if a relatively higher (and not statistically significant) percentage of observations fell in areas with slope gradient between 15 and 35%. These relationships were coherent with the tectonic history of the area (Plio-Pleistocene uplift and subsequent valley formation due to cutting by the main rivers and streams) and with the stronger resistance to weathering and erosion of metamorphic rock relative to granite. The higher percentage of granite in the 1, 0, and −1 landform units reflected the presence of erosional glacis between the hilly and plain sectors. These landforms were not found on metamorphic rock.

Most of observations on slope deposits, DV, were on >15% sloping concave areas (48% on landform unit −3 and 22% on landform unit −2), while colluvial deposits, DC, were associated with generally gently sloping (<15%) areas. Holocene alluvial deposits, AL, were in more than 80% of observed cases in flat areas (<2.5%). All these results were coherent with the well-known relationships existing between topography and sediments (Huggett and Cheesman, 2002). Pleistocene deposits, DP, were associated with flat to gently sloping areas. Some DP can also be found in sloping areas associated with convex morphologies. These relationships were consistent with the fact that large parts of DP were present in the form of depositional glacis.

3.2. The influence of landform and parent material on soil types

To assess the influence of landform and parent material on the distribution of WRB RSGs (IUSS Working Group WRB, 2015) only the most commonly found parent materials and soil groups were considered (N = 1175). The slope deposits were distinguished between Holocene, DVO, and Pleistocene, DVP, and the more developed soils, Luvisols, Alisols and Lixisols, were considered as a group (AliLuvLix); other sporadic RSGs were not considered.

As regards the distribution of soil types on slopes (Fig. 3a), Leptosols were much more frequent on convex (N = 223) than on concave slopes (N = 90). On the contrary, the frequency of Regosols and Cambisols was higher on concave slopes (206 vs 156 observations for Regosols and 118 vs 78 observations for Cambisols). Consequently, there was a clear indication that shallower soils (Leptosols) were more frequent in areas with a predominant erosional character (convex slopes), while deeper soils (Regosols and Cambisols) were more frequent in areas where depositional processes predominate (concave slopes), similarly to what found by Catani et al. (2010). As expected, Fluvisols were mostly present in flat areas (N = 29), together with Regosols, Cambisols and highly developed soils (Luvisols, Alisols and Lixisols). The preferential distribution of these highly developed soils on landforms with slopes not greater than 15% (N = 120) reflected their more usual development on stable landform positions (Scarciglia et al., 2011). Arenosols, being mainly formed on Holocene alluvial deposits, AL, and on colluvial deposits, DC, (Fig. 3b), were mainly found in flat areas (N = 10). As regards Umbrisols and Phaeozems, their absence or lower presence, respectively, in flat areas was due to intensive agricultural activity and the consequent reduction of organic carbon content in soils of these areas.

The relationships between WRB RSGs and parent material are shown in Fig. 3b. On Holocene alluvial deposits, AL, only 41% of the 318 observations were classified; of these 28% (N = 38) were Fluvisols, and 39% Regosols (N = 54). The other observations were mostly Cambisols (N = 37). On colluvial deposits, DC, the observed soils (N = 51) were classified in 70% of cases as Regosols and Cambisols (35% each). Phaeozems represented 20% of observations (N = 10). On Pleistocene deposits, DP, 95% of observations (N = 109) were of highly developed soils, mainly Luvisols (N = 91), Alisols (N = 16) and Lixisols (N = 2). These soils were occasionally associated to Cambisols (N = 4) and, more rarely, to Regosols, in more eroded sites, and to Phaeozems, in areas with denser natural vegetation cover. The soils observed on Holocene slope deposits, DVO (N = 156), were mainly Cambisols and Regosols. Umbrisols were sporadic (N = 18) associated

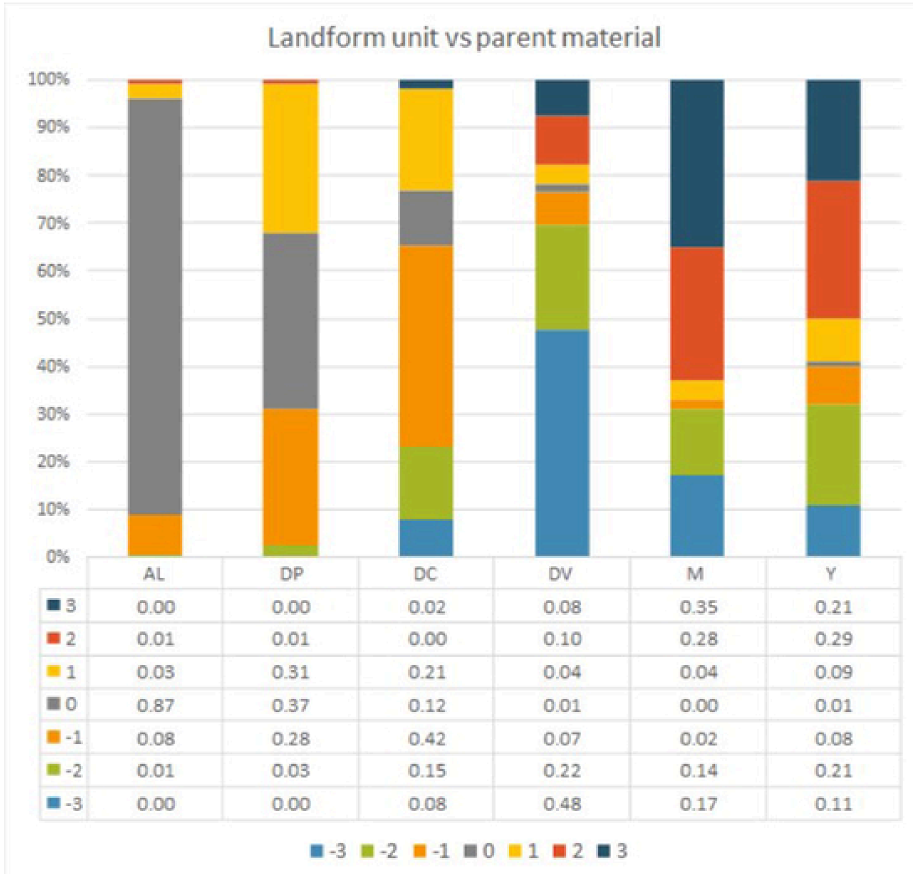

Fig. 2. Relationships between landform units and main parent materials: AL, Alluvial deposits (Holocene); DP, Pleistocene deposits; DC, Colluvial deposits; DV, Slope deposits; M, Terrigenous metamorphic rocks; Y, Granites. For landform unit's description see Table 1.

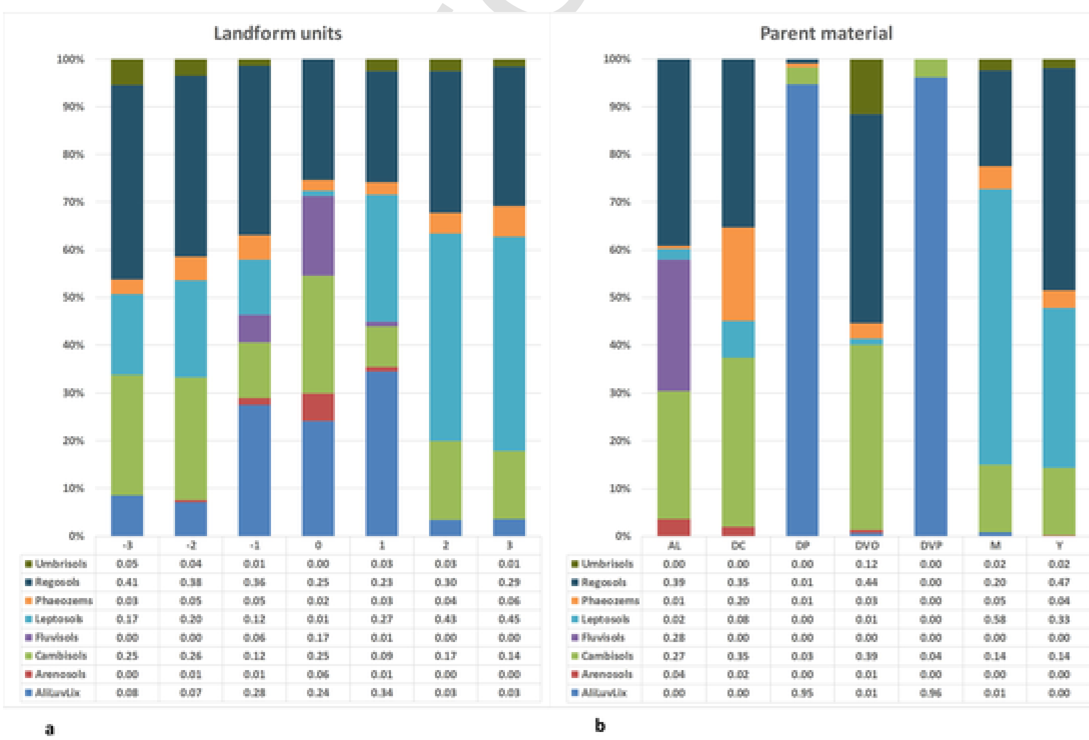

Fig. 3. Relationships between WRB RSGs and landform units (a) and parent materials (b). For landform unit's description see Table 1. AL, Alluvial deposits (Holocene); DP, Pleistocene deposits; DC, Colluvial deposits; DV, Slope deposits; M, Terrigenous metamorphic rocks; Y, Granites.

to forests and uncropped areas. On Pleistocene slope deposits, DVP, 96% of observations (N = 52) were classified as Luvisols. On terrigenous metamorphic rocks, M, the observed soils (N = 253) were Leptosols in 58% of cases (N = 146). They were significantly ($p < 0.05$) more common on convex slopes with >15% slope. Regosols (20% of observations) were more common on concave slopes with slope between 15% and 35%. About 10–20% Cambisols occurred in all landform units, with a relatively higher occurrence on unit 2.

The distribution of soils on granite parent material, Y, was dominated by Regosols (47%, N = 191) and Leptosols (33%, N = 137). It is interesting to note that the relative proportions of Leptosols and Regosols were significantly inverted ($p < 0.05$) as compared to soils developed on M. As for landform units, no significant difference was observed in soil type distribution on different morphologies, even if Leptosols were more frequent on convex slopes.

Two groups of most common parent materials were recognized in the area: sediments (AL, DC, DP, DVO and DVP) and hard rocks (M and Y). On sediments, Leptosols were very rare and limited to areas where the sedimentary cover, over hard rock, is very shallow, or where the volume of fine earth was <20% (soils where rock fragments were dominant). On the contrary, Leptosols were very frequent on M and frequent on Y, which are both hard rocks. As most of M in the area was represented by siliceous meta-sandstones, which are very resistant to weathering (Brady and Weil, 2008), Leptosols were more frequent on this parent material than on Y, where mineralogy and crystal grain size favored deeper weathering (Graham et al., 1997; Mareschal et al., 2015). Nevertheless, on both hard rocks, soil development was generally very weak: besides Leptosols, Regosols and Cambisols can be found, with a prevalence of the former. The apparent inconsistency of the degree of soil development with the old age of the substrates (both M and Y are Paleozoic), clearly highlighted a strong rejuvenation of the weathering front by morphodynamic processes, as M and Y were mainly associated to steeper slopes. The distribution of Phaeozems and Umbrisols was related to areas with denser natural vegetation cover. On sediments, the distribution of major WRB RSGs was related to the time factor (age of parent material). On Holocene deposits (AL, DC and DVO), very weakly developed soils (Regosols) and soils with at least an incipient subsurface soil formation (Cambisols) prevailed, while on Pleistocene deposits (DP and DVP) highly developed soils (mainly Luvisols) were more common. This is in agreement with the results of Carboni et al. (2006) and Scarciglia et al. (2011), who reported, for other areas of Sardinia, clay translocation and a high degree of weathering in soils formed on Pleistocene detrital deposits and their absence in soils formed on Holocene detrital deposits.

Fig. 4 shows the occurrence, on the different landform units, of Leptosols (presence of continuous hard rock within 25 cm from the soil surface) and soils with the following qualifiers: i) Lithic (continuous hard rock within 10 cm from the soil surface, only used for Leptosols), ii) Leptic (continuous hard rock within 25–100 cm from the soil surface), iii) Epileptic (continuous hard rock within 25–50 cm from the soil surface), and iv) Endoleptic (continuous hard rock within 50–100 cm from the soil surface). Soil profiles having a continuous rock layer within 50 cm from the surface (Epileptic) characterized concave surfaces with slope >5%, and represented 50% of the described profiles. The occurrence of Epileptic soil profiles decreased slightly with increasing slope, with a 39% frequency on slopes >35%. On convex surfaces with slopes >35% more than 50% of the profiles had a continuous rock layer within 25 cm, and the occurrence was >40% in the remaining two slope classes. The corresponding occurrences on concave surfaces were always below 30%, and decreased with increasing slope. These results confirmed, as already noted, the clear influence of slope curvature on soil depth and, consequently, on soil type.

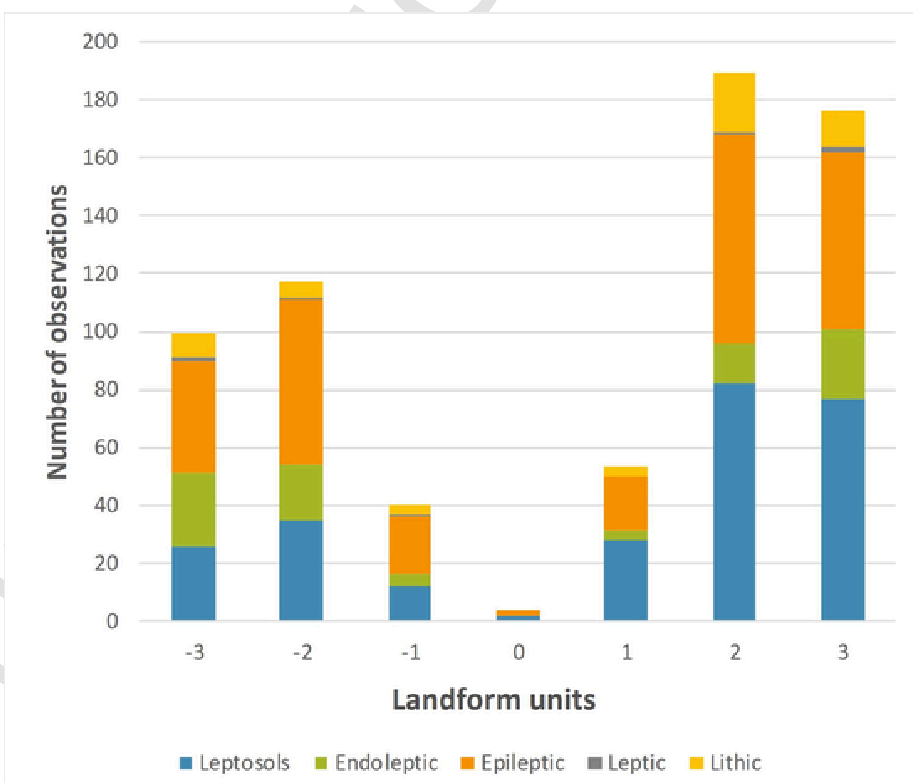

Fig. 4. Occurrence of Leptosols and soils with Lithic, Leptic, Epileptic and Endoleptic qualifiers on the different landform units. For landform unit's description see Table 1.

3.3. The influence of landforms on selected soil properties of functional relevance

We analyzed the influence of landforms on soil thickness and available water capacity (AWC), since both properties have a strong impact on functional aspects (e.g. biomass production and water storage, Fig. 5). In the study areas, soils were generally shallow to moderately deep (avg. depth 54 ± 1 cm) depending on landform and parent material and the main land uses. Soils on convex forms with slopes >15% were significantly ($p < 0.05$) shallower (avg. depth 38 ± 2 cm) than soils on other landform units. Accordingly, they had a lower AWC (about 30 mm), which significantly differed from that of soils on flat or gently sloping surfaces (about 60 mm on average). Soils on concave slopes were generally significantly deeper than those on convex ones with the same gradients, but the differences in AWC were not statistically significant (about 40 mm) as compared to all the other landform units, except that of soils on the plains (average AWC 71.5 ± 2.5 mm). These were significantly deeper (avg. depth 68.1 ± 2.3 cm) as compared to all the other landform units and had higher AWC that did not differ significantly from those on gently rolling areas (mean of about 58 mm for landform units 1 and −1).

These results confirmed the expected trend: soils on plains, where more deposition of material can take place, were deeper than soils on slopes, where soil development was counterbalanced by soil erosion (Scarpone et al., 2016). It was also confirmed that convex forms had a predominantly erosional character, and therefore had shallower soils than concave forms, which were associated with depositional processes and, consequently, deeper soils (Catani et al., 2010). A similar trend applied to AWC, which strongly depended on soil depth.

3.4. Relationships between parent material and soil properties of functional relevance for principal WRB RSGs

Considering only the most relevant WRB RSGs, namely Regosols (N = 382), Leptosols (N = 299) and Cambisols (N = 237), we analyzed four properties of functional relevance in the topsoil horizon, i.e. thickness, texture, pH and organic carbon content, in terms of parent material. A table summarizing the descriptive statistics for the soil properties of each RSG is available in the supplementary materials; since not all soil profiles had analytical data available, the descriptive statistics refer to datasets that were smaller than those presented in Table 2.

As for Regosols, the dataset composition was rather imbalanced, with only four types of parent material adequately represented (AL, DVO, M and Y). The thickness of the topsoil horizon (avg. 22.9 ± 1.7 cm) was significantly greater for the soils developed on alluvial deposits (AL, avg. 34.4 ± 6.5 cm) than for those formed on terrigenous metamorphic rock (M, avg. 23.7 ± 4.1 cm) or on granitic rock (Y, avg. 17.2 ± 1.7 cm). Moreover, it was also significantly thicker on Holocene slope deposits (DVO, avg. 27.2 ± 5.6 cm) than on terrigenous metamorphic rock. Regosols' topsoil sand content (avg. 67.2 ± 2.9%) was significantly lower on metamorphic rock (M, avg. 50.9 ± 7.4%) with respect to sandstone (D, avg. 75.7 ± 4.7%), Holocene slope deposits (DVO, avg. 65.6 ± 9.1%), granitic rock (Y, avg. 71.3 ± 9.1%) and alluvial deposits (AL, avg. 70.0 ± 8.4%). Instead, topsoil clay content (avg. 10.8 ± 1.5%) was significantly higher on metamorphic rock (M, avg. 15.3 ± 3.3%) than on granite (Y, avg. 7.9 ± 1.4%). The same was observed for silt content (avg. 22.1 ± 2.2%), being significantly lower on granite (Y, avg. 20.8 ± 2.2%) and alluvial deposits (AL, avg. 16.8 ± 5.9%) than on metamorphic rock (M, avg. 33.7 ± 6.8%). The same trend was observed in terms of organic carbon content (avg. 2.56 ± 0.6%), with topsoil content being again significantly higher on metamorphic

Fig. 5. Average soil depth and AWC for landform units (vertical bars: standard errors). For landform unit's description see Table 1.

rock (M, avg. 2.92 ± 0.9%) than on granite (Y, avg. 1.89 ± 0.4%) and alluvial deposits (AL, avg. 1.24 ± 0.6%). pH values (avg. 6.64 ± 0.17) were significantly lower on sandstone (D, avg. 6.00 ± 0.44). As expected, higher pH values were related to basic vulcanite (A, 8.40), limestone and marble (K, avg. 7.40 ± 0.99), but the very low number of data on these parent materials (N = 3) did not allow any statistical comparison with data from the others.

Leptosols were numerically relevant only on two types of parent material, i.e. M and Y; topsoil thickness (avg. 14.8 ± 1.7 cm) and organic carbon content (avg 2.88 ± 0.46%) did not differ significantly between the two. However, all three textures were significantly different, with topsoils on granite having a higher sand content (Y, avg. 67.1 ± 2.7%) than those on metamorphic rock (M, avg. 51.8 ± 3.3%), and lower silt (24.1 ± 3.4% vs. 34.2 ± 3.4%) and clay contents (5.8 ± 6.2% vs. 14.0 ± 3.4%). As regards pH values, there were no significant differences between topsoils formed on the two above parent materials (M and Y), with pH values on Y (avg. 6.39 ± 0.19) slightly lower than those on M (avg. 6.65 ± 0.20).

The dataset for Cambisols was numerically more balanced among parent materials. The topsoil horizon was significantly thicker in soils developed on alluvial deposits (AL, avg. 28.7 ± 4.3 cm) than on metamorphic rock (M, avg. 18.4 ± 3.7 cm), Holocene slope deposits (DVO, avg. 17.3 ± 3.1 cm), limestone and marble (K, avg. 15.6 ± 5.5 cm), acid vulcanite (P, avg. 13.8 ± 3.7) and granite (Y, avg. 11.9 ± 1.6 cm). For soil texture, we observed the same trend as that for Leptosols: all three textures were significantly different, with topsoil horizons on granite having a higher sand content (Y, avg. 69.0 ± 3.7%, N = 20) than those on metamorphic rock (M, avg. 46.4 ± 3.1%, N = 12), and lower silt (23.1 ± 3.5% vs. 38.1 ± 3.9%) and clay contents (7.9 ± 1.5% vs. 15.5 ± 1.8%). The two substrates also differed significantly in terms of organic carbon content, which was higher in soils developed on metamorphic rock (M, avg. 4.09 ± 1.1%) than on alluvial deposits (AL, avg. 1.51 ± 0.3%). As for pH, soils developed on granite (Y, avg. 6.61 ± 0.20) and on Holocene slope deposits (DVO, avg. 6.28 ± 0.28) were significantly more acid than those on alluvial deposits (AL, avg. 7.53 ± 0.36).

These results confirmed that, within the same WRB RSG, parent material had a significant effect on topsoil properties (thickness, texture, pH and organic carbon content) that have a strong impact on functional aspects (e.g. physical-hydrological properties and organic carbon stock). As expected, the topsoil horizon thickness was always significantly greater for soils developed on sediments with respect to those developed on hard rocks. The significantly higher clay and silt content and lower sand content in the topsoil horizons developed on M with respect to those on Y may be due to two reasons: i) although most of M in the area was represented by siliceous meta-sandstone, minor outcrops of meta-siltite, meta-pelite and phyllite were also present, and ii) coarser grain size of Y (Mareschal et al., 2015). As regards the organic carbon content, interpretation of results was limited by the important role of other soil forming factors, especially vegetation cover and land use, on this property. In general, lower organic carbon contents characterize parent material whose soils, for their depth and topographic position, are or were used for intensive agriculture (e.g. AL and DC). Detected pH changes agree with previous knowledge on the influence of parent material on this soil property (Reuter et al., 2008; Fabian et al., 2014).

3.5. Predicting soil groups occurrence in landform units

With the aim of providing a statistical model of the relationship among environmental variables, we calibrated stepwise multiple linear regressions (MLR) and general linear models (GLM). The aim of this analysis was to investigate the possibility of estimating the occurrence of the major WRB RSGs as a function of parent material and landform units. This would allow the soil mapping units to be associated to an estimate of their composition in terms of frequency of different soil types.

In the case of MLR, prior to calibration the data (N = 1263) were dummy-coded: as each category of landform unit and parent material was coded (0, 1) to indicate the presence or absence of the variable, this resulted in a set of 19 dummy variables (Table 3). Along with mean values, Table 3 reports the correlation coefficients between WRB RSGs occurrence, landform units and parent material.

The multiple regression coefficients obtained using a forward stepwise procedure are presented in a table provided in the supplementary material, while Table 4 summarizes regression statistics.

We calibrated GLM using both landform units (7 classes) and parent material (12 classes) as categorical predictors; the design had eight dependent variables. Table 4 gives the GLM statistics, while the supplementary materials contain a table with the GLM coefficients. The two approaches provided very similar results, and in terms of error statistics it was not possible to observe significant differences in any WRB RSG, as GLM resulted in an RMSE reduction of between 0.3 (AliLuvLix) and 0.01% (Fluvisols). In both approaches, predictions in terms of root mean square error (RMSE) were more accurate for those WRB RSGs occurring mostly on specific landform or parent material units, or on certain combinations of the two. This was the case, for example, for Arenosols (RMSE = 0.097), Fluvisols (RMSE = 0.147) and the Alisols-Luvisols-Lixisols group (RMSE = 0.102). The opposite was observed for those WRB RSGs which are likely to occur on different parent materials and in different landform units, e.g. Leptosols (RMSE = 0.37), Cambisols (RMSE = 0.37) and Regosols (RMSE = 0.44).

In order to assess the prediction capacity of the training models, we tested them in a sub-area (106.6 km^2) of the Pula pilot area (Fig. 6) characterized by a high degree of soil variability, in terms of both landform units and parent material, with a total of 62 soil mapping units given by the combination of parent material and landform units. Fig. 6 shows the distribution of landform units and parent material in the sub-area. There were 168 classified soil observations representing 27 mapping units for a total of 92.2 km^2 (86.5% of the area); these were the basis for calculating the observed occurrence of the different WRB RSGs in each of the mapping units. Table 5 summarizes the error statistics for each mapping unit in terms of RMSE, and mean absolute error (MAE).

Models showed MAEs between 2.2 and 21.3%, and RMSEs between 4.2 and 34.9%. Absolute average differences were 0.4 and 0.2% for RMSE and MAE, respectively. The two approaches provided very similar results in all mapping units with the exception of two small ones on limestone and marble in convex areas with slope > 35%, where misclassification errors were higher for GLM (RMSE 33.0%, MAE 18.7%) than for MLR predictions (RMSE 19.1%, MAE 11.3%). The opposite was observed in concave areas on the same parent material, but the differences in misclassification with the two methods were 5.1% in terms of MAE and 3.5% in terms of RMSE.

In terms of WRB RSGs, observations in the sub-area had the following distribution: Leptosols 30.6%, Ali-Luvi-Lixisols 19.7%, Cambisols 19.1%, Regosols 16.8%, Phaeozems 10.4%, Umbrisols 1.7%, Fluvisols 1.2% and Arenosols 0.6%. Table 6 reports detailed results in terms of WRB RSGs' occurrence in the most representative mapping units of the test area. In all mapping units, with the exception of AL 0, the higher observed occurrence was reproduced by the models' results, while second and third ranks were often inverted, with the models giving results closer to the observed marginal probabilities in the whole dataset.

Differences in the estimation of the second and third ranks with respect to observations had different origins. In DP 1, the overestimation of Leptosols was related to the presence, in the whole calibration dataset, of observations made in the apex area of the depositional glacis. Such observations were lacking in the pilot area. In DP −1,

Table 3
Descriptive statistics of the dummy-coded variables (N = 1263); significant coefficients are in italics and bold. Mean: mean occurrence fraction. LFU, landform unit: for description see Table 1. Parent material: for description see Table 2.

	WRB RSG	*AliLuvLix*	*Arenosols*	*Cambisols*	*Fluvisols*	*Leptosols*	*Phaeozems*	*Regosols*	*Umbrisols*
	Mean	*0.131*	*0.011*	*0.191*	*0.030*	*0.249*	*0.042*	*0.321*	*0.025*
		Correlations							
	Mean								
Landform unit									
LFU −3	0.159	***−0.060***	−0.046	***0.072***	***−0.077***	***−0.081***	−0.026	***0.081***	***0.081***
LFU −2	0.157	***−0.077***	−0.025	***0.075***	***−0.076***	−0.047	0.018	0.053	0.027
LFU −1	0.109	***0.149***	0.011	***−0.066***	***0.057***	***−0.108***	0.015	0.025	−0.024
LFU 0	0.138	***0.130***	***0.177***	***0.059***	***0.320***	***−0.220***	−0.038	***−0.059***	***−0.064***
LFU 1	0.092	***0.201***	−0.007	***−0.084***	−0.040	0.013	−0.026	***−0.060***	0.001
LFU 2	0.186	***−0.138***	−0.051	−0.028	***−0.084***	***0.204***	0.001	−0.024	0.001
LFU 3	0.159	***−0.124***	−0.046	−0.050	***−0.077***	***0.199***	0.049	−0.026	−0.029
Parent material									
A	0.014	−0.047	−0.013	−0.007	−0.021	0.023	***0.141***	−0.026	−0.019
AL	0.109	***−0.136***	***0.084***	***0.071***	***0.503***	***−0.184***	***−0.061***	0.052	***−0.056***
D	0.019	−0.003	***0.318***	−0.052	−0.025	−0.027	−0.029	0.028	−0.022
DP	0.091	***0.765***	−0.034	***−0.125***	***−0.056***	***−0.182***	−0.053	***−0.212***	−0.051
DC	0.040	***−0.080***	0.017	***0.086***	−0.036	***−0.081***	***0.158***	0.014	−0.033
DVO	0.122	***−0.138***	−0.017	***0.189***	***−0.066***	***−0.204***	−0.018	***0.094***	***0.216***
DVP	0.042	***0.515***	−0.022	***−0.081***	−0.037	***−0.121***	−0.044	***−0.144***	−0.034
F	0.021	−0.053	−0.014	***0.146***	−0.024	−0.010	−0.029	−0.043	−0.022
K	0.016	−0.049	−0.013	0.020	−0.022	0.029	***0.100***	−0.033	−0.020
M	0.200	***−0.183***	−0.053	***−0.060***	***−0.088***	***0.379***	0.014	***−0.128***	−0.005
P	0.002	−0.015	−0.004	−0.019	−0.007	0.023	−0.008	0.015	−0.006
Y	0.324	***−0.270***	***−0.057***	***−0.085***	***−0.122***	***0.136***	−0.019	***0.214***	−0.026

Table 4
Error statistics for MLR and GLM estimates (ME, mean error; RMSE, root mean square error; MAE, mean absolute error).

WRB RSG	Multiple R	Multiple R^2	Adjusted R^2	p	ME	RMSE	MAE
MLR							
Arenosols	0.366	0.134	0.129	0.00	−0.0020	0.0973	0.0210
Fluvisols	0.504	0.254	0.250	0.00	−0.0018	0.1475	0.0453
AliLuvLix	0.953	0.909	0.908	0.00	−0.0009	0.1021	0.0217
Regosols	0.359	0.129	0.121	0.00	−0.0024	0.4358	0.3824
Cambisols	0.324	0.105	0.097	0.00	−0.0007	0.3706	0.2753
Leptosols	0.516	0.266	0.259	0.00	−0.0053	0.3703	0.2799
Phaeozems	0.255	0.065	0.060	0.00	0.0000	0.1939	0.0752
Umbrisols	0.227	0.051	0.048	0.00	−0.0007	0.1530	0.0476
GLM							
Arenosols	0.367	0.135	0.122	0.00	−0.0027	0.0972	0.0216
Fluvisols	0.504	0.254	0.244	0.00	−0.0019	0.1474	0.0455
AliLuvLix	0.953	0.909	0.907	0.00	−0.0016	0.1020	0.0225
Regosols	0.360	0.130	0.117	0.00	−0.0023	0.4355	0.3819
Cambisols	0.331	0.109	0.096	0.00	−0.0007	0.3696	0.2740
Leptosols	0.517	0.267	0.257	0.00	−0.0051	0.3699	0.2792
Phaeozems	0.257	0.066	0.053	0.00	−0.0005	0.1937	0.0756
Umbrisols	0.228	0.052	0.038	0.00	−0.0007	0.1530	0.0475

the overestimation of Regosols was related to the presence, in the calibration dataset, of observations made in small depressions filled with new sediments. These observations were lacking in the pilot area. In M 3 and M −3 the overestimation of Regosols and underestimation of Cambisols could have been due to other soil forming factors than those considered, such as a denser wood cover in the pilot area with respect to the whole study area. This could also explain the underestimation of Umbrisols and overestimation of Cambisols in Y 3 and Y −3. In M 3 and M −3, Ali-Luv-Lixisols, formed in small Pleistocene slope deposits within the terrigenous metamorphic rocks, were more frequent in the pilot area. As for AL 0, MLR overestimated the presence of Regosols and underestimated the Cambisols.

Fig. 7 a-d shows maps of the MLR estimated occurrences for the four most widespread WRB SRGs in the area, i.e. Regosols, Leptosols, Cambisols and Ali-Luv-Lixisols. The patterns were clearly distinct and coherent with the relationships observed between soil groups and specific combination of terrain attributes and parent materials. The Ali-Luv-Lixisols occurred on depositional glacis and fluvial terraces on Pleistocene deposits (DP). Cambisols were mostly estimated on Holocene alluvial deposits (AL) and Holocene colluvial deposits

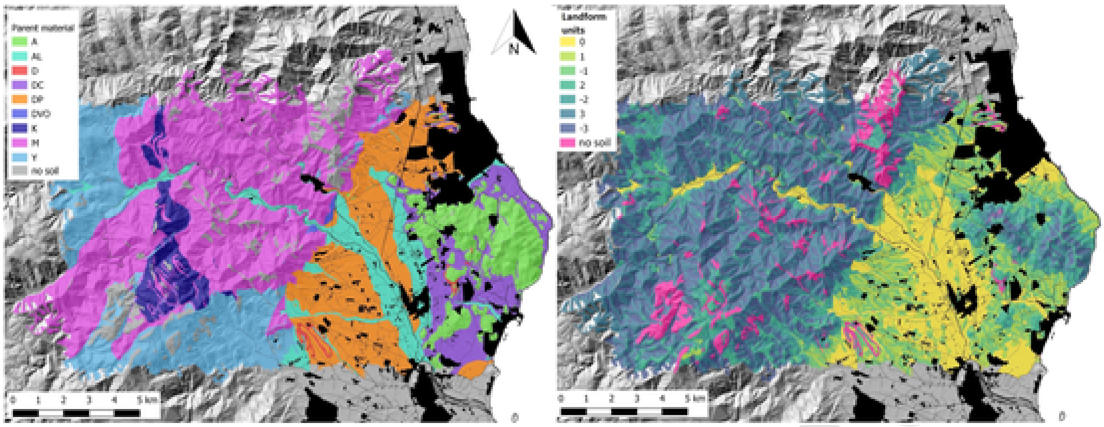

Fig. 6. Occurrence of parent material (left) and landform units (right) in a sub-area of the test area Pula. For landform unit's description see Table 1. A, basic vulcanites; AL, Alluvial deposits (Holocene); D, sandstones; DP, Pleistocene deposits; DC, Colluvial deposits; DVO, Holocene Slope deposits; K, limestones; M, Terrigenous metamorphic rocks; Y, Granites.

Table 5
Error statistics (RMSE, root mean square error; MAE, mean absolute error) for the mapping units (MU) in the Pula validation sub-area.

MU	Num	Area (km ^{2})	RMSE_GLM	RMSE_MLR	MAE_GLM	MAE_MLR
AL 0	7	7.6	13.40%	16.20%	11.50%	11.00%
DP 0	7	3.7	5.90%	5.40%	3.50%	3.10%
DP 1	7	5.3	12.50%	12.10%	7.10%	7.10%
DP −1	8	4.2	4.20%	4.20%	2.30%	2.20%
DP −2	3	0.7	29.20%	29.10%	16.10%	16.10%
DC 1	7	1.8	10.50%	10.20%	8.50%	8.30%
DC −1	9	3.2	21.60%	22.10%	15.60%	16.00%
DC 2	3	0.1	9.80%	9.30%	6.00%	5.80%
A 2	6	2.5	9.70%	9.70%	5.60%	5.70%
A −2	4	1.7	22.30%	23.60%	13.10%	13.80%
A 3	4	1.5	16.70%	16.10%	9.90%	9.60%
A −3	2	0.8	33.10%	34.90%	20.30%	21.30%
M 2	6	1.9	24.82%	24.81%	16.17%	16.07%
M −2	4	1.6	12.10%	12.20%	7.80%	7.50%
M 3	36	20.3	9.14%	8.76%	5.42%	5.38%
M −3	27	14.5	12.22%	12.27%	8.02%	8.03%
K 3	3	2	33.00%	19.10%	18.70%	11.30%
K −3	4	1.7	16.20%	19.70%	8.80%	13.90%
Y −1	3	0.6	16.00%	15.20%	11.00%	10.30%
Y 2	3	3.2	11.40%	11.60%	6.90%	7.00%
Y −2	3	2.6	16.30%	16.80%	11.60%	11.80%
Y 3	6	5.1	13.10%	13.40%	9.70%	9.80%
Y −3	6	5.7	13.10%	12.90%	8.00%	7.90%
All	168	92.2	17.82%	17.39%	10.05%	9.94%

(DC). They also occurred, in the hilly sector and, with a lower estimate, on terrigenous metamorphic rocks (M), limestone and marble (K), basic vulcanite (A) and granite (Y). Leptosols were estimated in the hilly sector and mostly occurred on terrigenous metamorphic rock (M) and, with a lower estimate, on limestone and marble (K), basic vulcanite (A) and granite (Y). Regosols were estimated in both the hilly and plain sectors. They mostly occurred on Holocene alluvial deposits (AL) and sandstone (D) and, with lower occurrence, on granite (Y). They were also estimated, with lower values, on basic vulcanite (A), Holocene colluvial deposits (DC), limestone and marble (K), and terrigenous metamorphic rock (M).

In general, both MLRs and GLMs led to poor results whenever soil-classes overlapped in the feature space, or if the correlation between soil-classes and predictors was low. This agreed with the findings of Hengl et al. (2007) who compared four semi-automated interpolation methods to produce soil-class maps from 4125 profile observations in Iran using multiple auxiliary predictors such as terrain parameters and remote sensing indices. In our study, the data had a density of 2 observations per km^2, which was coherent with a cartographic scale of 1:100,000. In this case, soil units were necessarily associations or complexes and mapping units identified pedolandscapes, which could not be detailed any further without increasing sampling density in the field or trying to include more predictors in the model. For example, Grinand et al. (2008) used a classification tree analysis to predict soil distribution in unvisited locations using terrain factors, spectral reflectance as derived from LANDSAT ETM imagery, land cover and lithology maps as predictors at regional scale.

Regardless of sampling density, the predictive capacity of training models remained quite low when extrapolated to an independent validation area. At national scale, Lorenzetti et al. (2015) assessed the class frequencies for WRB RSGs on the 1:5,000,000 map of Italian soil regions using different data mining techniques and 10 predictors, and found that the most reliable approaches were better at predicting the absence rather than presence of a given soil type. In the case study presented in this work, we tested at a finer scale the possibility of defining the class frequencies for WRB RSGs for land units automatically derived with a semi-automatic geomatic approach. Our results showed that both models were more reliable in predicting the absence rather than presence of a given soil type, with an average agreement with observed non-occurrences in the test area equal to 54 and 61%, respectively, for GLMs and MLRs.

4. Conclusions

In the two pilot areas of Pula and Muravera, the distribution of soil types varied with landform and parent material. The relationships between soil types and landforms reflected the influence of slope gradi-

Table 6
Observed (OBS) and GLM and MLR estimated relative percentage of the WRB RSGs in some of the mapping units (MU) of the test area.

MU	N	WRB	AliLuvLix	Arenosols	Cambisols	Fluvisols	Leptosols	Phaeozems	Regosols	Umbrisols
AL 0	7	**OBS**	0.00	0.00	0.43	0.29	0.00	0.00	0.29	0.00
		GLM	0.16	0.08	0.28	0.04	0.14	0.07	0.22	0.02
		MLR	0.00	0.12	0.24	0.04	0.00	0.01	0.60	0.00
DP 1	7	**OBS**	0.71	0.00	0.29	0.00	0.00	0.00	0.00	0.00
		GLM	0.90	0.00	0.00	0.00	0.09	0.00	0.00	0.01
		MLR	0.88	0.00	0.00	0.00	0.10	0.00	0.00	0.01
DP −1	8	**OBS**	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		GLM	0.91	0.00	0.00	0.01	0.00	0.01	0.08	0.00
		MLR	0.91	0.00	0.00	0.01	0.00	0.01	0.08	0.00
M 3	36	**OBS**	0.08	0.00	0.17	0.00	0.67	0.00	0.08	0.00
		GLM	0.01	0.00	0.13	0.00	0.61	0.06	0.18	0.02
		MLR	0.01	0.00	0.14	0.00	0.60	0.06	0.17	0.02
M −3	27	**OBS**	0.19	0.00	0.26	0.00	0.37	0.07	0.11	0.00
		GLM	0.01	0.00	0.16	0.00	0.49	0.03	0.29	0.02
		MLR	0.01	0.00	0.15	0.00	0.50	0.04	0.29	0.02
Y 3	6	**OBS**	0.00	0.00	0.00	0.00	0.17	0.00	0.67	0.17
		GLM	0.00	0.00	0.13	0.00	0.37	0.06	0.44	0.01
		MLR	0.00	0.00	0.14	0.00	0.36	0.05	0.42	0.02
Y −3	6	**OBS**	0.00	0.00	0.00	0.00	0.17	0.00	0.50	0.33
		GLM	0.00	0.00	0.16	0.00	0.26	0.02	0.55	0.02
		MLR	0.00	0.00	0.15	0.00	0.26	0.03	0.55	0.02

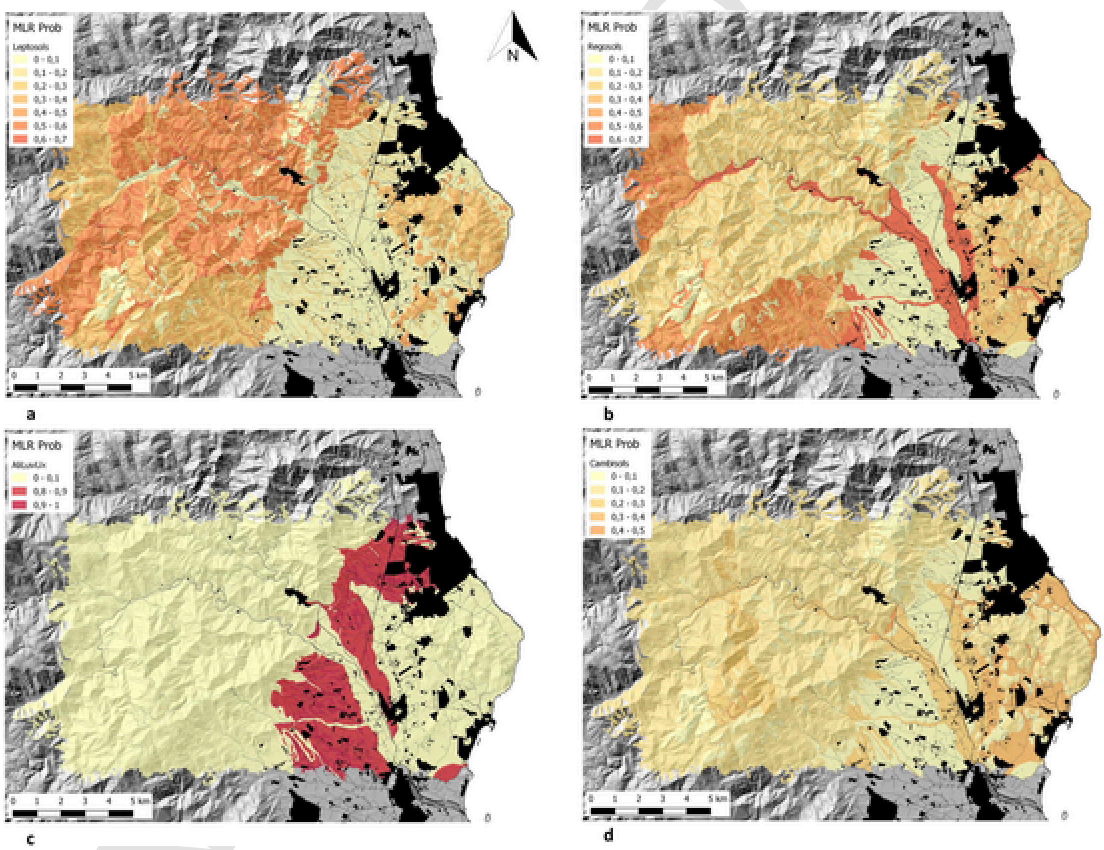

Fig. 7. MLR estimates maps of occurrence for WRB RSGs: a) Leptosols; b) Regosols; c) Ali-Luv-Lixisols; d) Cambisols.

ent and curvature. On steeper slopes, due to morphodynamic processes, mainly very weakly (Leptosols and Regosols) and weakly (Cambisols) developed soils were present. Leptosols (shallow soils) were more frequent where a predominant erosional character prevailed (convex slopes), whilst Regosols and Cambisols (deeper soils) occurred more frequently where depositional processes predominated (concave slopes). Many different soil types may be present in flat areas, mainly depending on the type and age of parent material. In these areas, Regosols and Cambisols prevailed on Holocene deposits, while more developed soils (mainly Luvisols) largely prevailed on Pleistocene deposits. On hard rocks, soil development was generally very weak and Leptosols were very frequent on terrigenous metamorphic rock and frequent on granite. In addition to Leptosols, Regosols were more widespread than Cambisols on these parent materials.

Landforms strongly influence soil depth and available water capacity. Indeed, soils on plains are deeper than soils on slopes, where convex terrain forms have shallower soils than concave terrain forms. A similar trend applied to the available water capacity, which strongly depended on soil depth. In Regosols, Leptosols and Cambisols, which were the most common WRB Reference Soil Groups in the two pilot areas, the parent material had a significant effect on topsoil properties (thickness, texture, pH and organic carbon content) of soils belonging to the same WRB RSG. Consequently, since these properties have a strong impact on functional aspects (e.g. physical-hydrological properties and organic carbon stock), it is essential to differentiate these soils according to the parent material.

The orders of occurrence of the different WRB RSGs in the map units, defined with the geomatic approach, were very similar for the two approaches tested. The patterns obtained were in both cases coherent with the relationships observed between soil groups and specific combination of terrain attributes and parent materials, with higher accuracy for those units where WRB RSGs were more tightly associated with specific combinations of the two pedogenesis factors considered. In such cases, the probability of occurrence of specific WRB RSGs was estimated with a greater reliability, while in the presence of more complex response patterns, according to our results, both models were more reliable in predicting the absence rather than the presence of a given soil type.

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

We are grateful to three anonymous reviewers, whose critical comments and suggestions greatly helped to improve the quality of the manuscript. We are also grateful to Agris Sardegna for the availability of soil data collected in the Muravera pilot area. We also thank Dr. Alison Garside for her revision and editing of the English language. This work was supported by the RAS under the decree DGR n. 56/36 dated 29.12.2009.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.catena.2020.104818.

References

Arangino, F., Aru, A., Baldaccini, P., Vacca, S., 1986. I suoli delle aree irrigabili della Sardegna, in scala 1:100.000. Assessorato Regionale alla Programmazione Bilancio ed Assetto del Territorio, Ente Autonomo del Flumendosa, Cagliari.

Arrouays, D, McBratney, A, Bouma, J, Libohova, Z, Richer-de-Forges, A C, Morgan, C L S, Roudier, P, Poggio, L, Mulder, V L, 2020. Impressions of digital soil maps: The good, the not so good, and making them ever better. Geoderma Regional 20, e00255. doi:10.1016/j.geodrs.2020.e00255.

Aru, A., Baldaccini, P., Delogu, G., Dessena, M.A., Madrau, S., Melis, R.T., Vacca, A., Vacca, S., 1990. Carta dei suoli della Sardegna, in scala 1:250.000. Dipartimento Scienze della Terra Università di Cagliari, Assessorato Regionale alla Programmazione Bilancio ed Assetto del Territorio, SELCA, Firenze.

Barca, S., Serri, R., Rizzo, R., Forci, A., Calzia, P., Pertusati, P.C. (Eds.), 2009. Note illustrative della Carta Geologica d'Italia alla scala 1:50.000. Foglio 565 Capoterra. ISPRA Servizio Geologico d'Italia, Regione Autonoma della Sardegna, A.T.I. S.EL.CA L.A.C. SystemCart, Roma, 208 pp.

Birkeland, P W, 1999. Soils and Geomorphology. third ed. Oxford University Press, New York.

Brady, N.C., Weil, R.R., 2008. The Nature and Properties of Soils, fourteenth ed. Prentice Hall/Pearson Education, Upper Saddle River, New Jersey.

Brungard, C W, Boettinger, J L, Duniway, M C, Wills, S A, Edwards, T C, 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239, 68–83. doi:10.1016/j.geoderma.2014.09.019.

Buol, S W, Southard, R J, Graham, R C, McDaniel, P A, 2011. Soil Genesis and Classification. sixth ed. John Wiley, Chichester, UK.

Carboni, S, Palomba, M, Vacca, A, Carboni, G, 2006. Paleosols provide sedimentation, relative age, and climatic information about the alluvial fan of the River Tirso (Central-Western Sardinia, Italy). Quat. Int. 156–157, 79–96. doi:10.1016/j.quaint.2006.05.005.

Carmignani, L, Conti, P, Pertusati, P C, Barca, S, Cerbai, N, Eltrudis, A, Funedda, A, Oggiano, G, Patta, E D, Ulzega, A, Orrù, P (Eds.), 2001. Note illustrative della Carta Geologica d'Italia alla scala 1:50.000. Foglio 549 Muravera. Servizio Geologico d'Italia, Regione Autonoma della Sardegna, Istituto Poligrafico e Zecca dello Stato, Roma, p. 140.

Catani, F, Segoni, S, Falorni, G, 2010. An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. Water Resour. Res. 46, W05508. doi:10.1029/2008WR007450.

Edward, E, Walter, W, Kevin, S, Susan, D, Linda, J, Christman, M, West, M, Kramer, M, 2012. Analysis of generalized linear mixed models in the agricultural and natural resources sciences. Soil Science Society of Agronomy, Madison, DC.

Evans, D M, Hartemink, A E, 2014. Digital soil mapping of a red clay subsoil covered by loess. Geoderma 230, 296–304. doi:10.1016/j.geoderma.2014.03.013.

Fabian, C., Reimann, C., Fabian, K., Birke, M., Baritz, R., Haslinger, E., The GEMAS Project Team, 2014. GEMAS: Spatial distribution of the pH of European agricultural and grazing land soil. Applied Geochemistry 48, 207-216. https://doi.org/10.1016/j.apgeochem.2014.07.017.

Finke, P., Hartwich, R., Dudal, R., Ibàñez, J., Jamagne, N., King, D., Montanarella, R., Yassoglou, N., 2001. Georeferenced Soil Database for Europe. Manual of Procedures, Version 1.1. European Soil Bureau Scientific Committee, Italy. 166 pp.

Graham, R C, Anderson, M A, Sternberg, P D, Tice, K R, Schoeneberger, P J, 1997. Morphology, porosity, and hydraulic conductivity of weathered granitic bedrock and overlying soils. Soil Sci. Soc. Am. J. 61, 516–522. doi:10.2136/sssaj1997.03615995006100020021x.

Gray, J M, Humphreys, G S, Deckers, J A, 2009. Relationships in soil distribution as revealed by a global soil database. Geoderma 150, 309–323. doi:10.1016/j.geoderma.2009.02.012.

Grealish, G.J., Fitzpatrick, R.W., Hutson, J.L., 2015. Soil survey data rescued by means of user friendly soil identification keys and toposequence models to deliver soil information for improved land management. Geo. Res. J. 6, 91-81. https://doi.org/10.1016/j.grj.2015.02.006.

Grinand, C, Arrouays, D, Laroche, B, Martin, M P, 2008. Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190. doi:10.1016/j.geoderma.2007.11.004.

Grunwald, S, 2006. What do we really know about the space-time continuum of soil-landscapes. In: Grunwald, S (Ed.), Environmental Soil-Landscape Modeling - Geographic Information Technologies and Pedometrics. CRC Press, New York, pp. 3–36.

Grunwald, S, 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152, 195–207. doi:10.1016/j.geoderma.2009.06.003.

Hengl, T, Toormanian, N, Reuter, H I, Malakouti, M J, 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma 140, 417–427. doi:10.1016/j.geoderma.2007.04.022.

Hudson, B D, 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56, 836–841. doi:10.2136/sssaj1992.03615995005600030027x.

Huggett, R, Cheesman, J, 2002. Topography and the Environment. Prentice Hall, London.

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 Reports No. 106. FAO, Rome.

Kempen, B, Brus, D J, Stoorvogel, J J, Heuvelink, G B M, de Vries, F, 2012. Efficiency comparison of conventional and digital soil mapping for updating soil maps. Soil Sci. Soc. Am. J. 76, 2097–2115. doi:10.2136/sssaj2011.0424.

Lane, P W, 2002. Generalized linear models in soil science. Eur. J. Soil Sci. 53, 241–251. doi:10.1046/j.1365-2389.2002.00440.x.

McBratney, A B, Mendonça Santos, M L, Minasny, B, 2003. On digital soil mapping. Geoderma 117, 3–52. doi:10.1016/S0016-7061(03)00223-4.

MacMilland, R.A., Shary, P.A., 2009. Landforms and landform elements in geomorphometry. In: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Develop. Soil Sci. 33, 227-254.

Madrau, S, Deroma, M, Loj, G, Baldaccini, P, 2006. Carta Ecopedologica della Sardegna, in scala 1:250.000. Dipartimento di Ingegneria del Territorio. Università degli Studi di Sassari, Sassari.

Mareschal, L, Turpault, M P, Ranger, J, 2015. Effect of granite crystal grain size on soil properties and pedogenic processes along a lithosequence. Geoderma 249–250, 12–20. doi:10.1016/j.geoderma.2015.02.009.

Marrone, V A, Vacca, A, Vecchio, G, Apollaro, C, Caruso, A M, Scarciglia, F, Aru, A, De Rosa, R, 2008. Pedopaesaggi e suoli del foglio IGMI 549-Muravera (Sardegna sud-orientale). In: Ottonello, G (Ed.), GEOBASI e Il Foglio IGMI N 549-Muravera. Pacini Editore, Pisa, pp. 59–85.

Minasny, B, McBratney, A B, 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. doi:10.1016/j.geoderma.2015.07.017.

Mosleh, Z, Salehi, M H, Jafari, A, Borujeni, I E, Mehnatkesh, A, 2016. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environ. Monit. Assess. 188, 195. doi:10.1007/s10661-016-5204-8.

Nelder, J, Wedderburn, R, 1972. Generalized linear models. J. Royal Statistical Soc.Ser. A (General) 135, 370–384. doi:10.2307/2344614.

RAS (Regione Autonoma della Sardegna), 2006. Piano Paesaggistico Regionale: Norme Tecniche di Attuazione e Allegati. Settembre 2006, Available on the Internet at: http://www.sardegnaterritorio.it/documenti/6_83_20060907193936.pdf (accessed 19.05.2020).

RAS (Regione Autonoma della Sardegna), 2008. Piano Paesaggistico Regionale: Linee Guida per l'adeguamento dei Piani Urbanistici Comunali al PPR e al PAI. In: Prima Fase e Il Riordino delle Conoscenze. Assetto Ambientale. Luglio 2008, Available on the Internet at: http://www.sardegnaterritorio.it/documenti/6_83_20081007095546.zip (accessed 19.05.2020).

RAS (Regione Autonoma della Sardegna), 2010. Carta geologica - Elementi areali. Available at: http://www.sardegnageoportale.it/webgis2/sardegnamappe/?map=base (accessed 19.05.2020).

Reuter, H I, Rodriguez Lado, L, Hengl, T, Montanarella, L, 2008. Continental-scale digital soil mapping using European soil profile data: soil pH. In: Böhner, J, Blaschke, T, Montanarella, L (Eds.), SAGA - Seconds Out, Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, Heft 19. Universität Hamburg, Institut für Geographie, pp. 91–102.

Saxton, K E, Rawls, W, Romberger, J S, Papendick, R I, 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50 (4), 1031–1036. doi:10.2136/sssaj1986.03615995005000040039x.

Scarciglia, F, Tuccimei, P, Vacca, A, Barca, D, Pulice, I, Salzano, R, Soligo, M, 2011. Soil genesis, morphodynamic processes and chronological implications in two soil transects of SE Sardinia, Italy: Traditional pedological study coupled with laser ablation ICP-MS and radionuclide analyses. Geoderma 162, 39–64. doi:10.1016/j.geoderma.2011.01.004.

Scarpone, C, Schmidt, M G, Bulmer, C E, Knudby, A, 2016. Modelling soil thickness in the critical zone for Southern British Columbia. Geoderma 282, 59–69. doi:10.1016/j.geoderma.2016.07.012.

Shary, P A, Sharaya, L S, Mitusov, A V, 2002. Fundamental quantitative methods of land surface analysis. Geoderma 107, 1–32. doi:10.1016/S0016-7061(01)00136-7.

Sierra, M, Martínez, F J, Sierra, C, Aguila, J, 2009. Correlations between pedological parameters in relation to lithology and soil type in Almería (SE Spain). J. Arid Environ. 73 (4–5), 493–498. doi:10.1016/j.jaridenv.2008.10.014.

Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed. USDA-Natural Resources Conservation Service, Washington, DC. 360 pp.

Stoorvogel, J J, Kempen, B, Heuvelink, G B M, de Bruin, S, 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149, 161–170. doi:10.1016/j.geoderma.2008.11.039.

Vacca, A, Loddo, S, Melis, M T, Funedda, A, Puddu, R, Verona, M, Fanni, S, Fantola, F, Madrau, S, Marrone, V A, Serra, G, Tore, C, Manca, D, Pasci, S, Puddu, M R, Schirru, P, 2014. A GIS based method for soil mapping in Sardinia, Italy: A geomatic approach. J. Environ. Manag. 138, 87–96. doi:10.1016/j.jenvman.2013.11.018.

van Delden, H, van Vliet, J, Rutledge, D T, Kirkby, M J, 2011. Comparison of scale and scaling issues in integrated land-use models for policy support. Agric. Ecosyst. Environ. 142, 18–28. doi:10.1016/j.agee.2011.03.005.

Yaalon, D.H., 1989. The Earliest Soil Maps and Their Logic. Bulletin of the International Society of Soil Science. International Society of Soil Science, Wageningen p. 24.

Zhang, G L, Liu, F, Song, X D, 2017. Recent progress and future prospect of digital soil mapping: a review. J. Integr. Agric. 16, 2871–2885. doi:10.1016/S2095-3119(17)61762-3.

Zevenbergen, L W, Thorne, C R, 1987. Quantitative analysis of the land surface topography. Earth Surf. Process. Landf. 12, 47–56. doi:10.1002/esp.3290120107.

Zeraatpisheh, M, Ayoubi, S, Jafari, A, Finke, P, 2017. Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. Geomorphology 285, 186–204. doi:10.1016/j.geomorph.2017.02.015.