Boletín de la Asociación de Geógrafos Españoles N.º 75 - 2017, págs. 605-611

I.S.S.N.: 0212-9426

INFLUENCE OF ENVIRONMENTAL VARIABLES ON ESTIMATING SOIL ORGANIC MATTER AT REGIONAL SCALE ON A SEMIARID ENVIRONMENT. (REGION OF MURCIA, SPAIN)

Pedro Pérez-Cutillas[1, 2]
Gonzalo G. Barberá[2]
Carmelo Conesa García[1]
[1]Departament of Geography, University of Murcia. pedrope@um.es, cconesa@um.es
[2] Department of Soil and Water Conservation CSIC-CEBAS. perezcutillas@cebas.csic.es, gbarbera@cebas.csic.es

I. INTRODUCTION

The soil organic matter (SOM) conditions substantially the basic properties of soils and, particularly, important hydrological and geomorphological functions, since it favors the aggregation of the soil. It allows the formation of clay-humic complexes which act as the core of soil aggregates (Lal et al., 1994), increasing their porosity and, thus, the infiltration and percolation of fluids. Consequently, it increases the water retention capacity of the soil (Brady, 1984), while the risk of run-off and erosion decreases (Van Beers, 1980). In turn, the erosive processes directly affect the mobilization of carbon stocks in the soil (Martínez-Mena et al., 2012; Nadeu et al., 2012; Boix-Fayos et al., 2015). Moreover, the behavior of the organic carbon (OC) present as SOM is especially relevant in relation to its management and treatment, and its influence on ecosystem services (MEA 2005). In connection with the study of the global carbon cycle and its implications for global warming (Jones et al., 2005), there is an extensive scientific literature. Several papers published in this respect consider knowledge of the distribution of OC as essential and coincide in affirming that the carbon stocks in soil are greater than those of the atmosphere and the biosphere as a whole (Grace, 2004).

Estimation of the spatial distribution of the SOM could contribute to decision making in the management of these resources and minimize the harmful effects on the environment. In order to progress to more-detailed soil mapping, in the last few years different techniques to predict the value of a given soil variable over large areas (from now on, regional) have been developed. In the early 1980s, the estimation of spatially distributed soil variables by applying kriging and co-kriging techniques (McBratney et al., 1981) or Geographically weighted regressions (GWR) (Wang et al., 2013; Wang et al., 2014; Zeng et al., 2016), based primarily on the spatial dependence of the study variables, was addressed. The usefulness of these techniques is greater for not-so-large areas and quite intensive and regular sampling (Chen et al., 2000; Vašát et al., 2012). In an estimation at the regional level -understood as a large area of thousands of square kilometers- with widely-dispersed samples, the information provided by the possible spatial autocorrelation between samples can be much reduced (Burrough et al., 1997; Western et al., 2004).

Since the density of the samples is often not so high at the regional scale, the use of other methods to predict the values of the soil variables in a simple and affordable way is required (Ließ et al., 2012). A particularly-effective alternative is to model the relationship between the soil variable of interest and environmental variables for which the spatially-distributed information is available (Mckenzie & Ryan 1999; Thompson et al., 2001; McBratney et al., 2003; Bou Kheir et al., 2010).

The importance of this relationship in the digital elaboration of soil mapping has already been reviewed by several authors, by the use of statistical methods applied to the spatial prediction of soil properties (Minasny et al., 2008; Greve et al., 2012). The results are cartographic products with continuous information obtained by statistical modeling of the soil properties as a function of the environmental variables which influence their development and spatial distribution.

Moreover, remote sensing techniques provide massive amounts of georeferenced quantitative data, allowing an overview of large areas and spectral values which show good correlations with the areas covered by different land uses, as well as having a very dynamic update capacity (Ben-Dor, 2002). The evaluation, characterization, and determination of soil properties by means of the use of remote sensing data have been widely applied in recent years (Rawls et al., 2004; Vrieling, 2006; Lagacherie et al., 2012; Poggio et al., 2013; Mirzaee et al, 2016). The capacity of spectrometry, under laboratory conditions, to predict important soil properties have been shown (Viscarra-Rossel et al., 2006; Ben-Dor et al., 2009; Gholizadeh et al., *2013*). Nevertheless, its direct application to the Earth's surface is not without certain constraints and shortcomings, which make it advisable, in this case, to focus it use on obtaining specific indices related to environmental variables.

In this work, the ability of GIS tools to manage large volumes of properly georeferenced environmental information (field and remote sensing data) is exploited. By means of GIS, each sample of the database is associated with the environmental variables corresponding to the sampling site. Linking the SOM values to the aforementioned environmental variables, used as predictors, the regression model is built in order to predict the SOM content. Finally, the model is integrated into the GIS using map algebra to obtain an estimate of SOM at each point of the region. This model allows one to suggest the main environmental processes that, at this scale, control the SOM content.

II. STUDY AREA AND ENVIRONMENTAL CONDITIONS

The study area is located in the Murcia Region (11313 km^2, INE, 2013), within the Segura river basin (18208 km^2, CHS, 2013) in Southeast Spain (0º 21" to 2º 54" W; 37º 14" to 39º 07"

N) (Figure 1). Overall, this land is somewhat uneven due to the presence of numerous mountain ranges belonging to the Cordillera Bética mountain range, lined up in an ENE-WSW direction, often at altitudes above 1,000 m. Here, a collection of valleys, basins, plains, and plateaus together configure a contrasting topography and a singular territory with a wide variety of landscapes.

The study area has a semi-arid Mediterranean climate (Capel-Molina 2000). The average rainfall of the basin is around 375 mm/year, with annual average values of 472 mm at the headwaters and 317 mm in the area closer to the sea, and there is high seasonal and inter-annual rainfall irregularity, prolonged droughts, and often torrential rains.

This entire region receives a strong annual insolation, with an average of 2,800 hours of sunshine per year, peaking in July (340 h) and with a minimum in December (160 h) - according to the Alcantarilla Observatory, close to the city of Murcia (Alonso-Sarría 2007). Due to the rugged regional relief, the local variations in the radiation received are notable. Only in the plains (coastal plains, pre-littoral depressions, and high plateaus) do the contrasts decrease, so that the homogeneity is greater.

Desertification and soil loss due to erosion are two of the most-important causes of soil degradation in this area, which results in a loss of the main physical, chemical, and biological components (Boix-Fayos et al., 2005). According to the National Soil Erosion Inventory, prepared by the Ministry of the Environment (MMA 2002-2012), the average soil erosion in the Segura river basin is around the Spanish average, 24.53 ton ha^{-1} yr^{-1}, equivalent to a loss of about 2 mm depth per year. In this basin there is a great diversity of soils, produced by the action of various processes influenced by multiple environmental factors. Among them, the weather conditions, lithology, and relief, which largely condition the erosion and the soil processes, stand out (Alvarez Rogel et al., 2001).

III. MATERIALS AND METHODS

III.1. Analytical data from the LUCDEME project

All the empirical information on the SOM content in the soils was obtained from data of the Combat Desertification in the Mediterranean project (LUCDEME) (ICONA 1986), developed in the 1980s and early 1990s. From this project, a database was obtained from a series of field samplings corresponding to analytical determinations in the arable soil; that is, the topsoil (0-20 cm approximately). A set of 1126 georeferenced samples are distributed according to a grid size of 3 x 3 km, in both agricultural soils and uncultivated land (Fig. 1).

III.2. Environmental variables used in the modeling

In a strict sense, not all the variables used in the modeling processes of this work may be called directly-obtained environmental variables. Some of them are indices developed from spectral signals of remote sensing - which can be related only indirectly to soil characteristics. Nevertheless, this denomination is kept in order to unify all the information processed in the GIS analyses and statistical treatments. These variables have been structured in a series of data sets, which are described in the following blocks of variables (the sequence number of each variable is shown in bold and in brackets):

III.2.1. Topographical and hydrological variables

These variables have been obtained from the Digital Elevation Model (DEM) (**1**) of sensor ASTER (NASA - METI 2013). From DEM were obtained the slopes (**2**), (Burrough & Mcdonnell 1998) and curvature values (Moore et al., 1991).The curvature can be broken down into the profile curvature (**3**), and the perpendicular curvature (**4**) (Zeverbergen & Thorne 1987). The flow accumulation (**5**) represents the number of pixels of the watershed to a particular pixel (Tarboton et al., 1991). The distribution of this variable is strongly skewed to the right, so its logarithm is used for the modeling.

III.2.2. Type of soils and lithology

The lithological and pedological information came from different sources. For the LUCDEME project soil map (E. 1:100,000) (**6**), with the spatial distribution of taxonomic units classified according to the FAO (1974) System. The soil map mapping the Saline Phase **(7)** was used to identify areas whose electrical conductivity was > 2 mmhos / cm. The lithology map **(8)** was derived using data from the National Geological Map 1:50,000 (MAGNA) of the Geological and Mining Institute of Spain (IGME); its geological mapping was reclassified to give lithological values through interpretation of the legend for each geological formation that appears in the report accompanying each map.

III.2.3. Climatic variables

The climatic variables used for the study area are the layer of precipitation (mm) **(9)** and temperature (°C) **(10)** (LBI 2013). As well as several variables related to solar radiation (Rich et al., 1994), the computation of the radiation was performed for a full year, yielding three sets of values: the direct solar radiation **(11),** the diffuse solar radiation **(12) and** direct solar radiation **(13)** (Fu & Rich 2000).

III.2.4. Additional variables obtained by remote sensing

Landsat satellite images for two dates (14/02/2009 and 24/07/2009) were used to obtain additional information. These images correspond to two different seasons of the same year, winter and summer. In this way the possible seasonal bias is minimized. For each period the Normalized Difference Vegetation Index (NDVI) (**14** and **15**, data for winter and summer, respectively) was estimated. This index allows determination of the quantity, quality, and development of vegetation - according to the measured intensity of the radiation of wavelengths from the Red (R) and Near Infrared (NIR) parts of the electromagnetic spectrum that the vegetation emits or reflects (Townshend et al., 1985).

A series of algorithms were also applied to the images selected for mineralogical indicators (Sabins, 1999; Crosta et al., 2003), by combining the following standardized functions: 'Clay Minerals Index' (CMI) (**16** and **19**), 'Ferrous Minerals Index' (FMI) (**17** and **20**), and "Iron Oxide Index" (IOI) (**18**, **21**). For each of the functions the calculations were performed for the summer season (July) and winter (February).

Table 1 shows the summary values of the variables.

III.3. Predictive model for estimation of the SOM on a regional scale

The GIS procedures were developed with ArcGIS v.10 (ESRI, Redlands, USA). From the coordinates of the samples of the database the values of 21 environmental variables were obtained to a pixel resolution of 400 m. The values of organic matter available for each item were listed in a table corresponding to the selected environmental variables. This table was used as an input for the language 'R' (R CORE TEAM 2013) for statistical modelling by regression. A regression model in which the predictors are GIS layers can easily be represented by map algebra (Pérez-Cutillas, 2013).

To obtain the SOM model based on the environmental variables, a linear stepwise forward regression using the Akaike Information Criterion (AIC) for selecting best models was applied (Akaike, 1974). This method provides a relative measure of the information lost when a particular model is used to describe reality, so finally allowing one to select the model that best fits the actual data. Mathematically, it can be shown that the AIC measures the distance between the fitted model and the perfect model that accurately represents reality, even if this perfect model is never met with. Thus, of two models fitted to the same data, the one with a lower AIC value is preferable because it is closer to the perfect or ideal model. It is of note that a small decrease in the calculated value of the AIC represents a major improvement in the quality of the model.

As the response to the quantitative variables can be nonlinear, for all of these quadratic terms were used in the models. For example, for precipitation we included in the construction of the models not only the variable *precipitation* but also the quadratic term *precipitation*^{2}. In this way the model can easily adjust different curved forms.

IV. RESULTS

IV.1. Regional-scale modeling of SOM using environmental variables

The best predictive model generated by the method described, according to the AIC values obtained for each combination of variables, is:

'Slope' [Slp], 'profile curvature' [CuP], 'perpendicular curvature' [CuPP], 'Flow accumulation' [FAc], 'Temperature' [Tmp], 'Diffuse radiation' [RaDif], 'Clay Minerals Index' [IOI], 'Ferrous Minerals Index' [FMI], 'Iron Oxide Index' [IOI], 'Vegetation Index' [NDVI] and 'LUCDEME Soil Taxonomy' [SLu].

Figure 2 shows the spatial representation for the estimation of the SOM with the variables and coefficients obtained in the model.

IV.2. Influence and effect of environmental variables on the SOM content

Figures 3 and 4 give an overview of the influence of the variables in the prediction of SOM. Of the topographical and hydrological aspects considered in the SOM model, the most-important are the variables [SLP], [CuP], [CuPP], and [FAc]. The main effect exerted by the variable [SLP] in the model results in an increase in the percentage of SOM, associated with increased topographic slope. In [Cup] convex shapes have a higher content of SOM, whereas in the concave surfaces of the profile of maximum slope, the content declines. For [CuPP] the opposite occurs: SOM concentrations are higher in the concave surfaces of profiles perpendicular to the maximum slope. Finally, the variable [FAc] has a quadratic effect, with a stable content of SOM in zones with a smaller drainage area and a linear increase beginning from a threshold flow accumulation.

The variable [Tmp] describes an increase in the levels of SOM with increasing temperature, a process that tends to stabilize when the average annual temperature is around 17-18°C. Meanwhile, the OM content decreases with an increase in [Radi] stability above a threshold of 0.25 MWh/m^2-year.

The largest number of variables that influence the estimate of the SOM were obtained by remote sensing. The mineralogical index [IOI], for both February and July, corresponds to a clear reduction of the content of SOM. This decrease is evident as the iron-containing components detected by the sensor give way to higher values of mineralogical alteration, as also happens with the variable [CMIjul] in areas detected as having high levels of altered clay minerals. Meanwhile, [FMIjul] denotes the opposite effect: an increase in the SOM content as the alteration of iron minerals becomes more evident. Finally, [NDVIjul] shows an increased content of SOM in line with the increased photosynthetic response values associated with the presence of vegetation masses.

The effects observed for the coefficients of the qualitative variables in the model are shown in Figure 4. The effects of different levels of the soil variable [SLU] are given in relation to the *'Albic arenosols'* class, which is arbitrarily attributed level 0. Areas where the soil is in the Saline Phase recorded an increase in OM content, though rather moderate. In the case of the variable (SLU), all levels show increased SOM indices with respect to the reference class, but the surfaces of the *'Lithosols'* and their combination with the *'Calcic xerosols'* class have the highest values. Important also are the *'Calcic fluvisols'*, which are related significantly with the SOM.

V. DISCUSSION AND CONCLUSIONS

Several recently-published papers feature efficient prediction models for the organic carbon (OC) content in the surface layers of soil, relating the OC with features extracted from a DEM (slope, curvature, etc.) and soil maps (Meersmans et al., 2009;. Schwanghart & Jarmer 2011; Doetterl et al., 2013). The main objective of this data is to serve as an input layer for modeling processes at the regional level and in its current state it seems appropriate for this purpose, but future improvements are likely.

A correlational and observational model as developed here should be interpreted with caution. Regression models may suggest underlying processes, but if they do not arise from an experimental approach it is difficult to speak of causality. Moreover when environmental variables are strongly correlated with each other, as in our case, it is normal that two highly-correlated variables that independently have similar effects show, in the regression, coefficients of opposite sign due to a 'compensatory' effect (one of the two variables is redundant). In such situations, the AIC sometimes tends to select overly-complex models.

The topographic variables (Berhe, 2012; Doetterl et al., 2012) have a significant influence on the prediction of the SOM content, especially the slope (Slp) and the curvatures of the ground (CuP and CuPP). In all three cases, there is a direct relation to the decrease in SOM in flat or low-slope areas. It is likely that this effect is largely associated with the fact that the majority of the flat land is occupied by agriculture, while high SOM contents tend to be registered in steeper areas occupied by natural vegetation. This direct relationship between vegetation and the predicted SOM is observed equally for NDVIjul variables (Paustian et al., 1997; Mishra et al., 2010) and FAc. In the first case, the expected SOM increase in accordance with the increased occupation of the land by natural vegetation is observed. There are, however, certain exceptions, as some agricultural uses - such as horticultural crops - can provide a high NDVI value, inducing the model to overestimate SOM levels for these areas. Regarding the variable FAc, whose highest values represent the areas of highest flow accumulation, in these semi-arid environments it reaches its highest expression in the areas with more favorable water balance, where the natural vegetation is developed more greatly.

The relationship between SOM and temperature is counterintuitive, since at a water-deficient site, such as the one studied here, it seems contradictory that the SOM content would be greater at higher temperatures; therefore, this aspect should be improved in future models. The map of Figure 2 shows that the highest contents of SOM are predicted to occur in mountainous areas, although the altitude, negatively related to temperature, does not feature in the model. The slope assumes that role, since steeper areas are not only home to the natural vegetation but are also representative of the mountainous areas with lower temperatures and greater precipitation. Moreover, the NDVI also identifies areas with more vegetation in the areas of more-favorable water balance; thus, the net effect of the temperature should 'discount' all these effects. It is also noteworthy that, regardless of other environmental variables, this soil mapping helps to improve predictions, in a fashion similar to the results described by Zhang et al., (2012) - according to which the introduction of categorical variables, such as soil types, improves the accuracy of models used to predict the SOM.

Finally, it remains to be elucidated why certain variables related to the remote sensing of the soil mineralogical composition contribute to the prediction of OM in the soil. This issue should be studied in detail, to see if spectral ratios can be related to other ecosystem attributes which *a priori* have more connection with the formation and accumulation of SOM. However, it is interesting to explore whether mineralogical indices are related to soil fertility.