

Using spatial data mining to analyze area-diversity patterns among soil, vegetation, and climate: A case study from Almería, Spain

Enrico Feoli , Rufino Pérez-Gómez , Cecilio Oyonarte , Juan J. Ibáñez

A B S T R A C T

Area-pedodiversity, area-vegetation diversity, area-bioclimate belts, diversity and pedodiversity-vegetation diversity-bioclimate diversity relationships were analyzed for the most arid region of Western Europe (SE Spain). The study is novel in that it analyzed these relationships using as operational geographic units (OGUs) the set of polygons belonging to the same character state (i.e., typology) that defines a zone in a given thematic map. We considered three thematic maps: 1) the map of soil associations (SMU) with 303 different soil associations (i.e., 303 areas or zones); 2) the map of potential vegetation (PNV) with 40 vegetation series or zones and 3) the map of bioclimate belts (BB) with 7 bioclimate zones. Using GIS tools, we analyzed the contents (richness) of soil mapping units (polygons) with regard to vegetation series and bioclimate belts as well as the contents of vegetation series and bioclimate belts with regard to SMU. The results indicate that the relationships between the area and the richness of zones follow a power law, the “fingerprint of fractal geometry”, irrespective of the way the area has been defined and their relative magnitude (the areas defined by SMU are the smallest, while the areas defined by bioclimate belts are the largest). The results also indicate a significant correlation between the α diversities of the zones and between their β diversities. We conclude that the methods used to measure such correlations, based on α and β diversities are useful to investigate and quantify the relationships between the pedosphere, vegetation and climate.

1. Introduction

Data mining techniques (Han and Kamber, 2006) and geographic information system (GIS) technology can be combined to extract useful information from existing databases and thematic maps, even if they are at different scales. The study of area-geodiversity relationships is a relatively new area of research that has concentrated mainly on area-pedodiversity relationships. The results of these studies have proved the fractal nature of pedodiversity (see Ibáñez et al., 2005, 2009; Ibáñez and Bockheim, 2013) analogous to what has been found in studies of species abundance and diversity in ecology (Harte et al., 1999; May, 1975; Sugihara and May, 1990). In contrast, the relationships between soil and vegetation have long been of interest to researchers (e.g., Duchaufour, 1998; Pugnaire et al., 2004). Analysis aimed at understanding the relationships between geomorphology and vegetation also has a long history (e.g., Howard and Mitchell, 1985; Stallins, 2006; Viles,

1998), as does the study of relationships between soil and climate (e.g., Cerdá, 1998). However, such studies have not considered explicitly the concept of diversity as a function of area. Only recently it was shown that α pedodiversity (α diversity is the diversity within a collection: Feoli et al., 2013; Whittaker, 1975) of land systems in terms of soil types or pedotaxa is correlated to α diversity of plants (Ibáñez et al., 2014; Ibáñez et al., 2009; Ibáñez and Feoli, 2013; Petersen et al., 2010) and of plant communities (Ibáñez et al., 2016; Ibáñez et al., 2009).

This study applies a novel method for documenting the fractal nature of α pedodiversity, α vegetation diversity and bioclimate diversity. It also tests and quantifies the correlations between the vegetation system, the pedosystem and the bioclimate system using the concept of β diversity (i.e., the difference between the collections of a set, Feoli et al., 2013; Whittaker, 1975). The approach is novel in that it analyzes such relationships using as operational geographic units (OGUs) the set of polygons belonging to the same character state (i.e., typology) that defines a zone in a given thematic map. The zones are homogeneous areas with respect to a given criterion (or set of criteria) but could be heterogeneous with respect to other criteria, i.e., one zone defined by a certain criterion may contain more zones defined by other criteria. It follows

that given a land system, the richness of one of its zones (defined on the basis of certain criteria) in terms of zones defined by other criteria, may depend on its heterogeneity with respect to such zones but also by its extent: the larger the area of a zone, the higher the likelihood that it would include more zones defined by other criteria.

The specific aims of this paper are (i) to test if a power law, the “fingerprint of fractal geometry” (Bordá-de-Agua et al., 2002; Harte et al., 1999; Ibáñez et al., 2009; Mandelbrot, 1977; May, 1975; Ostling et al., 2003; San-José and Caniego, 2013; Šizling et al., 2011; Sugihara and May, 1990), can describe the area-richness relationships of vegetation series, pedotypes and bioclimatic belts (α diversities), where the areas are defined in turn by the areas of soil associations of basic pedotypes (SMU), the areas of clusters of SMU based on their basic pedotype content, the areas of vegetation series or potential natural vegetation (PNV) and the areas of bioclimatic belts (BB); and (ii) to test the strength of the correlation between the β diversities (Magurran, 2004; Whittaker, 1975) of zones of a map layer described by zones of other layers. This means, according to Mantel (1967), Mantel and Valand (1970) and in terms of similarity theory (Feoli and Orlóci, 2011), quantifying the correlation between the different sets of variables used to describe independently the same set of zones, or in other words to measure the predictivity of one set of variables with respect to another set (Feoli and Orlóci, 2011). The exercise is performed for data from the land system of the Almería province, the driest area on the European continent and the subject of recent studies concerning pedodiversity (Ibáñez et al., 2015).

2. Materials and methods

2.1. Location and characterization of the area

The study area is located in the South-Eastern part of the Iberian Peninsula. It corresponds to the whole of Almería province. The landscape of Almería shows a large variety of lithologies from Paleozoic to Holocene, including volcanic rocks (e.g., Villalobos-Megía, 2003). It has a rugged relief with numerous SW-NE oriented mountain ranges with very steep slopes, interposed by deep valleys. The Sierra Nevada is the most important mountain range reaching altitudes of around 3400 m at a short distance from the coast (Aguilar-Ruiz et al., 2004). From a physiographic point of view, Almería can be sub-divided into three main units: mountain ranges, Neogene depressions and coastal lands. The mountain ranges isolate the province from the rest of Andalusia and from the Atlantic humid circulation (Aguilar-Ruiz et al., 2004). As a consequence, rainfall is low and comes from the East (Mediterranean circulation), mainly in autumn in erratic storms. These circumstances favor evaporation and result in a water deficit that extends throughout most of the year, with the exception of the inland mountain ranges (Aguilar-Ruiz et al., 2004). From 800 to 1000 m above sea level, a more humid climate permits the growth of Mediterranean mountain xeric vegetation, including the typical dry Mediterranean forests and mountaintop shrubs, which only remain in small sites as consequence of prolonged land use. In contrast, in lowland depressions (badlands, alluvial fans, etc.) and coastlands (dunes, wetlands, and ephemeral riverbeds called “ramblas”, etc.), morphogenic systems are typically arid (Simón et al., 2005).

Low rainfall and high temperatures produce a sparse vegetation cover that is not capable of preventing intense erosion during storm events. As a result of continuous soil erosion, young and shallow soils (Regosols and Leptosols) are common on the slopes. In contrast, eroded sediments are dominant on gently sloping surfaces of lowlands (piedmonts and alluvial fans covered by more or less angular colluvium usually dissected by ravines, gullies and rivers) where the most evolved soils are found, such as Luvisols, Calcisols, Solonchaks and Gypsisols (Ibáñez et al., 2016).

Table 1
Main properties of georeferenced databases analyzed in the Almería Province.

Georeferenced data bases	No. classes	Scale map	No. polygons	Legend units
Potential vegetation map (vegetation, PNV)	40	1:10,000	5864	40
Map of bioclimatic belts (BB)	7	1:400,000	1914	7
Soil associations (SMUs)	303	1:100,000	2248	303

2.2. Data

We considered three thematic maps of the province of Almería (Table 1):

- (1) The soil map, which shows the distribution of 303 soil type associations (SMU), at a scale of 1:100,000, defined according to the WRB classification (FAO, 1998).
- (2) The map of potential natural vegetation that is based on 40 vegetation series (PNV), at a scale of 1:10,000, defined by Rivas-Martínez (2005, 2007) (see Loidi and Fernández-González, 2012 for further details).
- (3) The map of bio-climatic belts (BB), at a scale of 1:400,000, that have also been defined according to the criteria of Rivas-Martínez (2005, 2007).

The pedological map was taken from the one published by the Spanish government (Aguilar et al., 2004). The maps of vegetation series and bioclimatic belts were taken from the IDEA Andalucía Portal (Junta de Andalucía, 2013a, 2013b). Table 1 shows the main features of the cartographic databases used in this study.

These three maps correspond to three environmental systems of the province of Almería, namely the pedosystem, the vegetation system and the bioclimatic system. The three maps have been produced independently and at different times. However, thanks to the obvious correlations between soil, vegetation and climate, the polygons of the three maps, individually or in combination, may have a high degree of overlap. Among the three thematic maps, the one of bioclimatic belts has the highest dependence to one of the other two maps, namely the one of vegetation series, since it is based, according to Rivas-Martínez (2005, 2007) on the combination of vegetation and climatic data.

The three maps were georeferenced and adjusted to a common scale before being stored in a GIS environment (ArcGIS 10.1) as three independent overlapping layers. In this way, we can describe each polygon of one layer by the polygons of the other two layers.

GIS functions have been used to obtain the following four data matrices:

- (1) **Xps**, a matrix describing the SMUs by the pedotaxa or basic pedotypes (PT). The matrix has 47 rows (p = index for PTs) and 303 columns (s = index for SMUs). The number of basic pedotypes for each SMU ranges between 1 and 5 and there is no correlation between their number and the extent of the area of the SMUs.
- (2) **Xvs**, a matrix describing the SMUs by the presence of 40 vegetation series (PNV). The matrix has 40 rows (v = index for PNV) and 303 columns (s).
- (3) **Xbs**, a matrix describing the SMUs by the presence of the bioclimatic belts (BB). The matrix has 7 rows (b = index for BB) and 303 columns (s).
- (4) **Xvb**, a matrix describing the BBs by the PNVs. The matrix has 40 rows (v) and 7 columns (b).

From the matrix **Xps** we obtained three new matrices by clustering the SMUs according to their similarity based on the presence-absence

of the basic pedotypes (PT). These three matrices: X_{pc} , X_{vc} and X_{bc} , give the description of clusters (CS) of SMUs ($c = \text{index for CS}$) by the basic pedotypes (p), the vegetation series (v) and the bioclimatic belts (b). Among the several possible classifications, we chose the 53 clusters (CS) of SMUs obtained from the dendrogram given by the method of average linkage clustering applied to the similarity matrix based on Jaccard's function (Podani, 2000), because among the several classifications we have obtained by clustering methods, the number of clusters was the closest to the number of basic pedotypes (i.e. 47). The separation between the 53 clusters is highly significant according to the test of evenness of the eigenvalues ($E = 0.76$, $p < 0.001$) suggested by Feoli et al. (2009), so that the clusters can be considered as homogeneous with respect to their content of basic pedotypes. In this case, the correlation between the area of clusters and the number of pedotypes (PT) becomes significant.

For each SMU, vegetation series (PNV), bioclimatic belt (BB) and cluster (CS) we have the extent of the area in km^2 , accordingly, the matrices may be grouped into three orders of magnitude: (i) the matrices describing the SMUs by vegetation series (PNV) and by bioclimatic belts (BB) (i.e., X_{vs} and X_{bs}). This order is defined by the SMUs that show an average extent of 29 km^2 , with a standard deviation of 55.2 km^2 ; (ii) the matrices describing the vegetation series (PNV) by SMUs and by bioclimatic belts (i.e. X_{sv} and X_{bv} , the transposed matrices of X_{vs} and X_{bs}). This order is defined by the PNVs that show an average extent of 170 km^2 , with a standard deviation of 312.9 km^2 ; and (iii) the matrices describing the bioclimatic belts (BB) by SMUs and by PNVs (i.e., X_{sb} , the transposed matrix of X_{bs} , and the matrix X_{vb}). This order is defined by the BBs that show an average extent of 990 km^2 with a standard deviation of 821 km^2 .

The 53 clusters (CS) of SMUs represent another set of zones, i.e., OGU that have more or less the same order of magnitude as the PNVs because the average extent of the clusters is 165.2 km^2 with a standard deviation of 251.62 km^2 .

The areal extents of the four sets of OGUs (SMU, PNV, BB and CS) are not normally distributed: the SMUs are the smallest units with 84% of the distribution under 50 km^2 . The percentage of PNVs under 50 km^2 is 65%, while for CS it is 43% and for BB it is 14%.

2.3. Methods

The power law ($Y = cx^z$, where Y is, in our case, the number of zones per area x; c and z are empirical constants) has been applied to the following relationships: (1) area SMUs–number of PNVs; (2) area SMUs–number of BBs; (3) area PNVs–number of SMUs; (4) area PNVs–number of BBs; (5) area BBs–number of SMUs; (6) area BB–number of PNVs; (7) area CS–number of PNVs; (8) area CS–number of BBs; (9) area CS–number of PTs. The online regression program XURU (<http://www.xuru.org/rt/toc.asp>) was used to perform the regression analysis. The goodness of the fit was tested by the R^2 value obtained following logarithmic data transformation (e.g., Kleinbaum et al., 1988). The α diversities of the zones of one layer in terms of the zones of the other layers were compared by the correlation coefficient (r). For this correlation analysis the number of PTs for each PNV and BB and the number of PNVs and BBs for each PT was calculated by multiplying the matrix X_{ps} by the matrices X_{vs}^T and X_{bs}^T , respectively, and summing the binary values of the resulting matrices by rows in the first case and by columns in the second case.

The β diversity of the zones of each map was described by the matrix of similarity between the zones based on their description with the zones of the other maps (the higher the similarity between two zones, the lower their β diversity). The following similarity matrices were obtained according to the Jaccard's index (Podani, 2000):

- (1) the matrix S_{ps} of the similarity between the SMUs described by the basic pedotypes (PT);

- (2) the matrix S_{vs} of the similarity between the SMUs described by vegetation series (PNV);
- (3) the matrix S_{bs} of the similarity between the SMUs described by bioclimatic belts (BB);
- (4) the matrix S_{sv} of similarity between the PNVs described by SMUs;
- (5) the matrix S_{bv} of similarity between the PNVs described by bioclimatic belts (BB);
- (6) the matrix S_{sb} of similarity between the BBs described by SMUs;
- (7) the matrix S_{vb} of similarity between the BBs described by PNVs;
- (8) the matrix S_{pc} of similarity between the 53 clusters (CS) described by basic pedotypes (PT);
- (9) the matrix S_{vc} of similarity between the CSs described by PNV;
- (10) the matrix S_{bc} of similarity between the CSs described by BB;
- (11) the matrix S_{as} of similarity between the SMUs in terms of their extent (area);
- (12) the matrix S_{av} of similarity of the PNVs in terms of their extent;
- (13) the matrix S_{ab} of similarity between the BBs in terms of their extent;
- (14) the matrix S_{ac} of similarity between the CSs in terms of their extent.

For each of these S matrices we calculated the β diversity by the evenness of their eigenvalues following the suggestions of Feoli (2012). The evenness ranges between 0 in case of complete similarity (β diversity = 0) and 1 in case of zero similarity (β diversity = 1, i.e., the matrix of similarity is the identity matrix, in fact if there is a perfect correspondence between two sets of zones, one zone of one set would contain one zone of the other set, and vice versa).

Mantel's test was applied between the couples of the similarity matrices S of the same order (Mantel, 1967; Mantel and Valand, 1970).

The matrices S , the evenness of their eigenvalues and the correlations between the S matrices by Mantel's test were obtained by the software MATEDIT (Burba et al., 2008).

3. Results

Table 2 presents the results of the regression analysis using the power law. Table 3 shows the Pearson's correlation coefficients between the α diversities of the zones of one map described by the number of zones of the other maps. The correlation is significant in all of the cases ($p < 0.01$) except for the correlation between the number of PTs and the number of PNVs, and the number of PTs and the number of BBs in the SMU.

Table 4 presents the evenness of the eigenvalues of the similarity matrices between the zones and the clusters; the higher the evenness, the higher the dissimilarity between the zones and between the clusters in terms of combinations of zones or in terms of areas.

Table 5 shows the correlations between the β diversities derived from Mantel's test considering the zones defined by SMU, PNV and BB,

Table 2

Regression analysis for the area–richness relationships for soil map units (SMU), vegetation series (PNV), bioclimatic belts (BB), by the use of the power function (d.f. = degrees of freedom).

	Equation	R^2	d.f.	p-Value
Area SMU–n of PNV	$Y = 1.91x^{0.30}$	0.52	301	<0.01
Area SMU–n of BB	$Y = 1.26x^{0.21}$	0.48	301	<0.01
Area PNV–n of SMU	$Y = 7.59x^{0.35}$	0.81	38	<0.01
Area PNV–n of BB	$Y = 1.78x^{0.11}$	0.49	38	<0.01
Area BB–n of SMU	$Y = 8.36x^{0.36}$	0.98	5	<0.01
Area BB–n of PNV	$Y = 2.77x^{0.26}$	0.90	5	<0.01
Area CS–n of PNV	$Y = 1.9x^{0.38}$	0.79	51	<0.01
Area CS–n of BB	$Y = 1.05x^{0.26}$	0.67	51	<0.01
Area CS–n of PT	$Y = 1.14x^{0.27}$	0.56	51	<0.01

Table 3

Pearson's correlation coefficients (r) between the α diversities of the zone. Legend: CnxSMU = r for the richness (n) of SMU in PNV (vegetation series), PT (basic pedotypes) and BB (bioclimatic belts); CnxCS = r for the richness of clusters (CS) in PT, PNV and BB; CnxPT = r for the richness (n) of PT in BB, PNV and CS; CnxPNV = r for the richness (n) of PNV in PT, SMU, BB and PNV; CnxBB = r for the richness of BB in PT, PNV, SMU and CS.

CnxSMU	nPNVxSMU	nPTxSMU	nBBxSMU	
nPNVxSMU	1	-0.188	0.710	
nPTxSMU	-0.188	1	0.065	
nBBxSMU	0.710	0.065	1	
CnxCS	nPTxCS	nPNVxCS	nBBxCS	
nPTxCS	1	0.653	0.694	
nPNVxCS	0.653	1	0.770	
nBBxCS	0.694	0.770	1	
CnxPT	nBBxPT	nPNVxPT	nCSxPT	
nBBxPT	1	0.838	0.762	
nPNVxPT	0.838	1	0.908	
nCSxPT	0.762	0.908	1	
CnxPNV	nPTxPNV	nSMUxPNV	nBBxPNV	nCSxPNV
nPTxPNV	1	0.907	0.605	0.952
nSMUxPNV	0.907	1	0.687	0.933
nBBxPNV	0.605	0.687	1	0.629
nCSxPNV	0.952	0.933	0.629	1
CnxBB	nPTxBB	nPVxBB	nSMUxBB	nCSxBB
nPTxBB	1	0.826	0.882	0.964
nPNVxBB	0.826	1	0.978	0.932
nSMUxBB	0.882	0.978	1	0.960
nCSxBB	0.964	0.932	0.960	1

while Table 6 presents the results of Mantel's test by comparing the β diversities of the clusters. The correlations (r) between the similarity of the zones in terms of areas and in terms of the combinations of zones are the following: Sas with Sps = 0.007 ($p = 0.87$); Sas with Svs = 0.09 ($p < 0.05$); Sas with Sbs = 0.09 ($p < 0.05$); Sav with Svs = 0.23 ($p < 0.001$); Sav with Svb = 0.08 ($p = 0.03$); Sab with Sbs = 0.69 ($p < 0.001$); Sab with Sbv = 0.65 ($p < 0.001$). Table 6 presents the results of the Mantel's test applied to the matrices of similarity

Table 4

Evenness of the eigenvalues of the similarity matrices **S** between the zones of one thematic map described by zones of other thematic maps, i.e., the matrices of similarity between the clusters (CS) of SMU described by vegetation series (PNV) bioclimatic belts (BB) and basic pedotypes (PT) and the matrices of similarity between the zones (SMU, PNV and BB) and clusters (CS) in terms of their area.

Matrix	Evenness of the eigenvalues
Sps (SMU described by the PT)	0.67
Svs (SMU described by PNV)	0.69
Sbs (SMU described by BB)	0.36
Ssv (PNV described by SMU)	0.95
Ssb (BB described by SMU)	0.91
Sbv (PNV described by BB)	0.64
Svb (BB described by PNV)	0.89
Spc (53 CS described by PT)	0.90
Svc (53 CS described by PNV)	0.78
Sbc (53 CS described by BB)	0.43
Sas (303 SMU described by their area)	0.32
Sav (40 PNV described by their area)	0.60
Sab (7 BB described by their area)	0.64
Sac (53 CS described by their area)	0.47

Table 5

Mantel's test (r) between the matrices of similarity obtained by the description of each set of OGU's (zones): s = SMU, v = vegetation series, b = bioclimatic belts, by the other two categories and by the area a. The upper part (columns) of the 3×3 matrix is the correlation coefficient between the similarity matrices, the lower part (rows) is the probability of r .

	Sbs	Svs	Ssv	Sbv	Ssb	Svb
Sbs	1	0.52	1	0.53	1	0.93
Svs	0.00	1	Sbv	0.00	1	Svb
Ssv			1	1		
Sbv				1		
Ssb					1	
Svb						1

describing the clusters by the area (Sac), by basic pedotypes (Spc), by vegetation series (Svc) and by bioclimatic belts (Sbc).

4. Discussion

Studies on pedodiversity (Ibáñez and Bockheim, 2013) have made use of the same mathematical tools applied by ecologists over decades, demonstrating that biodiversity and pedodiversity at different scales are comparable (Saldaña, 2013; Ibáñez and Bockheim, 2013; Ibáñez et al., 2014, and references therein). We can say that notwithstanding differences in the nature of the collections—animals, plants, plant communities on one side and pedotypes, landforms, lithotypes on the other—the relationship between diversity and area is always significantly fitted by a power law. The type of OGU (zone) on which the diversity is calculated does not matter and neither does the order of magnitude of the area that is considered (pedotypes, vegetation series, clusters of SMUs, and bioclimatic belts). The parameter z of the curve always ranges between 0.20 and 0.40 (Bordá-de-Agua et al., 2002; Harte et al., 1999; Ibáñez et al., 2009; May, 1975; Ostling et al., 2003; San-José and Caniego, 2013; Šizling et al., 2011; Sugihara and May, 1990). The only relationship that does not fit the power curve is the area of SMU and the number of basic pedotypes (PT). This is a consequence of the fact that the areas of SMU are all very similar notwithstanding their high standard deviation; in fact, the evenness of the eigenvalue of the matrix of similarity between the SMU in terms of their area is the lowest among all the evenness values ($E = 0.32$, Table 4).

The correlations between the α diversities corresponding to the levels of magnitude defined by zones of different thematic maps (i.e., pedological maps, maps of vegetation series and maps of bioclimatic belts), are all significant except for the relationships between the number of PNVs and PTs and the number of BBs in the SMU. This is certainly a consequence of the relatively small area of the SMU that produces high random variation in the combinations of zones they may contain.

This finding is supported by the fact that all the correlations between the β diversities are highly significant (Tables 5 and 6) after excluding only the correlation between the matrices Sas (SMU described by areas) and Sps (SMU described by basic pedotypes). The correlation between the similarity of SMU in terms of area and in terms of PNV and BB are also the lowest, although they are still significant ($r = 0.09$ for both). Not high but still significant are the correlations between the matrices of similarity of SMU in terms of basic pedotypes (PT) and in terms of vegetation series (PNV) ($r = 0.18$, $p < 0.01$) and between the matrices of similarity of SMU in terms of PNV and BB ($r = 0.17$, $p < 0.01$).

Table 6

Mantel's test between the matrices of similarity obtained by the description of the 53 clusters of SMUs by the basic pedotypes (Spc), the vegetation series (Svc), the bioclimatic belts (Sbc) and the areas of the clusters (Sac). The upper part of the matrix shows the correlation coefficients between the similarity matrices, the lower part the probability of not being correlated.

	Spc	Svc	Sbc	Sac
Spc	1	0.39	0.42	0.31
Svc	0.00	1	0.67	0.36
Sbc	0.00	0.00	1	0.37
Sac	0.00	0.00	0.00	1

The highest β diversity in terms of evenness (Table 4) is reached for zones described by SMU: the β diversity of PNV described by SMU is 0.95, the β diversity of BB described by SMU is 0.91, and the β diversity of CS described by PT (the pedotypes describing the SMU) is 0.90. This means that in the province of Almería there is a high correlation (correspondence) between the pedotypes, the vegetation series and the bioclimatic belts. In terms of β diversity the highest correlation is shown between the vegetation series and SMU when they describe the bioclimatic belts (BB) ($r = 0.93$, Table 5), i.e., at the highest level of magnitude. This means that, in the province of Almería, the climate is a very important factor in determining both pedogenesis and vegetation structure and dynamics.

Further studies on the effects of the climatic gradient, as expressed by the bioclimatic belts, on both vegetation series and on pedotypes would be necessary to explain in more detail the links and interactions between the three components of the land systems (i.e., pedosphere, vegetation and climate) and the diversity associated with their heterogeneity. Such studies should also consider the similarity between the SMUs in terms of the chemical and physical properties of the soil, the similarity between the vegetation series in terms of species and of functional traits and the similarity between bioclimatic belts in terms of climatic variables, especially precipitation and temperature.

5. Conclusions

Our data mining approach offered useful tools for analysing biogeodiversity at spatial extents covering several orders of magnitude. The fact that diversity-area relationships can always be fitted by a power law function proves again the fractal nature of the spatial pattern of the features of land systems and determines the linear correlations between the α diversities of the zones described by different variables, in our case soil associations (SMU) described by vegetation series (PNV) and bioclimatic belts (BB), vegetation series described by SMUs and BB, bioclimatic belts described by SMUs and PNV, and clusters of SMUs described by SMUs, PNV and BB. The use of similarity matrices allowed us to quantify the correlations between the factors that determine such spatial patterns. These results support and encourage further research in modelling pedodiversity and vegetation diversity relationships at different spatial scales and with respect to climatic gradients and human impacts. This knowledge is required for land management and conservation planning.

Acknowledgements

The authors acknowledge the University of Trieste (Department of Life Sciences) for supporting this research on data analysis of complex systems. The authors also thank two anonymous reviewers and the editor of the special issue for their helpful suggestions on how to improve the text.

References

Aguilar-Ruiz, J., Martín-Peinado, F., Sierra-Aragón, M., Ortíz-Silla, R., Oyonarte, C., 2004. Mapa Digital de Suelos: Provincia de Almería. Dirección General para la Biodiversidad. Ministerio de Medio Ambiente. NIPO, Spain (311-04-082-5).

Bordá-de-Agua, L., Hubbell, S.P., McAllister, M., 2002. Species-area curves, diversity indices, and species abundances distributions: a multifractal analysis. *Am. Nat.* 159, 138–155. <http://dx.doi.org/10.1086/324787>.

Burba, N., Feoli, E., Malaroda, M., 2008. MATEDIT: a software tool to integrate information in decision making processes. In: Neve, R., Barretta, J.W., Mateus, M. (Eds.), *Perspectives on Integrated Coastal Management in South America*. IST PRESS, Lisbon, pp. 123–127 (ISBN: 978-972-8469-74-0).

Cerdá, A., 1998. Relationships between climate and soil hydrological and erosional characteristics along climatic gradients in Mediterranean limestone areas. *Geomorphology* 25, 123–134. [http://dx.doi.org/10.1016/S0169-555X\(98\)00033-6](http://dx.doi.org/10.1016/S0169-555X(98)00033-6).

Duchaufour, P., 1998. *Handbook of Pedology: Soils, Vegetation, Environment*. Balkema, Rotterdam (ISBN-10: 9054107820).

FAO, 1998. World reference base for soil resources. *World Soil Resources Report Vol. 84*. FAO, Rome.

Feoli, E., 2012. Diversity patterns of vegetation systems from the perspective of similarity theory. *Plant Biol.* 146, 797–804. <http://dx.doi.org/10.1080/11263504.2012.740090>.

Feoli, E., Orlóci, L., 2011. Can similarity theory contribute to the development of a general theory of the plant community? *Community Ecol.* 12, 135–141. <http://dx.doi.org/10.1556/ComEc.12.2011.1.16>.

Feoli, E., Gallizia Vuerich, L., Ganis, P., Zerihun, W., 2009. A classificatory approach integrating fuzzy set theory and permutation techniques for land cover analysis: a case study on a degrading area of the Rift Valley, Ethiopia. *Community Ecol.* 10, 53–64. <http://dx.doi.org/10.1556/ComEc.10.2009.1.7>.

Feoli, E., Ganis, P., Ricotta, C., 2013. Measuring diversity of environmental systems 2013 In: Ibáñez, J.J., Bockheim, J. (Eds.), *Pedodiversity*. C.R.C. Press, Boca Raton, CA, pp. 29–58 (ISBN-10: 1466582774).

Han, J., Kamber, M., 2006. *Data Mining: Concepts and Techniques*. 2nd ed. Elsevier, Amsterdam, Boston (ISBN 13: 978-1-55860-901-3).

Harte, J., Kinzig, A.P., Green, J., 1999. Self-similarity in the distribution of abundance of species. *Science* 284, 334–336. <http://dx.doi.org/10.1126/science.284.5412.334>.

Howard, J.A., Mitchell, C.W. (Eds.), 1985. *Phytogeomorphology*. Wiley, N.Y. (ISBN 0-471-09914-7).

Ibáñez, J.J., Bockheim, J., 2013. *Pedodiversity*. C.R.C. Press, Boca Raton CA (ISBN-10: 1466582774).

Ibáñez, J.J., Feoli, E., 2013. Global relationships of pedodiversity and biodiversity. *Vadose Zone J.* 12, 3. <http://dx.doi.org/10.2136/vzj2012.0186>.

Ibáñez, J.J., Caniego, J., San-José, F., Carrera, C., 2005. Pedodiversity-area relationships for islands. *Ecol. Model.* 182, 257–269. <http://dx.doi.org/10.1016/j.ecolmodel.2004.04.005>.

Ibáñez, J.J., Arnold, R.W., Ahrens, R.J., 2009. The fractal mind of pedologists (soil taxonomists and soil surveyors). *Ecol. Complex.* 6, 286–293. <http://dx.doi.org/10.1016/j.ecocom.2009.05.007>.

Ibáñez, J.J., Zuccarello, V., Ganis, P., Feoli, E., 2014. Pedodiversity deserve attention in plant biodiversity research. *Plant Biol.* 148, 1112–1116. <http://dx.doi.org/10.1080/11263504.2014.980357>.

Ibáñez, J.J., Pérez-Gómez, R., Oyonarte, C., Brevik, E.C., 2015. Are there arid land soilscapes in southwestern Europe? *Land Degrad. Dev.* 26, 853–862. <http://dx.doi.org/10.1002/ldr.2451>.

Ibáñez, J.J., Pérez-Gómez, R., Ganis, P., Feoli, E., 2016. The use of vegetation series to assess α and β vegetation diversity and their relationships with geodiversity in the province of Almería (Spain) with watersheds as operational geographic units. *Plant Biol.* <http://dx.doi.org/10.1080/11263504.2016.1165755>.

Junta de Andalucía, 2013a. WMS Mapa del modelo de distribución de los pisos bioclimáticos de Andalucía (IDEA Andalucía) Reference to a web site: [http://www.juntadeandalucia.es/medioambiente/site/rediam/template.PAGE/informacionambiental/?jvax.portlet.tpst=d3b6d0b274a881654ca24ca260425ea0&jvax.portlet.prp_d3b6d0b274a881654ca24ca260425ea0=origen%3Ddarbol%26action%3DfichaMetadatoController%26paginaActual%3DBiogeogra%25C3%25ADa%26pagina%3D04010104%26idInternal%3D13352&jvax.portlet.begCacheTok=com.vignette.cachetoken&jvax.portlet.endCacheTok=com.vignette.cachetoken \(12/28/2015\)](http://www.juntadeandalucia.es/medioambiente/site/rediam/template.PAGE/informacionambiental/?jvax.portlet.tpst=d3b6d0b274a881654ca24ca260425ea0&jvax.portlet.prp_d3b6d0b274a881654ca24ca260425ea0=origen%3Ddarbol%26action%3DfichaMetadatoController%26paginaActual%3DBiogeogra%25C3%25ADa%26pagina%3D04010104%26idInternal%3D13352&jvax.portlet.begCacheTok=com.vignette.cachetoken&jvax.portlet.endCacheTok=com.vignette.cachetoken (12/28/2015)).

Junta de Andalucía, 2013b. Mapa de Series de Vegetación de Andalucía (Spatial Data Infrastructure of Andalusia Portal (IDEA Andalucía)). Reference to a website: [http://www.juntadeandalucia.es/medioambiente/site/rediam/template.PAGE/informacionambiental/?jvax.portlet.tpst=d3b6d0b274a881654ca24ca260425ea0&jvax.portlet.prp_d3b6d0b274a881654ca24ca260425ea0=origen%3Ddarbol%26action%3DfichaMetadatoController%26paginaActual%3DVegetaci%25C3%25B3n%2BPotencial%26pagina%3D04010102%26idInternal%3D16802&jvax.portlet.begCacheTok=com.vignette.cachetoken&jvax.portlet.endCacheTok=com.vignette.cachetoken \(12/28/2015\)](http://www.juntadeandalucia.es/medioambiente/site/rediam/template.PAGE/informacionambiental/?jvax.portlet.tpst=d3b6d0b274a881654ca24ca260425ea0&jvax.portlet.prp_d3b6d0b274a881654ca24ca260425ea0=origen%3Ddarbol%26action%3DfichaMetadatoController%26paginaActual%3DVegetaci%25C3%25B3n%2BPotencial%26pagina%3D04010102%26idInternal%3D16802&jvax.portlet.begCacheTok=com.vignette.cachetoken&jvax.portlet.endCacheTok=com.vignette.cachetoken (12/28/2015)).

Kleinbaum, D.G., Kupper, L.L., Muller, K.F., 1988. *Applied Regression Analysis and Other Multivariable Methods*. PWS Publishing Company, Boston (ISBN 0-871-50123-6).

Loidi, J., Fernández-González, F., 2012. Potential natural vegetation: reburying or reboring? *J. Veg. Sci.* 23, 596–604. <http://dx.doi.org/10.1111/j.1654-1103.2012.01387.x>.

Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell, Oxford, UK (ISBN: 978-0-632-05633-0).

Mandelbrot, B.B., 1977. *The Fractal Geometry of Nature*. W.H. Freeman and Co, N.Y., USA (ISBN 0-7167-0473-0).

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.

Mantel, N., Valand, R.S., 1970. A technique of nonparametric multivariate analysis. *Biometrics* 26, 547–558. <http://dx.doi.org/10.2307/2529108>.

May, R.M., 1975. Patterns of species abundance and diversity 1975 In: Cody, M.L., Diamond, J.D. (Eds.), *Ecology and Evolution of Communities*. Harvard Univ. Press, Cambridge, MA, USA, pp. 81–120 (ISBN 10: 0674224442).

Ostling, A., Harte, J., Green, J.L., Kinzig, A.P., 2003. A community-level fractal property produces power-law species-area relationships. *Oikos* 103, 218–224. <http://dx.doi.org/10.1034/j.1600-0706.2003.12600.x>.

Petersen, A., Gröngroft, A., Miehlisch, G., 2010. Methods to quantify the pedodiversity of 1 km² areas. Results from southern African drylands. *Geoderma* 155, 140–146. <http://dx.doi.org/10.1016/j.geoderma.2009.07.009>.

Podani, J., 2000. *Introduction to the Exploration of Multivariate Biological Data*. Backhuys, Leiden (ISBN 90-5782-067-6).

Pugnaire, F.I., Armas, C., Valladares, F., 2004. Soil as a mediator in plant-plant interactions in a semi-arid community. *J. Veg. Sci.* 15, 85–92. <http://dx.doi.org/10.1111/j.1654-1103.2004.tb02240.x>.

Rivas-Martínez, S., 2005. Notions on dynamic-catenal phytosociology as a basis of landscape science. *Plant Biol.* 139, 135–144 Reference to a website: <http://dx.doi.org/10.1080/11263500500193790>.

- Rivas-Martínez, S., 2007. Mapa de series, geoseries y geopermaseries de vegetation de España (Parte I). *Itineraria Geobotanica* Vol. 17, pp. 5–436 (ISSN 0213-8530).
- Saldaña, A., 2013. Pedodiversity and landscape ecology 2013 In: Ibáñez, J.J., Bockheim, J.G. (Eds.), *Pedodiversity*. CRC Press, Boca Raton, CA, pp. 105–132 (ISBN 9781466582774).
- San-José, F., Caniego, F.J., Ibáñez, J.J., Bockheim, J., 2013. Fractals and multifractals in pedodiversity and biodiversity analysis 2013 *Pedodiversity*. C.R.C. Press, Boca Raton, CA, pp. 79–104 (ISBN-10: 1466582774).
- Simón, M. (Ed.), 2005. *Almería. Factores formadores y suelos*. Editorial Universidad 543 de Almería, Almería (ISBN: 84-8240-767-3. <mailto:http://www.secs.com.es/wp-content/uploads/2016/02/2005-op-1.pdf> (accessed 5.8.2016)).
- Šizling, A.L., Kunin, W.E., Šizlingová, E., Reif, J., Storch, D., 2011. Between geometry and biology: the problem of universality of the species-area relationship. *Am. Nat.* 178, 602–611. <http://dx.doi.org/10.1086/662176>.
- Stallins, J.A., 2006. Geomorphology and ecology: unifying themes for complex systems in biogeomorphology. *Geomorphology* 77, 207–216. <http://dx.doi.org/10.1016/j.geomorph.2006.01.005>.
- Sugihara, G., May, R.M., 1990. Applications of fractal to ecology. *Trends Ecol. Evol.* 5, 79–86. [http://dx.doi.org/10.1016/0169-5347\(90\)90235-6](http://dx.doi.org/10.1016/0169-5347(90)90235-6).
- Viles, H.A. (Ed.), 1988. *Biogeomorphology*. Basil Blackwell, Oxford, UK (ISBN 0-631-15405-1).
- Villalobos-Megía, M.A., 2003. *Geology of the Arid Zone of Almería: South East Spain: An Educational Field Guide*. Fundación Gypaetus, Granada (2007. ISBN: 978-84-668 935194-7-6 http://www.juntadeandalucia.es/medioambiente/web/ContenidosOrdenacion/red_informacion_ambiental/PDF/Geodiversidad/Guia_geologica_sureste_almeriense_ingles.pdf (accessed 23.3.2016)).
- Whittaker, R.H., 1975. *Community and Ecosystems*. Mac Millan Publisher C.O., N.Y. (ISBN-10: 0024273902).