ANNALI DI BOTANICA
Ann. Bot. (Roma), 2017, 7: 11–24
Journal homepage: http://annalidibotanica.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA
DIPARTIMENTO
DI BIOLOGIA AMBIENTALE
ANNALI DI BOTANICA
COENOLOGY AND PLANT ECOLOGY
Volume 7, 2017
Published in Rome (Italy)
Pietro Romualdo Pirotta, founder, 1884

HOW DOES SPATIAL SCALE AFFECT SPECIES RICHNESS MODELLING? A TEST USING REMOTE SENSING DATA AND GEOSTATISTICS

Marcantonio M.[1], Martellos S.[2]*, Altobelli A.[2], Attorre F.[3], Tordoni E.[2],
Ongaro S.[2], Rocchini D.[5, 6, 7], Da Re D.[2], Chiarucci A.[4], Bacaro G.[2]

[1]*Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine,*
University of California, Davis, USA;
[2]*Department of Life Sciences, University of Trieste, Via L. Giorgieri 10, 34127, Trieste, Italy;*
[3]*Department of Environmental Biology, Sapienza University of Rome, P.le A. Moro 5, 00185, Roma, Italy;*
[4]*Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum University of Bologna,*
Via Irnerio 42, 40126, Bologna, Italy;
[5]*University of Trento, Center Agriculture Food Environment, 38010, S. Michele all'Adige (TN), Italy;*
[6]*University of Trento, Centre for Integrative Biology, 38123, Povo (TN), Italy;*
[7]*Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity*
and Molecular Ecology, 38010, S. Michele all'Adige (TN), Italy;
* *Corresponding author*: *Telephone*: *+39040558 3889; email: martelst@units.it*

(Received 31 January 2017; received in revised form 14 April 2017; accepted 24 April 2017)

Abstract – Spatially-explicit dataset of plant species occurrences collected in the Province of Siena (Central Italy) is analysed, with the aim of investigating *a*) the relative role of environmental factors in shaping spatial patterns of plant species richness, and b) how the spatial scale at which predictors have been sampled determines the explicative power of species richness models. The optimal spatial resolution of analysis was evaluated with respect to the total deviance explained by models, using a set of environmental and remotely sensed derived predictors calculated at different spatial scales. Results confirm the hypothesis that the predictive power of landscape structure is influenced by the spatial scale at which predictor variables have been sampled. Furthermore, the relevance of identifying a proper geographical scale of investigation, hence minimizing the redundancy in the predictor variables and maximising the explanatory power of the single groups of predictor variables, is highlighted as well.

Keywords: community structure, kriging, generalised linear models, spatial autocorrelation

Introduction

Identifying spatial patterns in species diversity is relevant for developing biodiversity conservation and monitoring strategies (Caley et al., 2014). In spatial ecology, statistical models relate ecological process or patterns (such as species diversity) to environmental properties, eventually allowing their representation as distributional maps (Elith et al., 2006; Elith & Graham, 2009; Blasi et al., 2011; Attorre et al., 2014; Martellos et al., 2014; Amici et al. 2015). Since plant species richness is one of the most common indicators of total species diversity, understanding its spatial variation is relevant to develop conservation and management strategies. Both explanatory and predictive power of ecological models are strongly influenced by the selection of appropriate predictors (Austin, 2002; Austin et al., 2006). At local or plot scale, patterns of plant diversity are associated to local factors, such as topography and soil type, as well as disturbance and competitive interactions (Tilman, 1982; Ellenberg, 1988, Ibáñez et al., 2014), while, macro-climate

doi: 10.4462/annbotrm-13804

and latitude shape plant diversity at larger spatial scales (Elith & Leathwick, 2009).

The analysis of how the spatial configuration of sampling units influences species richness estimates has become an important issue (Kühn, 2007), since species richness is one of the simplest and most popular diversity measures (Chao & Jost, 2012; Chiarucci et al., 2011, Bacaro et al. 2016). Spatial autocorrelation (SAC) is associated with biological patterns and processes, and it must be taken into account while modelling ecological data at large scale (Bacaro & Ricotta, 2007). Indeed, as stated by Kühn (2007), "if spatial autocorrelation is ignored we simply do not know if we can trust the results at all". Therefore, the presence of residual spatial autocorrelation should always be tested for in spatial ecology. If SAC is not taken into account, residual errors will contain unexplained spatial patterns, leading to Residual Spatial Autocorrelation (RSA). Hence, assumptions on independently and normally distributed errors (common to most statistical procedures) are violated, thus inducing biased Type I error estimates, due to degrees of freedom inflation (Clifford et al., 1989; Legendre et al., 2002). Furthermore, parameter estimates may be biased, or their magnitude inverted (Lennon, 2000; Bino et al., 2008, Bacaro et al. 2016). Hence, understanding the interaction between SAC and spatial scaling of species-environment relationships is a mandatory task in ecology (De Knegt et al., 2010, Bacaro et al. 2016).

Geostatistical principles are commonly applied in ecology for taking into account autocorrelated data structures (Bacaro & Ricotta, 2007; Bacaro et al., 2011; Jiménez-Alfaro & Iriondo, 2014). The inclusion of spatial variance is known to improve model effectiveness (Diggle & Ribeiro, 2007). In this paper, geostatistical modelling was applied to a dataset of vascular flora collected in the Province of Siena (Tuscany, Central Italy) in order to:

a) identify the environmental factors that underlay the spatial variation of plant species richness in the study area;

b) understand the influence of the spatial scale at which predictor variables are measured on the modelling process.

By using a multi-scalar approach, the predictive power of a set of variables on species richness was evaluated, once the spatial structure of sampled data was modelled. We moved from the hypothesis that the predictive power of variables changes with the spatial scale at which these explicative variables are calculated. Indeed, in some cases, the relationship between species richness and predictor variables can change in magnitude and/or direction when measured at different spatial scales.

Materials and methods

Study area

The study was carried out in the Province of Siena (Fig. 1), an area of 3,821 km^2 in Tuscany (Central Italy, centroid coordinates: longitude 11° 26′ 54″ E, latitude 43° 10′ 12″ N, *datum* WGS84). The area has a North-West / South-Est orientation and its elevation ranges between 59 and 1685 meters a.s.l.. It is characterized by low hills (89% of territory is between 59 and 546 meters), and protected areas cover 1,109 km^2 (29% of the whole Province). 49% of the territory is dominated by annual or perennial forage crops, while 24% is dominated by deciduous oakwoods (Bartolozzi et al. 1995). Other land use categories cover the rest of the area, creating a complex landscape mosaic. Such environmental diversification promotes the establishment of peculiar micro-climatic conditions, producing a wide range of habitats, some of which are included in the Habitat Directive 92/43/CEE.

Forests mostly occur in hilly and mountainous areas. The dominant species are oaks (*Quercus ilex* L., *Q. pubescens* Willd., *Q. cerris* L.), pines (*Pinus pinaster* Ait., *P. pinea* L.), chestnut (*Castanea sativa* Mill.), beech (*Fagus sylvatica* L.) and spruce (*Abies alba* Mill.).

Figure 1. Study area: province of Siena (Tuscany, Central Italy).

Sampling design and data collection

Few studies focused on patterns of plant diversity in Mediterranean forests at a regional scale, as well as on their relationships with environmental and spatial determinants (e.g. Brunet et al., 2000; Kolb & Diekmann, 2004), hence, sampling was focused on forest habitats. A composite dataset of 424 10x10 m plots was built by collecting data from different published vegetation surveys. Each sampling campaigns differed in the study area, shape (square) and area of the plots (10x10 m) was always the same. More in details, the whole analysed dataset were set up as follows:

1) 276 plots were randomly selected within the 21 Protected Areas (PA) included in the "Mo.bi.SIC." project (refer to Chiarucci et al., 2008 for details on sampling design).
2) further 93 plots were obtained by the "MonITO–TopModel" project, ended in 2001 (Chiarucci & Bonini, 2005). In this case, the sampling design was characterized by 20x20 m macro-plots, divided in four 10x10 m adjacent plots. Hence, in our analysis, each macro-plot was considered as 4 grouped 10x10 m plots.
3) 55 plots were sampled in the framework of a project aimed at investigating riparian *Alnus glutinosa* dominated woods (Landi & Angiolini, 2010).

Total species richness was considered as the main response variable. Data were collected between 2000 and 2011. Vascular plants names were standardized among the three studies by using Pignatti (1982) and Tutin et al. (1993). Field activities took place in late spring-early summer in all datasets.

Predictor variables

37 variables were considered as potential predictors (Table 1). They were grouped in: 1) environmental (climate, topography, geology and primary productivity); 2) landscape (human disturbance, fragmentation and land use dynamics), and 3) spatial (plot coordinates).

Table 1. Descriptive statistics of predictor variables. Units of measure are reported in the last column.

Spatial scale	**100**		**250**		**500**		**750**		**1000**		**Units**
Variable	*Mean*	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*	
Slope	11.75	6.1	11.56	5.01	11.24	4.06	11.19	4.01	11.03	3.29	degrees
Elevation	442.3	207.05	442.16	204.66	443.25	203.59	235.45	110.34	168.04	57.58	meters
Temperature	13.2	1.05	13.21	1.04	13.2	1.05	13.21	1.04	13.2	1.03	°C / month
NDVI mean	0.42	0.12	0.42	0.09	0.42	0.08	0.4	0.09	0.35	0.12	-
NDVI st.dev.	0	0	0.06	0.04	0.07	0.04	0.11	0.04	0.15	0.05	-
Insolation time	12.88	0.97	12.91	0.82	12.97	0.68	12.99	0.63	13.03	0.56	hours
Rainfall	71.20	3.82	71.19	3.82	71.25	3.79	71.25	3.8	71.24	3.82	mm / month
Perimeter sum	587.7	240.65	2200.68	1086.16	6859.64	3517.93	15435.32	5619.87	24948.41	10687.69	meters
Perimeter mean	858.98	361.97	876.67	386.78	902.04	395.12	950.45	401.12	1048.48	490.82	meters
Artificial rate	0.0004	0.0040	0.09	0.14	0.09	0.13	0.07	0.06	0.03	0.03	-
Agricultural rate	0.08	0.22	0.12	0.24	0.13	0.24	0.13	0.23	0.15	0.2	-
Mean perimeter area ratio	1	0.9	1.13	1.38	2.15	1.46	5.39	4.77	8.54	5.23	-
Number of patch	2	1.55	4.42	3.81	10.4	10.7	19.45	18.6	32.94	30.4	-
Area mean	6853.1	3289.56	25717.1	19697.63	48002.17	45070.85	51876.98	47654.34	57653.95	53611.38	meters^2
Mean shape index	6.01	2.53	9.18	4.59	14.47	7.4	20.66	9.82	27.28	11.58	-

Environmental features

Mean annual precipitation and mean annual temperature (as well as their standard deviation) were derived from climatic maps of Central Italy at 5 km resolution (provided by LaMma, research center for the Environmental Monitoring and Modelling for Sustainable Development). Slope, aspect and insolation time were derived from a 20 m resolution Digital Terrain Model using GRASS modules r.slope.aspect and r.sun (Neteler et al., 2012). Geological classes were derived from a 1:100000 geologic map of Italy. Normalized Difference Vegetation Index (NDVI) was calculated from two ortho-Landsat ETM+ images (path 192, row 029-030, acquisition date June 20th 2000; Bands 1-5 and 7, spatial resolution 30 m), acquired from Global Land Cover Facility site hosted by the University of Maryland (http://glcf.umd.edu/, Tucker, 2004). June was chosen as reference month, since it is the one with the maximum vegetation spread in Mediterranean areas (*e.g.*, Rocchini & Vannini, 2010). NDVI was calculated as: $\lambda NIR - \lambda R / \lambda NIR + \lambda R$, where λNIR is the reflectance in the NIR part of the spectrum (0.76 – 0.90 µm electromagnetic window) and λR *is* the reflectance in the Red part of the spectrum (0.63 – 0.69 µm electromagnetic window). Both NDVI standard deviation - as a proxy of environmental heterogeneity (*e.g.*, Levin et al., 2007; Kumar et al., 2009) - and mean - as a proxy of Net Primary Productivity - were used (Rocchini et al., 2010).

Landscape features, complexity and forest change

Land use categories were derived from the IV levels of Corine Land Cover Map 2006 (http://land.copernicus.eu/). A high resolution map for land use (scale 1:10.000) was used to extract the following landscape variables: number of patches, average patch area, patch perimeter sum, average patch perimeter, mean shape index, mean perimeter area index, agricultural ratio and artificial ratio. These are assumed to be proxies of landscape fragmentation in its various components (habitat loss, habitat reduction, species diversity decrease; Amici et al., 2015). The variable "forest landscape change" refers to changes in land use in the province of Siena occurred between 1954 and 2000. It was derived from maps (Geri et al., 2010, 2011) developed through a cross-classification procedure among georeferenced aerial photos taken in 1954-1955 by the Italian military geographic institute (for details on the classification procedure refer to Geri et al. (2010, 2011). Three land-use classes were taken into account: "Forests", "Agricultural areas" and "Semi-natural areas".

Spatial features

Latitude and Longitude were used to test the presence of spatial autocorrelation in the distribution of plant species richness (Dormann, 2007; Kühn, 2007).

Grain resolutions of cells adopted

To investigate whether and how predictors calculated at different spatial scales could influence the explicative power of models, the study area was divided into grids with cells of different size: 100, 250, 500, 750 and 1000 m. The finer grid was chosen taking into account the coarsest resolution of the predictor variables. Only those cells covered by forests for more than 50% (according to IV level of Corine Land Cover 2006) were used for prediction. Predictors' means and standard deviations calculated at each spatial scale are presented in Table 1.

Model development

A geostatistical modelling approach was applied to take in account both the combined effect of ecological predictors and SAC on species richness. A gridded map of observed species richness was produced assembling the plots from the three datasets for each grid scale (as defined above). The modelling process followed three steps: 1) data normalization; 2) development of a generalized linear spatial model; 3) redundancy analysis.

Data normalization

Counts (the number of species in a cell) are usually modelled assuming a Poisson distribution (together with log link function in order to avoid predicted values lower than 0). However, over-dispersion (occurring when the ratio between the mean and the variance of the response variable exceeds the value of 1) implies the need for normalizing the dataset, and to deal with transformed Gaussian models (Csontos et al., 2007). Thus, since the number of species showed over-dispersion at all the spatial scales considered, a Box-Cox normalization was used (Legendre & Legendre, 1998). The lambda (λ) parameter was estimated by maximizing the log-likelihood profile. Furthermore, a Bonferroni outliers test (Cook & Weisberg, 1982) was applied to remove potential outliers from the model.

Generalized linear spatial model

The response variable (species richness) in each grid cell was expressed as:

$$(x_i , y_i) : i = 1,...., n \quad [1]$$

where x_i identifies the spatial location (in two-dimensional space, longitude and latitude) and y_i is the plant richness value associated with xi. A geostatistical (isotropic) model can be defined as:

$$Y_i = S(x_i) + Z_i : i = 1,....., n \quad [2]$$

where

$$\{ S(x) : x \in R \} \qquad [3]$$

is a Gaussian process with a spatially varying mean $\mu(x)$ defined by a classical linear regression model such as:

$$\mu(x) = \beta_0 + \beta_j p_j \qquad [4]$$

with p_j ($j=1,....,s$) expressing the j^{th} explanatory variable p.

Its variance is given by:

$$\sigma^2 = \mathrm{Var}\{S(x)\} \qquad [5]$$

and by a positive-defined correlation function

$$\rho(u) = \mathrm{Corr}\ \{S(x), S(x')\} \qquad [6]$$

defining the way correlation decays to zero for increasing distances occurring between observations at locations x and x'. The term Z_i in the model formulation represents mutually independent $N(0, \tau^2)$ random variables (see Diggle & Ribeiro, 2007) for mathematical and statistical details.

A generalized linear spatial model was then fitted as follows:

a) Explanatory variables for modelling the large-scale variation in plant diversity were chosen via Akaike Information Criterion (AIC).

b) A reduced linear model describing the spatially varying mean related with the number of plant species was then calculated. The best predictor subset was obtained, and regression coefficients estimated.

The residuals of the model were tested for spatial autocorrelation, and a suitable family of correlation functions was chosen (Hoeting, 2006). The spatial relationships for residuals were modelled computing an empirical semivariogram for a vector of distance classes *h*. The following parameters for depicting autocorrelative spatial structure (theoretical semivariogram) were estimated (Diggle & Ribeiro, 2007): nugget (τ^2, the intercept of the variogram), sill ($\tau^2 + \sigma$, expressed as the difference between the asymptote and the nugget), and range (φ, indicating the distance at which the theoretical variogram reaches its maximum). A practical range was also defined as the distance at which the correlation function reaches the value of 0.05. Since the estimation of spatial parameters strongly depends on the selection of the correlation function $\rho(u)$, different fits of a parametric Matern (1986) function of order *k* were obtained. Hence, the correlation function which maximized the likelihood estimation was selected. The estimates of parameters in the trend surface (model spatial component) were updated using an optimization function (Nelder & Mead, 1965), followed by an estimation of maximum likelihood of covariance parameters using residuals (Ribeiro & Diggle, 2001). In this process, the inclusion of one or more explanatory variables could drastically change or reduce the correlation structure of residuals (Hoeting, 2006). Cross-validation statistics by leave-one-out procedure were used to assess the bias and the accuracy of the final spatial model. Universal kriging (Krige, 1976) was then used to predict expected plants richness (and its variation) in each grid cell. For each cell, the kriging error (r) was calculated as the difference between the predicted species richness (v*) and the observed value (v):

$$r = v^* - v \qquad [7]$$

Kriging errors were standardised and used for comparing errors at different spatial scales, and for selecting the models with minimized bias.

Redundancy analysis ordination

Redundancy analysis (RDA, Legendre & Legendre, 1998) was used to assess the amount of variation explained by the three groups of predictors (environmental, landscape and spatial). The following components of variation were extracted from each model:

1. Pure effect of Environmental factors;
2. Pure effect of Landscape factors;
3. Pure effect of Spatial factors;
4. Joint effect of Environmental and Landscape factors;
5. Joint effect of Environmental and Spatial factors;
6. Joint effect of Landscape and Spatial factors;
7. Unexplained variance.

Variance partitioning was performed for each spatial scale.

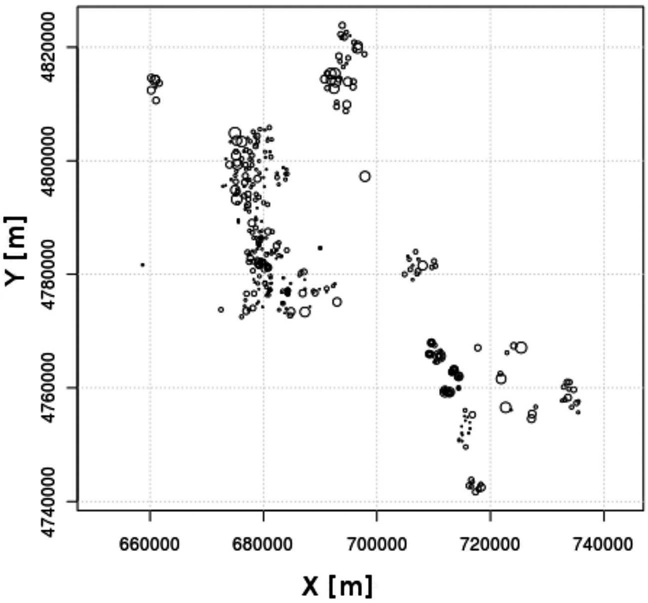


Figure 2. Sampling points depicted as circles with radius proportional to plant species richness. Coordinates are in UTM -WGS84 zone 32N.

The pure effect of each explanatory variables group was extracted for investigating its contribution to the total explained variance at different spatial resolutions explored (Amici et al., 2015). Analyses were performed by using the R software (R Development Core Team, 2015). The vegan package (Oksanen et al., 2011) was used for RDA analysis, while the geoR package (Ribeiro & Diggle, 2001) for the geo-spatial analyses.

RESULTS

Plants species richness

Mean species richness for the whole dataset was 24.46 (minimum 3, maximum 70, see Figure 2 for species richness distribution in the survey area). Count data resulted not normally distributed (Fig. 3), and, for this reason, Poisson models were applied. However, over-dispersion parameter resulted higher than 1 and the Box-Cox normalization was applied. Model coefficients (obtained by maximum likelihood analyses) were calculated for each spatial scale, and are reported in Table 2. Figure 4 shows the distribution of observerd vs. predicted species richness values at each spatial scale. At the coarsest scale, maximum likelihood was the smallest, while its R^2 value was the highest (Table 2).

Some model predictors showed constant coefficients estimates, while others varied, sometimes also changing their magnitude (Table 3). All coefficients for land use classes showed a stable pattern, while the significance of geologic classes varied with the spatial scale. Landscape and environmental variables increase their contribute to the total explained variance at the coarser spatial scales (starting from 750 m), while being scarcely or not significant at finer scales. In general, the individual role of predictors changed changing the spatial scale, with the exception of land use classes.

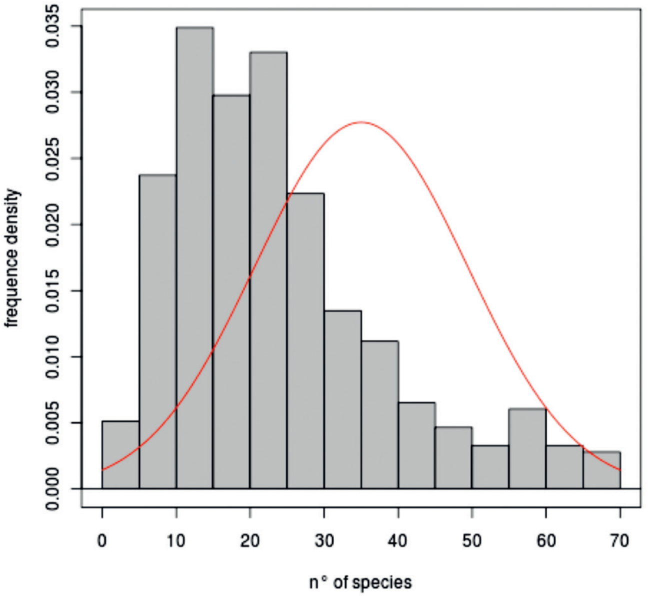
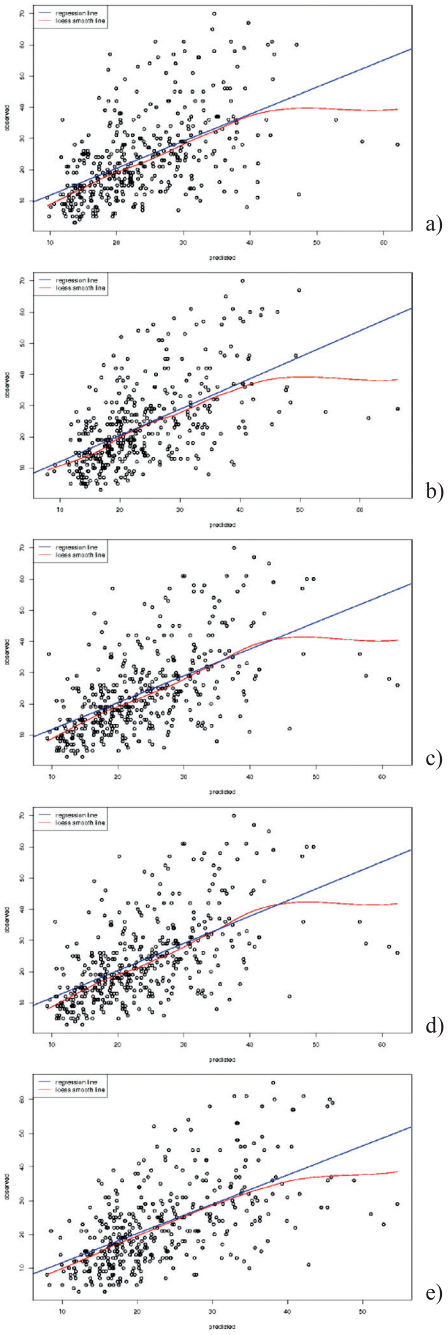


Figure 3. Frequency distribution of plants species richness values. The red line suggest the expected normal distribution curve.



a)

b)

c)

d)

e)

Figure 4. "Observerd vs. predicted" species richness value for all the predictive model developed. Regression line is in blue, while LOESS smooth line in red.

Table 2. Predictors' coefficients and their significance at different spatial scale. At bottom spatial parameters derived from the maximum likelihood analysis are listed.^{1} Please refers to the Corine land cover IV legend (ISPRA).^{2} Please refers to 1:100000 geologic map of Italy legend (ISPRA)

Category	Variable	100 meters	pr(>|t|)	250 meters	pr(>|t|)	500 meters	pr(>|t|)	750 meters	pr(>|t|)	1000 meters	pr(>|t|)
	intercept	-2.0996		-3.5892		-3.1874		-10.7135		-23.1809	*
Environmental	313 ^{1}	-	-	-	-	0.2947		-	-	-	-
	3111 ^{1}	-	-	-	-	-	-	-	-	0.1541	***
	3112 ^{1}	-	-	-	-	-	-	-	-	0.3638	***
	3113 ^{1}	-	-	-	-	-	-	1.2854	*	-	-
	3114 ^{1}	0.3403	*	-	-	-0.7802		0.3061	.	-	-
	3115 ^{1}	-	-	-0.7580	***	0.2750	**	-0.8640	.	-1.0244	***
	3111 ^{1}	0.1612	**	0.1311	.	0.5057		0.2237	*	-	-
	3112 ^{1}	0.3377	**	0.3073		0.2612	**	0.4109	*	-	-
	312 ^{1}	-	-	-	-	1.4964		-	-	-	-
	3116 ^{1}	1.4457	*	1.4913	.	-0.6884	*	-	-	2.0320	**
	Flood deposits ^{2}	-0.6849	*	-0.8065	*	-0.2459	*	-	-	-0.7661	*
	Sands and congl. ^{2}	-	-	-	-	-	-	0.2581	.	-0.4258	*
	Clays ^{2}	0.3814		-	-	0.4811	.	0.2526	*	0.3442	*
	Limestone and dolomite ^{2}	0.5438	**	0.4657		-	-	0.5529	.	0.4516	*
	Congl. and sandstones ^{2}	-0.2087		-0.2530	*	0.0033	.	-0.2515	*	-0.3016	*
	Slope	-	-	-	-	-	-	0.0309		0.0373	
	Elevation	-	-	0.0020	***	0.7311	***	0.0037	**	0.0073	***
	Temperature	0.0010	*	0.6418	**	1.4156	*	0.9429	***	1.7676	***
	NDVI mean	0.5066	*	1.5720	*	-0.2273	***	0.8976	***	1.1343	**
	NDVI standard deviation	-	-	0.5430	*	0.8900	.	-	-	-	-
	Insolation time	-	-	-	-	-0.0110	*	-	-	-	-
	Rainfall	-0.0097	***	-0.0011	**	-	-	0.0054	**	-	-
Landscape	Perimeter sum	-	-	-	-	-	-	-	-	0.1095	***
	Perimeter mean	-	-	-	-	-	-	-	-	0.4198	**
	Artificial rate	-	-	-	-	-	-	0.5187	**	0.6490	**
	Agricultural rate	-	-	-	-	-	-	0.6046	*	1.4595	***
	Mean per. area ratio	-0.0001		-0.0001		-	-	0.0006	**	-	-
	Number of patch	0.2722	*	-	-	-	-	-0.0092	.	-0.0124	***
	Area mean	-	-	-	-	0.0148	.	-	-	-	-
	Mean shape index	-	-	-	-	0.4028	**	-	-	-	-
Land use dynamics	Forest - Forest	-0.0146	.	-0.2203	.	-0.2783	**	-0.2095	*	-04054	***
	Forest - Semi-natural areas	0.2801	***	-	-	-	-	-	-	0.2601	**
	Forest - Agricultural	-	-	0.4625	**	0.4501	*	0.3656	**	0.4006	.
	Semi-natural - Semi-natural	-	-	0.6528		-	-	-	-	0.7046	.
	Agricultural - Forest	-	-	0.6558	*	1.3087	***	0.4024	*	0.5790	*
	Agricultural - Semi-natural	1.1833	*	1.4720	***	1.3476	***	1.5859	***	1.3298	***
	Agricultural - Agricultural	0.6674	***	0.7647	***	-	-	0.5973	**	0.6000	*
Spatial	τ2	0.7177		0.8427		0.8300		0.7642		0.8708	
	σ2	0.5610		1.2160		0.5543		0.4812		0.7764	
	Φ	11739.5375		18800.9998		26647.7449		29407.4980		58815.0037	
	R-squared	0.5399		0.5705		0.5755		0.5896		0.6059	
	Number of variables	17		19		21		23		24	

Spatial autocorrelation parameters showed high variability. The trend for φ (maximum distance of spatial autocorrelation) was a steady increase of spatial autocorrelation distance, when increasing the sampling spatial scale (Table 2). Total explained variance for each model was higher after the inclusion of spatial autocorrelation (Table 2). Table 3 shows values of variance related to each group of variables.

Table 3. Variance partition table. Total explained variance (R^2) was derived by cross-validation procedure of spatialised models. Spatial variance was calculated subtracting total variance, explained by not spatialized linear model, from total variance, explained by spatialized models. Other variance values (Environmental, Landscape, Environment al+Landscape) were calculated by RDA, performed on not spatialised linear models.

Spatial scale	Environ.	Landscape	Env. + Land.	Space	Total
100 meters	0.21	0.09	0.25	0.29	0.54
250 meters	0.27	0.09	0.29	0.28	0.57
500 meters	0.28	0.08	0.30	0.28	0.58
750 meters	0.27	0.09	0.32	0.26	0.59
1000 meters	0.30	0.12	0.37	0.24	0.61

Table 4. Descriptive statistics of species richness predicted values for the finest and the coarsest spatial scales. Descriptive statistics of the mean prediction error and of its standardised value are reported at the bottom.

Spatial Scale	100 meters	1000 meters
Predicted Species Richness		
Min	0.2	8.0
Max	75.2	54.6
Median	30.9	21.6
Mean	31.0	23.6
St.dev.	9.9	8.3
Standardized error		
MSE	-0.015	0.001
VSE	0.016	1.112
Error		
Mean	-0.1273	-0.1231
Variance	128.1552	103.3721

The variance explained by spatial structure was smaller at larger spatial scales, while pure environmental variance showed its maximum at 500 m. Models at the coarser scale had the highest values of total explained variance, and of environmental-plus-landscape explained variance (Table 3). In general, the spatial arrangement of plots is an explicative factor, especially at smaller spatial scales, while it is less relevant at coarser scales. The amount of variance explained by spatial structure ranged between 24% and 29% (Table 3).

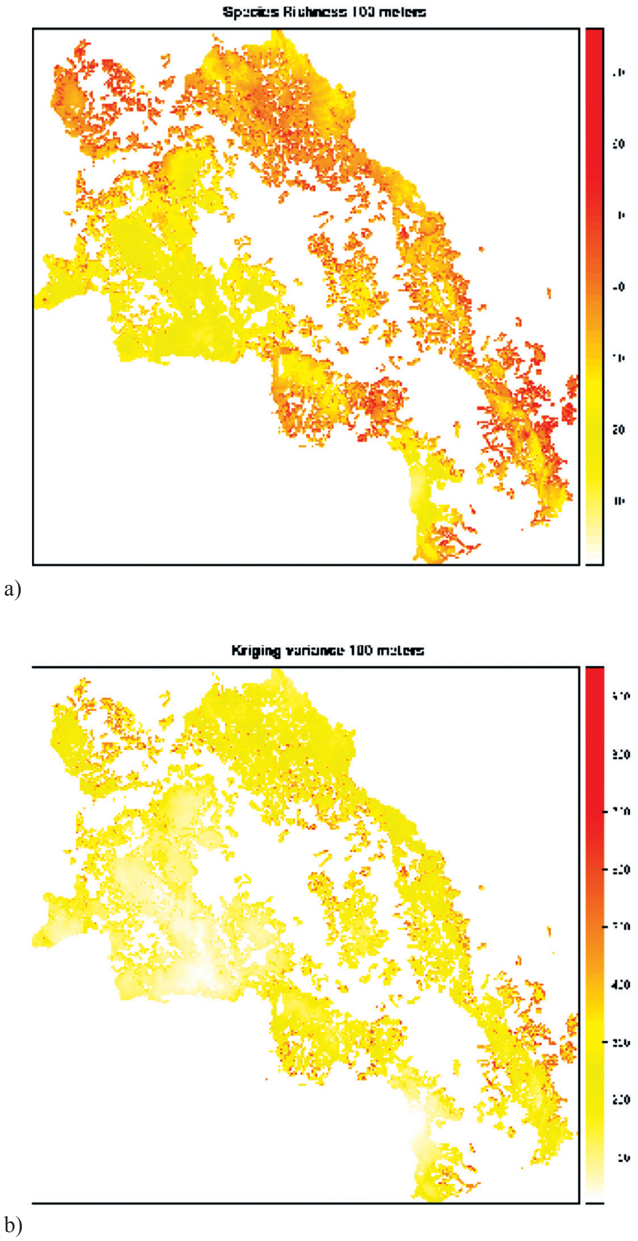
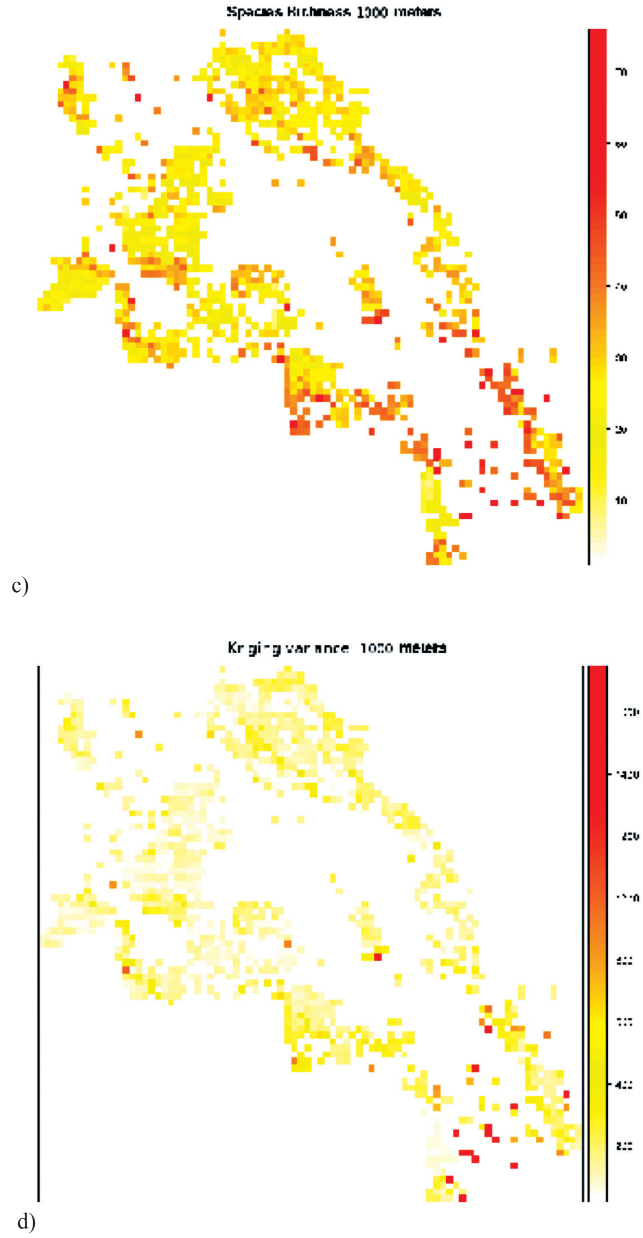


Figure 5. Predicted Plant species richness distribution maps at a) 100 m and c) 1000 m cell resolution. b) and d) present the associated kriging error variance.

Predicted Plant Species Richness

Maps in Figure 5 show the predicted species richness as well as its error variance at the finer (100 m) and coarser (1000 m) spatial scale. Descriptive statistics of predicted species richness are reported in Table 4, along with kriging error and its standardised value in cross–validation. The values for the larger spatial scale are always lower. The mean of standardized errors was similar, but the variance was higher for the 100 m spatial scale (Table 4). Different species richness values for the same areas (Figure 5a and 5c) are evident especially in the North-Est side of the study area. The variance associated with prediction (Figure 5b and 5d) shows the highest values at the coarser spatial scale, which, on the contrary, has the lowest average values of species richness (Table 4).



Discussion

The different role of predictors in determining plant species richness

Species richness is an emergent property of natural systems, which arises from multiple ecological determinants (Currie, 1991). In this study, predictive models explained more than 50% of the total variance in the distribution of plant species richness, a robust result, especially if taking into account the high heterogeneity of the landscape characterising the study area (De Dominicis, 1993).

Model prediction power considerably increased when spatial autocorrelation was taken into account, stressing its key role in species diversity modelling. The importance of distance-based relationships is related to the dispersal ability of species, as well as to the complex inter-specific relations which can occur within or between communities (Pottier et al., 2007, Bacaro et al. 2012). Another explanation can be related to the spatial structure of the sampling design (Fortin & Dale, 2005). Since the sampling cells were derived from projects focused on monitoring protected areas, some portions of the survey area were over-sampled, thus possibly increasing spatial autocorrelation (Dormann, 2007).

Environmental variables explained a higher variance than landscape variables at any spatial scale. The relevant role of geology is justified by its influence on the structure of ecosystems, since the study area is relatively heterogeneous from a geo-lithological point of view.

Plant species richness resulted also strongly related to temperature, as can be deduced in Table 2 (it was included in all the models performed), and already highlighted in several other works (Rosenzweig, 1995; Levin et al., 2007), which relate a low number of species in an area with the limit of available energy (Currie, 1991). Mean temperature, positively correlates with species richness in our models. A similar trend – but with a different relevance – was observed between elevation and temperature, while varying spatial scale. These two variables are strictly related, and so is their effect on species richness (Rey-Benayas, 1995). Shorter growing season at higher elevations, together with colder average temperature, limits species richness in forest habitat (Bacaro et al., 2008). However, temperature is a proximal environmental predictor for species richness, while elevation is a distal one (Austin, 2002). This may explain why temperature model coefficient had always a greater value than altitude. Furthermore, the coefficient value of temperature increased almost monotonically with spatial scale, maybe due to the nature of temperature data, derived by spatial interpolation of sparse meteorological stations. While producing smoothed temperature surfaces, representative of regional climatic trends, this interpolation overlooks temperature spatial variability at smaller scales.

NDVI is one of the most important environmental factors associated with species richness. It correlates well with plant biomass, net primary productivity, tree density and canopy cover, which are all relevant proxies of species diversity in forest ecosystems (Evans et al., 2005; Gillespie, 2005). On the contrary, NDVI standard deviation – a proxy of environmental heterogeneity (Levin et al., 2007; Kumar et al., 2009) – showed a relatively low importance at the coarser spatial scales. Thus, productivity, rather than heterogeneity, seems to be correlated with plant species richness in forest habitats.

Landscape structure, although secondary to environmental variables, contributed to explain the observed species richness spatial pattern. Consistently with previous studies (Amici et al. 2015), this investigation demonstrates that the relationship between local plant species richness in forests and landscape variables generally increases with increasing the spatial scale. A wider spatial scale is likely to increase the probability of including those landscape features which act as determinant or barriers for plant dispersal. Our results show a moderate increase in the explained variance for total species richness with landscape metrics calculated over an increasing extent.

Changes in land use are amongst the most relevant transformations of Earth's surface (Gillanders et al., 2008). Plant communities developed after afforestation of abandoned lands differ from ancient forests as far as soil properties modification (Glatzel, 1991), and relative colonisation rates of species (Brunet & Von Oheimb, 1998) are concerned. In our study, semi-natural land use classes derived from the transformation of agricultural areas showed a positive relationship with plant species richness.

In the same way, the proportion of urban and agricultural surface were positively related with species richness at the coarser spatial scales. These patterns can be interpreted considering that within forest communities, forest specialized taxa are likely to be intrinsically more vulnerable to the surrounding landscape composition (e.g., the number and amount of different habitat types) and configuration (the spatial arrangement of such habitat types) than open habitat species, or species that are only partly dependent on forest habitat (e.g. those species which naturally occur at forest edges). Human presence in forest habitat causes forest fragmentation, and a relevant edge effect. Consequently, heterogeneity increases and more ecological niches become available (Marcantonio et al., 2013). While the exclusion of these two variables at smaller scales may seem counter-intuitive, the dominant land use of small forest patches may be trivial in affecting species diversity, since it is easily masked by the surrounding land use dynamics.

The role of spatial resolution in species richness modelling

Over the past years, the techniques for modelling ecological properties have become more and more effective, and detailed maps of environmental variables have become available (e.g. Nagendra & Rocchini, 2008). This is a relevant advantage when investigating the relationships between environment and species, since a detailed knowledge of the spatial variation of environmental variables allows a more detailed investigation of diversity distribution patterns. However, the predictive power of models is expected to vary by changing the resolution at which predictor variables are sampled. Therefore uncertainty, as well as the importance of certain variables, can increase or decrease as a function of spatial scale.

Modelling bias is a relevant issue when estimating the biodiversity of an area (Rocchini et al., 2010). Models can be relevant to research and to biodiversity management only if they are reliable, and errors are 1) assessed, 2) minimized and 3) reported in ecological literature (Rocchini et al., 2010). We assessed the kriging variance (KV), as well as mean and variance of prediction error, for each model. These uncertainty estimates are reported for the smallest and largest spatial scales only, since the other values were mid-points between the two. The maximum KV is measured in areas with maximum distance from sampling cells, or nearest to the boundary of the study area. This trend is particularly evident due to the clustering of sampling points inside protected areas, consequence of the clustered sampling designs.

Spatial richness pattern generated by the models become increasingly congruent with observed data at the coarser spatial resolution, as evidenced by other studies (Graham & Hijmans, 2006). The mean prediction error and its standardized value, estimated by cross-validation, were minimized at 1000 m spatial scale. Therefore, the model at the coarser spatial scale revealed the lowest associated degree of uncertainty, underlying the challenging nature of modelling species richness at finer spatial scales. Averaging predictors at larger spatial scale results in filtering out fine scale random noise generated by multiple sources of uncertainty (sampling, interpolation, etc), which otherwise may mask the existing relationships between these predictors and species richness.

Predictive variables and spatial scale

A central issue in ecology is related to the role of scale in determining patterns of species diversity (Scheiner, et al., 2000). In our study, increasing the spatial extent for averaging predictor variables values lead to geostatistical models with an increased predictive power. We observed a marked scale dependency of predictor variables in modelling procedures, which could indicate the presence of key ecosystem processes. This dependency should be considered in ecological modelling, and needs to be incorporated into any general theory which aim at explaining the relationship between species diversity and environment (e.g. Pastor et al., 1996). The importance of environmental and landscape variables increases with spatial scale, since factors such as disturbance and fragmentation affect species richness at a coarser scale. From a probabilistic point of view, using a wider spatial scale, landscape metrics are expected to include those elements acting as determinant or barriers for plants dispersal (Amici et al., 2015). Hence, the relationship between plant community and landscape predictors grows in strength with the spatial scale (Wu et al., 2000; Millington et al., 2003).

Conclusions

Kriging regression allows to couple the most common correlative model algorithms (e.g. GLM) with an intuitive and robust assessment of the spatial structure of the modelled variable, and model uncertainty is easily visualized through the kriging variance. Hence, we suggest this framework as a valuable tool to model and predict the spatial pattern of species richness.

We adopted geostatistical models to identify the best spatial scale for predicting plant species richness in relation to environmental and landscape variables measured at that specific spatial scale. The main outcome was that the variance explained by models increased with the spatial resolution of the predictor variables. At coarser spatial scales, landscape fragmentation and heterogeneity become key factors associated with species richness, linking richness with ecological processes that take place at medium-large spatial scale. However, the marginalization of noise when averaging the predictors across larger extent should be also considered as a factor contributing in decreasing uncertainty. A high variability in variable selection and in their importance was observed at different spatial scales. This could indicate that there is not a "better" spatial scale for sampling all predictors, and each predictor should be sampled individually at an adequate spatial scale. From an applied point of view, understanding the effects of specific predictors on species diversity at community scale could help in promoting effective environmental policies and landscape management practices. For instance, sustainable forest harvesting and management should take into account landscape structure. In addition, the knowledge of correlates of species diversity can help in finding possible proxies, or surrogates, useful in routine assessments, large-scale monitoring, and predictive modelling of plant diversity (Austin, 2002).

References

Amici V., Bacaro G., Geri F., Santi E., Filibeck G., Landi S., Scoppola A., Chiarucci A., 2015. Effects of landscape structure on local forest plant diversity: a multi-scale approach in a Mediterranean context. Ecological Complexity 21, 44-52.

Attorre F., Issa A., Malatesta L., Adeeb A., De Sanctis M., Vitale M., Farcomeni A., 2014. Analysing the relationship between land units and plant communities: The case of Socotra Island (Yemen). Plant Biosystems 148(3), 529-539.

Austin M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modelling 157 (2-3), 101-118.

Austin M.P., Belbin L., Meyers J.A., Doherty M.D., Luoto M., 2006. Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. Ecological Modelling 199 (2), 197-216.

Bacaro G., Ricotta C., 2007. A spatially explicit measure of beta diversity. Community Ecology 8 (1), 41-46.

Bacaro G., Rocchini D., Bonini I., Marignani M., Maccherini S., Chiarucci A., 2008, The role of regional and local scale predictors for plant species richness in Mediterranean forests. Plant Biosystems, 142 (3), 630-642.

Bacaro G., Santi E., Duccio R., Pezzo F., Puglisi L., Chiarucci A. 2011, Geostatistical modeling of regional bird species richness: exploring environmental proxies for conservation purpose. Biodiversity and Conservation 20 (8), 1677-1694.

Bacaro, G., Rocchini, D., Duprè, C., Diekmann, M., Carnesecchi, F., Gori, V., Chiarucci, A., 2012, Absence of distance decay in the similarity of plots at small extent in an urban brownfield. Community Ecology 13, 36-44.

Bacaro G., Altobelli A., Cameletti M., Ciccarelli D., Martellos S., Palmer M.W., Ricotta C., Rocchini D., Scheiner S.M., Tordoni E, Chiarucci A., 2016, Incorporating spatial autocorrelation in rarefaction methods: Implications for ecologists and conservation biologists. Ecological Indicators 69, 233-238.

Bartolozzi, L., Bussotti, F., De Dominicis, V. Ferretti, M., 1995, Program MONITO: Concepts, Structure and Results. Regione Toscana, Giunta Regionale Publisher, Firenze.

Bino G., Levin N., Darawshi S., Van Der Hal N., Reich-Solomon A., Kark S., 2008. Accurate prediction of bird species richness patterns in an urban environment using Landsat-derived NDVI and spectral unmixing. International Journal of Remote Sensing 29(13), 3675-3700.

Blasi C., Marignani M., Copiz R., Fipaldini M., Bonacquisti S., Del Vico E., Rosati L., Zavattero L., 2011. Important Plant Areas in Italy: From data to mapping. Biological Conservation 144(1), 220-226.

Brunet J., Von Oheimb G., 1998. Migration of vascular plants to secondary woodlands in southern Sweden. Journal of Ecology 86 (3), 429-438.

Brunet J., Von Oheimb G., Diekmann M., 2000. Factors influencing vegetation gradients across ancient-recent woodland borderlines in southern Sweden. Journal of Vegetation Science 11(4), 515-524.

Caley, M.J., Rebecca F., Kerrie M., 2014. Global Species Richness Estimates Have Not Converged. Trends in Ecology & Evolution 29(4), 187–188.

Chao A., Jost, L.,2012. Coverage-based rarefaction and extrapolation:standardizing samples by completeness rather than size. Ecology 93, 2533–2547.

Chiarucci A., Bacaro G., Rocchini D., 2008, Quantifying plant species diversity in a Natura 2000 Network: old ideas and new proposals. Biological Conservation 141(10), 2608-2618.

Chiarucci, A., Bacaro, G., Scheiner, S.M., 2011. Old and new challenges in using species diversity for assessing biodiversity. Philosophical Transaction of the Royal Society: Series B 366, 2426-2437.

Clifford P., Richardson S., Hemon D., 1989. Assessing the significance of the correlation between two spatial processes. Biometrics 45(1), 123-134.

Cook R.D., Weisberg S., 1982. Residuals and Influence in Regression. New York: Chapman and Hall. Retrieved from the University of Minnesota Digital Conservancy, http://purl.umn.edu/37076.

Csontos P., Rocchini D., Bacaro G., 2007. Modelling factors affecting litter mass components of pine stands. Community Ecology 8(2), 247-255.

Currie D.J., 1991. Energy and large-scale patterns of animal- and plant-species richness. American Naturalist 137:, 27-49.

De Dominicis V., 1993. La vegetazione. In: “La storia naturale della Toscana meridionale”. Monte dei Paschi di Siena, Amilcare Pizzi Editore, Siena, Italia, pp. 247-343.

De Knegt H. J., van Langevelde F. V., Coughenour M. B., Skidmore A. K., De Boer W. F., Heitkönig I. M. A., Prins H. H. T., 2010. Spatial autocorrelation and the scaling of species-environment relationships. Ecology 91(8), 2455-2465.

Diggle P., Ribeiro J.R., 2007. Model-based Geostatistics. Springer Science & Business Media.

Dormann C. F., 2007, Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16(2), 129-138.

Elith J., Graham C.H., 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32, 66-77.

Elith J., Leathwick J.R., 2009. Species distribution models: Ecological Explanation and Prediction across Space and Time. Annual Review of Ecology and Systematics 40, 1-20.

Elith J.H., Graham C.P., Anderson R., Dudík M., Ferrier S., Guisan A., Hijmans J.R., Huettmann F., Leathwick R.J., Lehmann A., Li J., Lohmann G.L., Loiselle A.B., Manion G., Moritz C., Nakamura M,. Nakazawa Y., Overton M.J., Peterson T.A., Phillips J.S., Richardson K., Scachetti-Pereira R., Schapire E.R., Soberón J., Williams S., Wisz S.M., Zimmermann E.N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29(2), 129-151.

Ellenberg, H., 1988, Vegetation Ecology of Central Europe. Cambridge University Press.

Evans K.L., Greenwood J.D, Gaston K.J., 2005. Dissecting the species–energy relationship. Proceedings of the Royal Society of London B: Biological Sciences 272(1577), 2155-2163.

Fortin M.J., Dale M.R.T., 2005. Spatial Analysis. A Guide for Ecologist. Cambridge, Cambridge University Press, London.

Geri F., Amici V., Rocchini D., 2011. Spatially-based accuracy assessment of forestation prediction in a complex Mediterranean landscape. Applied Geography 31(3), 881-890.

Geri F., Rocchini D., Chiarucci A., 2010. Landscape metrics and topographical determinants of large-scale forest dynamics in a Mediterranean landscape. Landscape and Urban Planning 95, 46-53.

Gillanders S.N., Coops N.C., Wulder M.A., Gergel S.E., Nelson T., 2008. Multitemporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends. Progress in Physical Geography 32(5), 503-28.

Gillespie T.W., 2005. Predicting woody-plant species richness in tropical dry forests: a case study from south Florida, USA. Ecological Applications 15, 27-37.

Glatzel G., 1991. The impact of historic land use and modern forestry on nutrient relations of Central European forest ecosystems. Fertilizer Research 27, 1-8.

Graham C. H., Hijmans R. J., 2006. A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography 15(6), 578-587.

Hoeting J.A., 2006. Model selection for geostatistical models. Ecological Applications 16(1), 87-98.

Ibáñez J.J., Zuccarello V., Ganis P., Feoli E., 2014. Pedodiversity deserves attention in plant biodiversity research. Plant Biosystems 148(6), 1112-1116.

Jiménez-Alfaro B., Iriondo J.M., 2014. Population dynamics of Aster pyrenaeus Desf., a threatened species of temperate forest edges: A view of meso- and micro-scales. Plant Biosystems 148(6), 645-654.

Kolb A., Diekmann M., 2004. Effects of environment, habitat configuration and forest continuity on the distribution of forest plant species. Journal of Vegetation Science 15(2), 199-208.

Krige D.G., 1976. Two-dimensional weighted moving average trend surfaces forest evaluation. Journal of the South African Institute of Mining and Metallurgy 66, 13-38.

Kühn I., 2007. Incorporating spatial autocorrelation may invert observed patterns. Diversity and Distribution 13(1), 66-69.

Kumar S., Simonson S., Stohlgren T., 2009. Effects of spatial heterogeneity on butterfly species richness in Rocky Mountain National Park, CO, USA. Biodiversity and Conservation 18(3), 739-763.

Landi M., Angiolini C., 2010. Osmundo-Alnion woods in Tuscany (Italy): A phytogeographical analysis from a west European perspective. Plant Biosystems 144(1), 93-110.

Legendre P., Dale M.R.T., Fortin M.J., Gurevitch J., Hohn M., Myers D., 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25(5), 601-615.

Legendre P., Legendre L., 1998. Numerical Ecology. Volume 24, (Developments in Environmental Modelling). 2nd Ed., Elsevier, Science B, Amsterdam.

Lennon J.J., 2000. Red-shifts and red herrings in geographical ecology. Ecography 23(1), 101-113.

Levin N., Shmida A., Levanoni O., Tamari H., Kark S., 2007. Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. Diversity and Distribution 13(6), 692-703.

Marcantonio M., Rocchini D., Geri F., Bacaro G., Amici V., 2013. Biodiversity, roads, & landscape fragmentation: Two Mediterranean cases. Applied Geography 42, 63–72.

Martellos S., Attorre F., Farcomeni A., Francesconi F., Pittao E., Tretiach M., 2014. Species distribution models backing taxa delimitation: the case of the lichen Squamarina cartilaginea in Italy. Flora 209(12), 698-703.

Millington A.C., Velez-Liendo X.M., Bradley A.V., 2003. Scale dependence in multitemporal mapping of forest fragmentation in Bolivia: implications for explaining temporal trends in landscape ecology and applications to biodiversity conservation. ISPRS Journal of Photogrammetry and Remote Sensing 57(4), 289-299.

Nagendra H., Rocchini D., 2008. High Resolution satellite imagery applied to biodiversity study in the tropics: the devil is in the detail. Biodiversity and Conservation 17(14), 3431-3442.

Nelder J.A., Mead R., 1965. A Simplex Method for Function Minimization. Computer Journal 7(4), 308-313.

Neteler M., Bowman M. H., Landa M., Metz M., 2012. GRASS GIS: A multi-purpose open source GIS. Environmental Modelling and Software 31, 124-130.

Oksanen J., Kindt R., Legendre P., O'Hara B., Stevens M. H. H., Oksanen M. J., Suggests M. A. S. S., 2007. The vegan package. Community Ecology package, 631-637.

Pastor J., Mladenoff D., Haila Y., Bryant J., Payette S., 1996. Biodiversity and ecosystem processes in boreal regions. In: Mooney H.A., Cushman J.H., Medina E., Sala O.E., Schulze E.D. (eds) Functional roles of biodiversity: a global perspective. pp. 33-69.Wiley Press, New York, USA.

Pignatti S., 1982. Flora d'Italia. Edagricole, Bologna, Italy.

Pottier J., Marrs R.H., Bédécarrats A., 2007. Integrating ecological features of species in spatial pattern analysis of a plant community. Journal of Vegetation Science 18(2), 223-230.

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rey-Benayas J.M., 1995. Patterns of diversity in the strata of boreal montane forest in British Columbia. Journal of Vegetation Science 6(1), 95-98.

Ribeiro P.J., Diggle P.J., 2001. GeoR: a package from geostatistical analysis. RNEWS 1(2), 15-8.

Rocchini D., Balkenhol N., Carter G.A., Foody G.M., Gillespie T.W., He K.S., Kark S., Levin N., Lucas K., Luoto M., Nagendra H., Oldeland J., Ricotta C., Southworth J., Neteler M., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. Ecological Informatics 5(5), 318-329.

Rocchini D., Vannini A., 2010. What is up? Testing spectral heterogeneity vs. NDVI relationship by quantile regression. International Journal of Remote Sensing 31(10), 2745-2756.

Rosenzweig M.L., 1995. Species Diversity in Space and Time. Cambridge University Press, New York.

Scheiner S.M., Cox S.B., Willig M.R., Mittelbach G.G., Osenberg C.W., Kaspari M., 2000. Species richness, species-area curves and Simpson’s paradox. Evolutionary Ecology Research 2(6), 791-802.

Tilman D., 1982. Resource competition and community structure. Monograph of Populations Biology 17(1), pp. 296.

Tucker C., Ostwald M. J., Chalup S. K., 2004. A method for the visual analysis of streetscape character using digital image processing. In Contexts of Architecture: Proceedings of the 38th Annual Conference of the Architectural Science Association ANZAScA and the International Building Performance Simulation Association, Launceston, Tasmania: Australia and New Zealand Architectural Science Association (pp. 134-140).

Tutin T.G., Heywood V.H., Burges N.A., Moore D.M., Valentine D.H., Walters S.M., Webb D.A., 1993. Flora Europaea, 2nd ed. Vol. 1, Cambridge University Press, Cambridge.

Wu J., Jelinski D. E., Luck M., Tueller P. T., 2000. Multiscale analysis of landscape heterogeneity: scale variance and pattern metrics. Geographic Information Sciences 6(1), 6-19.