
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Naimi, Babak; Hamm, Nicholas A. S.; Groen, Thomas; Skidmore, Andrew K.; Toxopeus, Albertus G.; Alibakhshi, Sara

ELSA: Entropy-based local indicator of spatial association

Published in:
Spatial Statistics

DOI:
[10.1016/j.spasta.2018.10.001](https://doi.org/10.1016/j.spasta.2018.10.001)

Published: 01/03/2019

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Naimi, B., Hamm, N. A. S., Groen, T., Skidmore, A. K., Toxopeus, A. G., & Alibakhshi, S. (2019). ELSA: Entropy-based local indicator of spatial association. *Spatial Statistics*, 29, 66-88.
<https://doi.org/10.1016/j.spasta.2018.10.001>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

1 *ELSA: Entropy-based Local Indicator of Spatial Association*

2

3 Babak Naimi^{1,2*}, Nicholas A.S. Hamm³, Thomas A. Groen⁴, Andrew K. Skidmore^{4,5}, Albertus G.

4 Toxopeus⁴, Sara Alibakhshi⁵

5

6 ¹ Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The
7 Netherlands;

8 ²Department for Migration and Immuno-ecology, Max Planck Institute for Ornithology, Am Obstberg
9 1, 78315, Radolfzell, Germany;

10 ³School of Geographical Sciences and Geospatial Research Group, Faculty of Science and Engineering,
11 University of Nottingham, Ningbo, China;

12 ⁴Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217,
13 7500 AE, Enschede, The Netherlands;

14 ⁵Department of Environmental Science, Macquarie University, NSW 2109, Australia

15 ⁶Department of Built Environment, School of Engineering, Aalto University, P.O. Box 14100, 00076
16 Espoo, Finland

17

18

19 ***Correspondence:** Babak Naimi, Institute for Biodiversity and Ecosystem Dynamics, University of
20 Amsterdam, Amsterdam, The Netherlands; Email: naimi.b@gmail.com

21

22 **Abstract**

23 Research on spatial data analysis has developed a number of local indicators of spatial association
24 (LISA), which allow exploration of local patterns in spatial data. These include local Moran's I and
25 local Geary's c , as well as G_i and G_i^* that can be used for continuous or interval variables only.
26 Despite numerous situations where qualitative (nominal/categorical) variables are encountered, few
27 attempts have been devoted to the development of methods to explore the local spatial pattern in
28 categorical data. To our knowledge, there is no indicator of local spatial association that can be used
29 for both continuous and categorical data at the same time.

30 In this paper, we propose a new local indicator of spatial association, called the entropy-based local
31 indicator of spatial association (ELSA), can be used for both categorical and continuous spatial data.
32 ELSA quantifies the degree of spatial association of a variable at each location relative to the same
33 variable at the neighbouring locations. This indicator simultaneously incorporates both spatial and
34 attribute aspects of spatial association into account. The values of ELSA vary between 0 and 1, which
35 denote highest and lowest spatial association, respectively. We compare ELSA to existing statistics
36 such as Local Moran's I and test the power and size of the new statistic. We also introduce the
37 "entrogram", a novel approach for exploring the global spatial structure within the entire area (like a
38 variogram). This study showed that the ELSA is consistent and robust, and is therefore suitable for
39 applications in a wide range of disciplines. The ELSA algorithm is made available as an R-package
40 (elsa).

41

42 1. Introduction

43 Spatial analysis is concerned with exploration and identification of associations over geographical
44 space. Such associations quantify the degree to which a value of a variable measured at one location
45 is dependent on the values of the same variable measured at a specific geographic distance from
46 that location (Cliff and Ord 1981; Goodchild 1986). If such dependency exists in a dataset, the
47 variable is said to exhibit spatial autocorrelation (Sokal and Oden 1978). Several statistics have been
48 developed to quantify spatial autocorrelation both globally and locally. Global measures provide a
49 statistic for the entire field under the assumption of spatial stationarity, *i.e.* the mean and
50 covariance do not vary over space. This assumption is often unrealistic. Recent advances have
51 addressed non-stationarity in the mean through spatially varying coefficient modelling (Finley 2011;
52 Gelfand et al. 2003; Hamm et al. 2015a). Heteroskedasticity in the variance can be addressed
53 through a weighting function (Hamm et al. 2012; Lark 2009) and further efforts have been directed
54 at modelling non-stationary covariance functions (Haskard and Lark 2009; Paciorek and Schervish
55 2006). Models that address non-stationary are often difficult to implement and there is a lack of
56 standard software tools. Further, these models all provide a global measure of autocorrelation. This
57 may be of limited relevance when local areas are of interest.

58 Research studies on spatial data analysis developed a number of local spatial statistics (Anselin 1995;
59 Boots 2003; Getis and Ord 1992; Ord and Getis 1995). In contrast to global measures, these
60 statistics allow exploration of local patterns in spatial association (Lloyd 2007). They do not rely on
61 the assumption of global stationarity. Anselin (1995) introduced a set of statistics, called local
62 indicators of spatial association (LISA), including local Moran's I and local Geary's c , that decompose
63 a single global measure into the contribution of each individual location to explore the locations that
64 are major contributors to the global autocorrelation. These statistics can be used to test if local
65 spatial clustering of similar values around the observation is significantly different from the global
66 mean. Getis and Ord (1992) and Ord and Getis (1995) also defined two local statistics, G_i and G_i^* ,

67 which are somewhat different from Anselin's LISAs, indicating local clustering of high and low values.
68 These allow detection of pockets of spatial association that may not be evident when using global
69 statistics (Getis and Ord 1992; Ord and Getis 1995). These methods, however, can be used for
70 continuous or interval variables only. Despite numerous situations where qualitative
71 (nominal/categorical) variables are encountered, only a few attempts have been devoted to develop
72 methods to explore the spatial pattern for categorical data. Examples include early works using joint
73 count statistics (Dacey 1968; Dale 2000; Iyer 1949; Moran 1948), extension of the Moran Coefficient
74 to a wide variety of probability models and linking continuous and discrete variables (Griffith 2010),
75 and more recent works including local indicators for categorical data (LICD) by Boots (2003) and
76 symbolic entropy by Ruiz et al. (2010). The purpose of LICD is to identify the nature and spatial
77 extent of local neighbourhoods that are distinctive or unusual compared to *a priori* expectation in
78 binary categorical data (Boots 2003, 2006). To our knowledge, there is no statistic of local spatial
79 association that can be used for both continuous and categorical data.

80 In this paper, we propose a new local indicator of spatial association, called the entropy-based local
81 indicator of spatial association (ELSA), for exploratory analysis and testing the significance of local
82 spatial association of both categorical and continuous spatial data. In doing so, we address the gaps
83 in the research literature highlighted above. Entropy has its root in thermodynamic and information
84 theory. Entropy based approaches have been applied in many disciplines, as a measure of
85 complexity in physics (Shannon and Weaver 1949), as a measure of diversity or structural complexity
86 in ecology (Anand and Orloci 1996; Ricotta and Anand 2006), and as a measure of information
87 content or uncertainty in information theory (Yeung 2008). This concept has also been used by
88 geographers, economists and social scientists to describe spatial phenomena (Batty 1974, 1976;
89 Heikkila and Hu 2006). Recently, some attempts have been conducted to use the concept of entropy
90 as a measure of spatial contiguity for qualitative data (Ruiz et al. 2010), or to detect spatially varying
91 multivariate relationships (Guo 2010). Matilla-García et al. (2011) highlighted the potential role that

92 entropy-based measures might play in detecting spatial structure. They used spatial symbolic
93 entropy for detecting the order of contiguity (spatial lag) of a spatial dependent process (Matilla-
94 García and Marín 2011).

95 Although entropy-based approaches have been used widely for categorical data (e.g. soil type,
96 classified data), there is a challenge in using entropy for exploratory analysis of continuous data. For
97 these data, the probability density function is often unknown, which is necessary to calculate
98 entropy (Guo 2010). A common solution to this problem in most practical applications is to construct
99 a contingency table by binning (discretizing) a continuous variable into a finite number of classes
100 (Guo 2003; Journel and Deutsch 1993), which then are used as categorical data in the entropy
101 measure. This, however, raises two new challenges. Firstly, binning continuous variables causes
102 information loss in data. Secondly, dissimilarity between binned data is not the same as it is in
103 categorical data. For example, when using a categorical land use map in spatial analysis, no
104 difference in the level of dissimilarity between pairs of classes is assumed. If a continuous variable
105 binned into, for example, five categories (C1, ..., C5), then C1 is more similar to C2 than to C5. In this
106 paper, we illustrate how ELSA addresses these challenges.

107 The main objectives of our study are: (1) to introduce a new statistic for measuring and testing local
108 spatial association (ELSA) that can be used for both continuous and categorical spatial data; (2) to
109 explore the application of ELSA for calculating local spatial association and comparing this with other
110 indicators; (3) to demonstrate the usability of ELSA to detect global patterns as well, and calculate a
111 variogram-like global spatial structure, named 'entrogram'. In addition, we developed a new
112 software package, 'elsa', in the R environment of statistical computing (R Development Core Team
113 2017) to freely provide the tools required for calculation of the ELSA statistic and the entrogram.

114

115

116 2. Entropy

117 In information theory, the concept of complexity is closely related to predictability. It is the amount
118 of information required to achieve an optimal prediction (Boschetti 2008). The entropy measure is
119 also known as the information content. The Shannon entropy has been defined as an average
120 amount of information to eliminate uncertainty, given by a finite number of events:

$$121 \quad H = - \sum_{k=1}^m p_k \log_2 p_k \quad (5)$$

122 where H measures the entropy of a system with a finite number of m possible events, and p_k
123 represents the probability of event k . H is at a maximum when all events occur in equal abundance
124 and can be quantified by $\log_2 m$. This measure can be standardized by dividing by $\log_2 m$, providing
125 a measure of relative entropy ranging between 0 and 1. The function is dimensionless and depends
126 only upon the number of events, not upon any other invariant property of the system to which it is
127 applied (Batty 1976).

128 3. ELSA statistic

129 The Entropy based Local indicator of Spatial Association, ELSA, extends the above described entropy
130 measure using a term that summarizes the attribute distance between a location and its
131 neighbourhood locations over a given geographical distance.

132 Assume $x = (x_1, x_2, \dots, x_n)'$ are n observations related to a spatial process at locations $u =$
133 $(u_1, u_2, \dots, u_n)'$. Further, denote by $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)$ the set of possible categories that x_i can take.
134 For a categorical variable, it is usually assumed that pairs of categories are equally dissimilar. There
135 are, however, situations where the level of dissimilarity varies for different pairs of categories. For
136 example, 'dense forest' and 'sparse forest' in a land use map are more similar than a pair of either of

137 these two classes and 'lake' (for more details, see the Section "ELSA for categorical data"). Likewise,
 138 when the values of a continuous variable are binned (discretized) into categories, the level of
 139 dissimilarity varies between categories. By ranking the binned values, the difference between rank
 140 numbers can be interpreted as the level of dissimilarity. These levels of dissimilarity between
 141 categories are taken into account in the calculation of ELSA.

142 ELSA (*E* statistic) at site *i* is defined as:

$$E_i = E_{ai} \times E_{ci} \quad (6)$$

$$E_{ai} = \frac{\sum_j \omega_{ij} d_{ij}}{\max\{d\} \sum_j \omega_{ij}}, j \neq i$$

$$E_{ci} = - \frac{\sum_{k=1}^{m_\omega} p_k \log_2(p_k)}{\log_2 m_i}, j \neq i$$

$$m_i = \begin{cases} m & \text{if } \sum_j \omega_{ij} > m \\ \sum_j \omega_{ij}, & \text{otherwise} \end{cases}$$

$$d_{ij} = |c_i - c_j|$$

143 where ω_{ij} is a binary weight which specifies whether the site *j* is within a specified distance (defines
 144 the neighbourhood size) from site *i*. d_{ij} describes the dissimilarity between x_i and x_j , which is
 145 calculated as the absolute difference of the ranks assigned to the categories at sites *i* and *j* (i.e., c_i
 146 and c_j), and $\max\{d\}$ is the maximum possible dissimilarity between any pair of observations in the
 147 entire dataset. This is discussed for continuous and categorical variables in the upcoming sections.
 148 There are *m* categories in the entire dataset, p_k is the probability of k^{th} category from the m_ω

149 categories within the local distance from site i , and m_i is the maximum possible number of
150 categories within the local distance from site i . This means that if the number of observations within
151 the local distance from site i , including site i , is greater than the number of categories in the entire
152 dataset ($\sum_j \omega_{ij} > m$), then m_i is equal to the number of categories, otherwise it is equal to the
153 number of observations ($\sum_j \omega_{ij}$) within the local distance from site i .

154 The first term in the equation 6 (E_{ai}) summarizes the attribute distance (dissimilarity) between site i
155 and the neighbouring sites. This coefficient is bounded between 0 and 1. Low values indicate high
156 similarity of site i to neighbouring sites, and high values indicate low similarity with neighbouring
157 sites.

158 The second term of the E_i statistic (i.e., E_{ci}) is the Shannon entropy (equation 5), normalized by
159 $\log_2(m_i)$. This term ranges between 0 and 1. By normalizing, values are invariant to the number of
160 categories present in a dataset. In other words, datasets with different numbers of categories are
161 comparable. For normalizing, m_i is defined in relation to the global number of categories in the
162 entire dataset. This term quantifies composition or diversity of the categories within the local
163 distance from site i , but it is not sensitive to the level of dissimilarities between pairs of
164 observations.

165 Fig. 1 shows the behaviour of the two terms in ELSA as well as ELSA itself at site i in the centre of a 3
166 x 3 window under five scenarios. When a location is surrounded by similar locations (example a),
167 both the dissimilarity and composition statistics (i.e., E_{ai} and E_{ci} , respectively) would be calculated
168 as 0. The composition in the binary maps in examples b and c, is the same, while the dissimilarity
169 between site i and its neighbour locations in the two maps is different. In example b, the site is
170 surrounded by 8 dissimilar locations to the site, while in the example c, the site is surrounded by 7
171 similar and only one dissimilar locations. Therefore, E_{ai} is at its maximum value in example b, while
172 it is much lower in example c. In the two maps (examples d and e) with three categories, the

173 dissimilarity is the same for both as the site i is surrounded by 8 dissimilar locations to the site. The
174 composition, however, is different in these two maps.

175 **[Fig. 1]**

176 These five examples show the complementarity of the two terms in the ELSA statistic in different
177 situations. The role of these terms would even be more important if the categories have different
178 levels of dissimilarity (see the upcoming section). It also suggests that either of these terms would be
179 informative on their own for the purpose they are designed for. For instance, if one needs to identify
180 locations with extreme or outlier values, compared to their neighbour locations, the E_{ai} term in
181 ELSA would be the statistic that fulfils the need, while if only the composition (i.e., diversity) of the
182 values within a local neighbourhood is the purpose of the study, the E_{ci} term in ELSA can be used.

183

184 3.1. ELSA for categorical data

185 Categorical variables are typically conceptualized as having no inherent ordering (Ahlqvist and
186 Shortridge 2010), which means that all pairs of categories are equally dissimilar. When using this
187 simplification of class differences and denoting the dissimilarity as d (as in equation 6): $\max(d) = 1$,
188 $d_{ij} = 1$ if $x_i \neq x_j$ and $d_{ij} = 0$ if $x_i = x_j$.

189 The assumption that all categories are equally dissimilar is often an oversimplification. Several
190 studies have been conducted to estimate a measure of dissimilarity between categories (Romme
191 1982; Uemaa et al. 2008). Categories may also be classified into a hierarchical structure. For
192 example, the United Nations Food and Agricultural Organization's (FAO) land cover classification
193 system arranges classes hierarchically (Di Gregorio and Jansen 2009). Such hierarchies may be used
194 to describe the level of dissimilarity. Consider a categorical map with four categories: 'mixed forest' (α_1),
195 'coniferous forest' (α_2), 'olive groves' (α_3) and 'vineyards' (α_4). These can be grouped under

196 primary categories such that 'mixed forest' and 'coniferous forest' belong to the primary category
 197 'forests' (denoted β_1) and 'olive groves' and 'vineyards' belong to 'agricultural areas' (denoted β_2).
 198 To calculate E_i for this categorical map, c_i (the rank number for site i) is set to 1 (always), and c_j is
 199 set to 1 (if sites i and j are the same), or 2 (if the categories in sites i and j are different, but belong
 200 to the same primary category), or 3 (if the categories in sites i and j are different and also belong to
 201 different primary categories). Consequently, the level of dissimilarity, d_{ij} , between two different
 202 subcategories of the same primary category is set as $d_{ij} = 1$ (e.g., $d(\text{coniferous forest, mixed forest})$
 203 $= d(\alpha_1, \alpha_2) = 1$), and between two different subcategories from different primary categories is set
 204 as $d_{ij} = 2$ (e.g., $d(\text{coniferous forest, vineyards}) = d(\alpha_1, \alpha_4) = 2$).

205 To develop this algorithm, we specify the dissimilarity for categorical data as

- 206 1) If $\alpha_i = \alpha_j$ then $c_i = 1$ & $c_j = 1$ therefore $d_{ij} = 0$
- 207 2) If $(\alpha_i \neq \alpha_j) \& (\beta_i = \beta_j)$ then $c_i = 1$ & $c_j = 2$ therefore $d_{ij} = 1$
- 208 3) If $(\alpha_i \neq \alpha_j) \& (\beta_i \neq \beta_j)$ then $c_i = 1$ & $c_j = 3$ therefore $d_{ij} = 2$

209 The level of dissimilarity between pairs of categories can be illustrated in a matrix (Table 1). This
 210 makes ELSA flexible enough to handle situations where the level of dissimilarity can be specified for
 211 pairs of categories or when categories are hierarchically ordered.

212 **[Table 1]**

213 The above situation can be extended to the case where there are more levels in the hierarchy. The
 214 maximum level of dissimilarity is then equal to the degree of hierarchy. In Fig. 2, the level of
 215 dissimilarity is presented schematically for a hierarchical system that contains 3 levels.

216 **[Fig. 2]**

217

218 3.2. ELSA for continuous data

219 A key step to calculate ELSA for continuous data is that the variable should be first categorized
220 (binned or discretized) into a number of categories; a procedure that may cause information loss.
221 Inspired by Morrison (1972), we propose an estimation of the optimum number of categories that
222 minimizes the information loss. This optimum is the minimum number of categories that is able to
223 reproduce the spatial data statistically (i.e., that minimizes the loss of information through
224 categorization). To find the optimum number of categories, our procedure uses Spearman's rank
225 correlation coefficient, ρ , as a measure of information between the continuous variable and the
226 categorized variable. If the amount of information is not affected through categorizing, the observed
227 correlation should be equal to one. Any loss of information would result in the observed correlation
228 to be less than one. Therefore, the magnitude of the difference $1 - \rho$ provides a measure of
229 information loss (Quester and Dion 1997). The procedure of selecting the optimum number involves
230 the following steps:

- 231 1) The categorization procedure starts with a minimum number of categories $m = 2$
- 232 2) The procedure assigns a rank number (between 1 and m , where m is the total number of
233 categories) to each category.
- 234 3) The ρ coefficient between the continuous values and the assigned ranks is calculated for
235 each iteration.
- 236 4) The steps 1 to 3 are repeated, every time by considering one more category (i.e., increasing
237 m), until a convergence threshold (e.g., 0.005) is reached. The convergence is defined as the
238 difference between the ρ coefficients of the current and previous iterations.
- 239 5) The one-standard-error rule (James et al. 2013) is applied to select the optimum number of
240 categories. First, the standard error of the ρ coefficients is calculated, and then the optimum
241 number of categories would be the lowest number for which the ρ coefficient is within one
242 standard error of the highest ρ coefficient.

243 It is assumed that the information loss due to the categorization is not substantial when the
244 optimum number is used. Fig. 3, illustrates an example of using this procedure for a continuous
245 variable.

246 **[Fig. 3]**

247 To calculate ELSA for the continuous map using equation 6, the original continuous data \mathbf{x} are then
248 mapped into the ranked categories: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)'$ with ranks c_1, c_2, \dots, c_m , where $c_1 = 1$ and c_m
249 $= m$. The maximum level of dissimilarity in the entire map is therefore: $\max(d) = c_m - 1$ and the
250 dissimilarity between the categories at two locations \mathbf{u}_i and \mathbf{u}_j is $d_{ij} = |c_i - c_j|$.

251

252 **4. Inference for ELSA statistic**

253 In this section we propose a non-parametric bootstrap randomization approach to test the local
254 spatial association against a null distribution. The approach is based on repeated resampling from a
255 distribution, \hat{F}_0 , which satisfies the relevant null hypothesis (Davison and Hinkley 1997). Suppose α
256 $= (\alpha_1, \alpha_2, \dots, \alpha_m)$ are the possible events, outcomes of a spatial process, that n observations $\mathbf{x} =$
257 $(x_1, x_2, \dots, x_n)'$ can take at locations $\mathbf{u} = (u_1, u_2, \dots, u_n)'$. The null surface can be constructed by
258 rearranging or shuffling the locations (Anselin 1995). Once the null surface \hat{F}_0 is constructed, a
259 Monte Carlo simulation with R runs is used to draw a sample with size n from the null distribution
260 through a bootstrap resampling procedure (sampling with replacement) for each run. The observed
261 ELSA statistic at site i (E_i) can be compared to R independent values of the statistic obtained from
262 the corresponding samples independently simulated under the null hypothesis (i.e., no spatial
263 autocorrelation). If these simulated values at site i are denoted by $E_{1i}^*, \dots, E_{Ri}^*$, then the probability of
264 accepting the null hypothesis (P -value) at site i (P_i) can be approximated by (Davison and Hinkley
265 1997):

$$P_i = \frac{1 + \#\{E_i \geq E_{ir}^*\}}{R + 1} \quad (7)$$

266 where $\#\{E_i \geq E_{ir}^*\}$ indicates number of times the observed ELSA at site i is greater than or equal to
 267 the ELSA values calculated for the bootstrap samples drawn from the null distribution.

268 4.1. Size and power of the test

269 We analysed the size and power of our non-parametric test based on the ELSA statistic to investigate
 270 whether the test correctly rejects the null hypothesis only when it should be rejected (Bivand 2009).

271 To do so, we conducted a comprehensive set of data simulations to generate both continuous and
 272 categorical variables with various levels of no to high spatial autocorrelation. We used an
 273 unconditional simulation to construct regular grids of 50×50 grid cells for each variable.

274 Unconditional simulation is a geostatistical technique that generates a realization of a spatially
 275 correlated variable, where the spatial correlation is defined by a variogram (Dungan 1999). The
 276 circulant-embedding algorithm (Dietrich and Newsam 1993) implemented in the RandomFields
 277 package v. 3.0.44 (Schlather 2009) in the R programming environment, v.3.0.1, was used to conduct
 278 the unconditional simulation. The effective variogram range (the maximum geographic separation
 279 where two points are expected to be autocorrelated) is determined by the scale parameter, ϕ . The
 280 variability is determined by the sill, $C + C_0$, where C is the partial sill and C_0 is the nugget and $\gamma =$

281 $C/(C + C_0)$ is the proportion of the variability that has spatial structure. An exponential variogram
 282 model with an effective autocorrelation range of $3\phi = 50$ cells and an arbitrary sill of 10 was used
 283 for all datasets. We controlled the degree of spatial autocorrelation by changing γ in the variogram
 284 models. Five levels of $\gamma = 0, 0.25, 0.5, 0.75,$ and 1 were used to define the variogram models (Fig. 4),
 285 giving a transition from no to high spatial autocorrelation, respectively. To generate the categorical
 286 variable, we followed a second step where the continuous spatially autocorrelated variable Y was

287 used to define a discrete spatial process as follows Ruiz et al. (2010). Let b_{ik} be i^{th} breaking point in
 288 discretizing Y into k categories, and defined as:

$$p(Y \leq b_{ik}) = \frac{i}{k} \quad \text{with } i < k \quad (8)$$

289 Let $A = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ and define the discrete spatial process as:

$$X_u = \begin{cases} \alpha_1 & \text{if } Y_u \leq b_{1k} \\ \alpha_i & \text{if } b_{i-1k} \leq Y_u \leq b_{ik} \\ \alpha_k & \text{if } Y_u \geq b_{k-1k} \end{cases} \quad (9)$$

292 where X_u is the value of discretized Y at location u . For each level of spatial autocorrelation, we
 293 generated three sets of categorical variables with $k=2, 3,$ and 4 categories, and a set of continuous
 294 variables (Fig. 5).

295 **[Fig. 4]**

297 **[Fig. 5]**

298 For each set of data, 999 replicates were simulated and a test with 999 (=R) runs was applied to each
 299 replicate at a level of significance $\alpha = 0.05$. We applied the test at the center of the image with
 300 different local neighbourhood sizes $Ne = 1.5, 3, 5, 10,$ and 15 cells. We then recorded the number of
 301 times that the null hypothesis is rejected (i.e., when $P \leq 0.05$) to quantify the rejection rate. We
 302 would expect that the statistic fails to reject the null hypothesis most of the times when the level of
 303 spatial autocorrelation is zero (size of the statistic). At the same time, we would expect that it rejects
 304 the null hypothesis more frequently (power of the statistic) as the level of autocorrelation goes up
 305 (Ruiz et al. 2010).

306

[Table 2]

307 The results of the numerical analyses are illustrated in Table 2 for different settings. For the all cases,
308 the power and size of the test showed that the statistic performed reasonably well. The power of
309 the test goes up when the level of spatial autocorrelation increases. The power was slightly higher
310 for the continuous variables compared to the categorical variables. The size was also slightly higher
311 for the continuous variables with smaller neighbourhood sizes, indicating a slightly greater risk for
312 false positive compared to the categorical variables. The results showed that increasing the number
313 of categories in the categorical variables has less effect on the power and size of the statistic. Not
314 surprisingly, there is a loss in the power of the statistic when the neighbourhood size is increased
315 because as the distance is increased, the level of spatial association in decreased. The reason is that
316 as the distance is increased, the level of spatial association is decreased (i.e., increasing the variance
317 in the variogram model).

318

319 **5. Application of ELSA to assess local spatial association**

320 In this section, we illustrate the use of ELSA by means of several examples using real and synthetic
321 datasets, covering both continuous and categorical variables.

322 *5.1. Experiment with categorical data with the same level of dissimilarity between classes*

323 A land cover map in a raster layer including 2769 grid cells of 1 x 1 km from southern Spain was used
324 for this experiment (Fig. 6-a). The map consists of 6 land cover classes. It is assumed that the level of
325 dissimilarity between pairs of classes is equal. In this experiment, the ELSA was calculated at each
326 cell within a local distance of 5 km (Fig. 6-b). As it was expected, the ELSA statistic becomes 0 for
327 homogenous areas with only one class (mostly at the western part of the area), and when the area

328 becomes more heterogeneous, the ELSA value becomes higher as it is obvious at the center and
329 eastern parts of the area.

330 **[Fig. 6]**

331 The three specified locations in Fig. 6 (i.e., A, B, and C) and the ELSA values at these locations show
332 how this statistic changes when the landscape changes. By looking at these locations on the land
333 cover map, it can be recognized that the level of heterogeneity changes from low to high from
334 location A to location C.

335

336

337 *5.2. Experiment with categorical data with non-equal level of dissimilarity between classes*

338 There are numerous situations where the level of dissimilarity between pairs of classes in a
339 categorical variable is not equal. To show how to deal with these situations, we illustrate two
340 experiments using both synthetic and real data.

341 *-Synthetic data:* We generated a synthetic categorical raster consisting of 1024 grid cells (32 rows ×
342 32 columns) and four classes (i.e., A1, A2, B1, and B2). The first and second two classes belong to the
343 main categories of A and B, respectively. So, A1 is more similar to A2 than to B1 or B2. The level of
344 dissimilarity between different pairs from the same main category (i.e., A1-A2 or B1-B2) is set to 1,
345 and from different main categories (e.g., A1-B1, A1-B2, etc.) is set to 2 (Fig. 7).

346

347 **[Fig. 7]**

348

349 The region is divided into four zones with different combinations of categories. This shows how ELSA
350 changes under these controlled situations. Zone 1 includes only one class (i.e., A1), Zone 2 includes a
351 random distribution of A1 and A2 (i.e., from the same primary category), Zone 3 includes a random
352 distribution of A1 and B1 (i.e., from two different primary categories), and Zone 4 includes random
353 distribution of all four categories.

354 The maximum level of dissimilarity ($\max \{d\}$) for this experiment is 2. We calculated ELSA at each
355 grid cell using a 3×3 window as the local neighbourhood (including diagonal neighbours, i.e.,
356 queen's case). We also calculated the mean ELSA for each zone to provide a base for comparison
357 (Fig.8). The results show that the mean ELSA over the two extreme zones are the minimum (Zone 1)
358 and maximum (Zone 4), respectively. The other two zones both follow the same structure (a random
359 distribution of the two classes, but with a different attribute distances). Since Zone 2 consists of two
360 more similar classes than Zone 3, it is expected that the ELSA statistic for the Zone 2 should be lower
361 than Zone 3. This is backed up by the empirical results (Fig. 8-b and Fig. 8-c).

362

363

[Fig. 8]

364

365 - *Real data*: We used the CORINE [Coordination of Information on the Environment of the European
366 Environmental Agency (EEA 2007)] 2006 land cover map from central Spain including 392336 grid
367 cells of 250 x 250 m. The land cover classes in the map were described using a hierarchical scheme
368 with three levels. The first level indicates the primary land cover class (e.g., agricultural area), which
369 are subdivided to more specific types of land cover classes at the second and third levels (e.g.,
370 permanent crops and vineyards at the second and third levels, respectively). A three-digit code is
371 used for each land cover, specifying the class at the three levels from left to right (e.g., 221 and 223

372 are 'vineyards' and 'olive groves' respectively, i.e., class 1 and 3 specified at the third level,
373 respectively, but both belong to the same classes of 'agricultural areas' [class 2] at level 1, and
374 'Permanent crops' [class 2] at level 2). A list of the classes at the three levels for the codes is
375 provided in Table A1 in Appendix A.

376 **[Fig. 9]**

377 The categories were ordered hierarchically at the three levels based on the three-digit codes. The
378 attribute distance (level of dissimilarity) between each pair of categories was calculated based on
379 their position on the hierarchical scheme. The maximum level of dissimilarity is 3 (e.g., between class
380 132 and 211). The dissimilarity between pairs of classes is illustrated in Fig. A1 (the table) & A2 (the
381 hierarchical view) in Appendix A.

382 We calculated ELSA at each grid cell within a local distance of 5 km (Fig. 10). A visual interpretation
383 of the three specified locations on the land cover map (Fig. 9) show that the local association among
384 the three locations is expected to be minimum and maximum at the locations B and C, respectively.
385 The values of the ELSA map at these locations are consistent with the visual interpretation.

386

387 **[Fig. 10]**

388

389 *5.3. Experiment with continuous data*

390 We simulated two synthetic continuous rasters with two levels of no and positive spatial
391 autocorrelation. We used unconditional simulation (see Section 4.1) to construct regular grids of 30
392 × 30 cells for each variable (Fig. 11 & 12). For each raster, we calculated a Moran scatterplot
393 (Anselin 1993), along with global Moran's I (Moran 1948) to quantify the global spatial

394 autocorrelation and summarize the overall pattern of spatial structure. We calculated ELSA at each
395 cell given a queen contiguity neighbourhood. For comparison, we also quantified other commonly
396 used local indicators of spatial association; G_i^* , local Moran's I and local Geary's c . These statistics
397 were compared pairwise to explore to what extent these measures are related. A Spearman
398 statistic was used to test and estimate a rank-based measure of association test between each pair
399 of statistics.

400 **[Fig. 11]**

401

402 **[Fig. 12]**

403

404 The results from the comparison between different local statistics (Fig. 11 & 12) indicate that when
405 the global spatial autocorrelation is positive, ELSA is only related significantly to local Geary's c ($\rho =$
406 0.675) while the other local indicators are not related either to ELSA or to each other according to
407 the pairwise comparison. When there is no global spatial autocorrelation, ELSA is related
408 significantly to both local Geary's c and local Moran's I . The former is a positive ($\rho = 0.807$), and the
409 later is a negative relationship ($\rho = -0.563$). Local Geary's c is also negatively related to local Moran's
410 I ($\rho = -0.509$). The results showed that the Local G^* statistic is not related to any of the other
411 statistics including ELSA. It has been argued that local statistics are sensitive to the existence of
412 global spatial autocorrelation in the dataset (Ord and Getis 2001), that might be the reason to have
413 varying behaviours between the different cases with the positive and no global spatial
414 autocorrelation. For example, we expect a negative relationship between local Moran's I and local
415 Geary's c , that is shown in the results only when there is no global spatial autocorrelation. The
416 relationship between ELSA and local Geary's c is also stronger when there is no autocorrelation

417 compared to the case with positive autocorrelation. The maps of P -values based on both ELSA and
418 local Moran's I showed that ELSA performed better by resulting in lower P -values for most of the
419 grid cells compared to the local Moran's I results, for the case with the positive spatial
420 autocorrelation, that might be because of the sensitivity of the local Moran's I to global spatial
421 autocorrelation.

422

423 **6. Application of ELSA to assess global spatial structure**

424 In this section we explore if global spatial structure can be assessed by employing ELSA into a
425 procedure of generating a sample variogram-like diagram, called 'entrogram'. The idea is that by
426 assessing ELSA with increasing local distances (lags) and putting the averaged values against these
427 distances in a diagram, we can explore spatial structure. The sample variogram is a well-known
428 approach for exploring spatial structure in continuous variables or binary categorical variables (i.e.
429 indicator variogram; Journel 1983). The semantic sample variogram has been developed for
430 multinomial categorical variables (Ahlqvist and Shortridge 2006). If ELSA can be used for this
431 purpose, then the spatial structure for both continuous and categorical variables can be explored
432 with the same technique. The entrogram for lag distance h is calculated as the mean of the ELSA
433 statistics at different sites within the distance equal to the lag size. The entrogram is calculated as
434 follows:

$$435 \quad E(h) = \frac{\sum_{i=1}^{n_h} E_i(h)}{n_h} \quad (7)$$

436 where $E(h)$ is the value of the entrogram for distance class h , $E_i(h)$ is the ELSA statistic at site i
437 within local distance h , n_h is the total number of sites within the distance h for which the ELSA
438 statistic is calculated.

439 In the following experiments we explored the behaviour of the entrogram for both continuous and
440 categorical data.

441

442 *6.1. Entrogram for categorical data*

443 We generated five synthetic categorical maps, on a 20 x 20 raster grid, to explore the capability of
444 the entrogram for calculating the spatial structure of categorical maps. The first three maps were
445 binary, one with randomly distributed classes, and the other two with spatially structured classes.
446 These binary maps provide the opportunity to compare the entrogram with the existing method of
447 exploring the spatial structure for a binary categorical variable (i.e. an indicator variogram). The two
448 spatially structured binary maps were constructed using an unconditional simulation with spherical
449 variogram models varying in their autocorrelation range. We used the range of 3 and 8 grid cells to
450 construct the maps with relatively low and high spatial autocorrelation, respectively (Fig. 13).

451 We quantified both the entrogram, using our developed R package (i.e. *elsa*), and the indicator
452 variogram using the *gstat* package v. 1.1-3 (Pebesma 2004) in the R development environment (R
453 Development Core Team 2017). For both, we used a lag size equal to one grid cell and the cutoff
454 values (number of lags) equal to 12 grid cells. The graphs are then visually interpreted and compared
455 (Fig. 12). The results showed that the entrograms indicate the same patterns as the indicator
456 variograms, and therefore can be used to interpret global spatial structure in binary categorical
457 maps.

458

459

[Fig. 13]

460

461 The last two categorical maps were generated with four classes. We assumed that these four classes
462 were equally dissimilar. The classes in the first map were spatially structured, giving a maximum
463 degree of spatial clustering, while in the second map they were distributed randomly. This dataset
464 allowed us to illustrate the capability of the entrogram for exploring the spatial structure of
465 multinomial data. We quantified the entrogram with a lag distance equal to one grid cell and the
466 cutoff value (number of lags) equal to 12 grid cells. The visual interpretation of the graphs (Fig. 14)
467 for the spatially clustered map shows that the mean ELSA value is low within the lower distances and
468 increases when the lag distance increases as would be expected. For the randomly distributed
469 classes, on the other hand, this value remains at the maximum level, showing there is no spatial
470 structure.

471

[Fig. 14]

472

473 6.2. Entrogram for continuous data

474 We generated five continuous raster maps with different ranges of spatial autocorrelation. We used
475 unconditional simulation to construct regular grids of 200×200 cells for each variable. Two
476 variogram models, a Spherical and a Gaussian model, were used with varying values for the sill ($C +$
477 C_0), nugget (C_0) and scale (ϕ) parameters for different maps. We assigned ϕ different values
478 including $\phi = 10$ and $\phi = 30$ grid cells giving relative low and high values for the spatial
479 autocorrelation. A total sill, $C + C_0 = 10$, was used for all the datasets, with $C_0 = 0$ for two maps and
480 $C_0 = 3$ and $C_0 = 5$ for the other two maps. Additionally, a white-noise surface ($\phi = 0$) was simulated,
481 giving a total of five maps with different levels of spatial autocorrelation. We then quantified and

482 visualized the empirical variogram and entrogram for each surface (Fig. 15). The visual comparisons
483 of the graphs showed that the entrogram describes a global spatial structure similar to the empirical
484 variogram.

485

486 **[Fig. 15]**

487

488 **7. Discussion**

489 ELSA allows for exploring and testing the local spatial association for both categorical and
490 continuous variables. This provides the opportunity of using one statistic for a study where both
491 types of variables are used, e.g., in species distribution modelling (Naimi and Araújo 2016; Naimi et
492 al. 2014; Naimi et al. 2011), environmental epidemiology (Araujo Navas et al. 2016; Hamm et al.
493 2015b), and for predicting soil properties (Hengl et al. 2015). The ELSA statistic measures the local
494 spatial associations within the same range (between 0 and 1) for both types of data, making the
495 outputs comparable.

496 We developed a nonparametric bootstrap test based on the ELSA statistic that is useful to test the
497 hypothesis of independence among spatially distributed quantitative (continuous) and qualitative
498 (categorical) data. Our extensive experiments to demonstrate the size and power of the statistic
499 under a range of conditions suggest the usefulness of the statistic for measuring the local spatial
500 association. In addition, we compared the results of our ELSA test based on a Monte Carlo
501 simulation with the well-known Moran's I test. The ELSA test behaved consistent when the global
502 spatial autocorrelation was either positive or zero. The other LISAs showed inconsistent behaviour,
503 i.e., local Moran's I and local Geary's c only showed the expected negative relationship when the

504 global autocorrelation was zero. The same holds with the p-values generated by the local Moran's *I*
505 (Fig. 11 & 12). It has been argued that testing the hypothesis of spatial dependence based on LISAs
506 will give incorrect significance levels in the presence of global spatial association (Anselin 1995). This
507 explains why the tests based on local Moran's *I* statistic did not generate the expected p-values
508 when the global spatial autocorrelation was positive. In contrast, our results showed that the ELSA
509 statistic was not sensitive to the presence of the global spatial association that can be considered as
510 an advantage of ELSA over the LISA statistics.

511 The ELSA calculation for categorical data supports different levels of dissimilarities. This allows
512 evaluation of the dissimilarity between nominal categories in a graded fashion. Several authors in
513 the field of landscape ecology have considered graded differences for measuring landscape
514 patchiness (DeGraaf and Yamasaki 2002; Desrochers et al. 2003). Use of this approach allows
515 transformation from a nominal to an ordered or numerical scale, and provides a foundation for
516 handling attribute distances for categorical data. This is important when quantifying spatial structure
517 for categorical data, since without this aspect all categories are considered equally dissimilar. This
518 approach, however, relies on the subjective evaluation of the class similarity (Ahlqvist and
519 Shortridge 2010), and requires additional data and a guiding theory. Multi-criteria decision making,
520 the analytic hierarchy process, and conjoint analysis are well-known frameworks for evaluating class
521 similarities in categorical maps (Ahlqvist and Shortridge 2010; Schwering 2008). Further studies are
522 needed to characterize and compare these methods, but in this study, we introduced a more
523 general and simple approach to handle class dissimilarities based on a hierarchical scheme of
524 classes. Given additional knowledge about the meaning of class definitions, a more formal
525 evaluation of the categorical dissimilarities can be used to calculate the ELSA statistic for categorical
526 maps.

527 To measure spatial association, a majority of studies were devoted to the development of statistics
528 for continuous data. In this study, we addressed three of these commonly used statistics and

529 explored how ELSA relates to them. Although all of these statistics have been used widely as
530 measures of local spatial associations, they quantify different properties. Therefore, the appropriate
531 technique for identification of the spatial association should correspond to the nature of the
532 question concerning dependence/independence (Getis and Ord 1996). Our results confirmed these
533 differences and revealed that ELSA is more related to local Geary's c . Local Geary's c (like in a
534 variogram) is based on differences between pairs of observations. Analogous to variance, entropy is
535 a measure of dissimilarity and diversity, and their equivalence has been explored and discussed in
536 several studies (e.g. Ebrahimi et al. 1999; Lindley 1956).

537 Despite the early specific work in geographical analysis (Batty 1974), entropy has been mainly used
538 in other fields, perhaps mostly in the fields of physics and information theory. The most relevant
539 research (to spatial data) in these disciplines introduced several extensions of the entropy measure
540 that apply to calculating the structural complexity or patterns of two-dimensional dynamical systems
541 (Feldman and Crutchfield 2003; Robinson et al. 2011). In recent years, there were several efforts to
542 develop some entropy-based methods for detecting (global) spatial association and patterns of
543 complexity for univariate data (Matilla-García and Marín 2011; Matilla-García et al. 2012; Pham
544 2010; Ruiz et al. 2010), or as a spatial weighted information measure of global and local spatial
545 association (Karlström and Ceccato 2000), or to discover different forms of local multivariate
546 relationships (Guo 2010). Similar to these methods, ELSA is also based on the entropy measure, but
547 can be considered as a novel approach that offers some unique features, as described in the
548 manuscript.

549 Together with ELSA, this study introduced the entrogram, an approach for exploring the global
550 spatial structure within the entire area (like a variogram). Such explorations are applied in many
551 fields, such as landscape ecology, geography, and soil science. The entrogram uses the ELSA statistic,
552 and therefore, has the advantage over variogram that can be used for different forms of categorical
553 data (e.g., binary-, multinomial-, or hierarchical-classes) as well as continuous data. The indicator

554 variogram is a known technique to measure the spatial variability of classes in categorical variables.
555 However, it has been argued that the binary treatment of categorical variables in this technique is an
556 unnecessary oversimplification, and that it should be replaced by ordered measures based on
557 semantic similarity evaluations (Ahlgvist and Shortridge 2010). Semantic variograms (Ahlgvist and
558 Shortridge 2006) were developed based on this concern and provide the capability to consider
559 semantic distance between categories in the calculation. We showed that the entrogram is capable
560 of this as well.

561 We showed that the entrogram can be used as an exploratory tool to characterise global spatial
562 structure. Variograms, however, can also be used for geostatistical modelling and prediction
563 (kriging). Although this study introduced the entrogram as an exploratory tool, future studies may
564 focus on investigating the use of both ELSA and the entrogram for spatial modelling. Moreover, we
565 believe that the ELSA statistic can be adapted for measuring local temporal autocorrelation, local
566 spatio-temporal autocorrelation and multivariate spatial and spatio-temporal autocorrelation.
567 These might be considered as the areas for future research.

568 Although raster datasets were used in this study to simplify the illustrations, ELSA can be quantified
569 for other types of spatial data (i.e., spatial points and polygons). Our new R package, `elsa`, offers the
570 functionalities to quantify ELSA, entrogram (as well as variogram and the other LISA statistics) for all
571 the spatial data types. The software is developed using both the R and C programming languages to
572 speed up the computations. All the functionalities in the package are followed by the help pages
573 where their usage and the relevant details, including some examples, are provided.

574

575 **8. Conclusion**

576 This paper focused on the development of an entropy-based statistic (ELSA) for the quantification of
577 local spatial association. The ELSA statistic presented in this paper showed to be a robust and
578 reliable method for identifying and testing the degree of spatial association. The method provides
579 the advantage of using one statistic for both continuous and categorical data. This makes
580 comparisons between spatial structures of both types of data possible. Another advantage of ELSA,
581 compared to the LISA statistics, is that it is not sensitive to the presence of the global spatial
582 autocorrelation. In addition, this statistic provides the ability to incorporate both spatial and
583 attribute aspects of spatial association into the statistic for both continuous and categorical data. It
584 can be tested for significance, as demonstrated by a power test in this article. By introducing the
585 'entrogram' we demonstrated that ELSA can also be used to measure global spatial structure of both
586 continuous and categorical data.

587 **References**

- 588 Ahlqvist, O., & Shortridge, A. (2006). Characterizing Land Cover Structure with Semantic Variograms.
589 In A. Riedl, W. Kainz, & G. Elmes (Eds.), *Progress in Spatial Data Handling* (pp. 401-415): Springer
590 Berlin Heidelberg
- 591 Ahlqvist, O., & Shortridge, A. (2010). Spatial and semantic dimensions of landscape heterogeneity.
592 *Landscape Ecology*, 25, 573-590
- 593 Anand, M., & Orloci, L. (1996). Complexity in plant communities: The notion and quantification.
594 *Journal of Theoretical Biology*, 179, 179-186
- 595 Anselin, L. (1993). *The Moran scatterplot as an ESDA tool to assess local instability in spatial*
596 *association*. Regional Research Institute, West Virginia University Morgantown, WV
- 597 Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27, 93-115
- 598 Araujo Navas, A.L., Hamm, N.A.S., Soares Magalhães, R.J., & Stein, A. (2016). Mapping Soil
599 Transmitted Helminths and Schistosomiasis under Uncertainty: A Systematic Review and Critical
600 Appraisal of Evidence. *PLOS Neglected Tropical Diseases*, 10, e0005208
- 601 Batty, M. (1974). Spatial Entropy. *Geographical Analysis*, 6, 1-31
- 602 Batty, M. (1976). Entropy in Spatial Aggregation. *Geographical Analysis*, 8, 1-21
- 603 Bivand, R. (2009). Applying Measures of Spatial Autocorrelation: Computation and Simulation.
604 *Geographical Analysis*, 41, 375-384

605 Boots, B. (2003). Developing local measures of spatial association for categorical data. *Journal of*
606 *Geographical Systems*, 5, 139-160

607 Boots, B. (2006). Local configuration measures for categorical spatial data: binary regular lattices.
608 *Journal of Geographical Systems*, 8, 1-24

609 Boschetti, F. (2008). Mapping the complexity of ecological models. *Ecological Complexity*, 5, 37-47

610 Cliff, A.D., & Ord, J.K. (1981). *Spatial processes: models and applications*. London: Pion

611 Dacey, M.F. (1968). A review on measures of contiguity for two and k-color maps. In B.J.L. Berry, &
612 D.F. Marble (Eds.), *Spatial analysis: a reader in statistical geography* (pp. 479-495). New Jersey:
613 Prentice-Hall Englewood Cliff

614 Dale, M.R. (2000). *Spatial pattern analysis in plant ecology*. Cambridge university press

615 Davison, A.C., & Hinkley, D.V. (1997). *Bootstrap methods and their application*. Cambridge university
616 press

617 DeGraaf, R.M., & Yamasaki, M. (2002). Effects of edge contrast on redback salamander distribution
618 in even-aged northern hardwoods. *Forest Science*, 48, 351-363

619 Desrochers, A., Hanski, I.K., & Selonen, V. (2003). Siberian flying squirrel responses to high-and low-
620 contrast forest edges. *Landscape Ecology*, 18, 543-552

621 Di Gregorio, A., & Jansen, L.J. (2009). *Land cover classification system: LCCS: classification concepts*
622 *and user manual*. Rome: Food and Agriculture Organization of the United Nations

623 Dietrich, C.R., & Newsam, G.N. (1993). A fast and exact method for multidimensional gaussian
624 stochastic simulations. *Water Resources Research*, 29, 2861-2869

625 Dungan, J. (1999). Conditional Simulation: An alternative to estimation for achieving mapping
626 objectives. In A. Stein, F. Meer, & B. Gorte (Eds.), *Spatial Statistics for Remote Sensing* (pp. 135-152):
627 Springer Netherlands

628 Ebrahimi, N., Maasoumi, E., & Soofi, E.S. (1999). Measuring informativeness of data by entropy and
629 variance. *Advances in Econometrics, Income Distribution and Scientific Methodology* (pp. 61-77):
630 Springer

631 Feldman, D.P., & Crutchfield, J.P. (2003). Structural information in two-dimensional patterns:
632 Entropy convergence and excess entropy. *Physical Review E*, 67, 051104

633 Finley, A.O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data
634 with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2, 143-
635 154

636 Gelfand, A.E., Kim, H.-J., Sirmans, C.F., & Banerjee, S. (2003). Spatial Modeling With Spatially Varying
637 Coefficient Processes. *Journal of the American Statistical Association*, 98, 387-396

638 Getis, A., & Ord, J.K. (1992). The analysis of spatial association by use of distance statistics.
639 *Geographical Analysis*, 24, 189-206

640 Getis, A., & Ord, J.K. (1996). Local spatial statistics: an overview. In P. Longley, & M. Batty (Eds.),
641 *Spatial analysis: Modelling in a GIS environment* (pp. 261-277). New York: John Wiley

642 Goodchild, M.F. (1986). *Spatial autocorrelation*. Norwich: Geo Books

643 Griffith, D.A. (2010). The Moran coefficient for non-normal data. *Journal of Statistical Planning and*
644 *Inference*, 140, 2980-2990

645 Guo, D. (2003). Coordinating computational and visual approaches for interactive feature selection
646 and multivariate clustering. *Information Visualization*, 2, 232-246

647 Guo, D.S. (2010). Local entropy map: a nonparametric approach to detecting spatially varying
648 multivariate relationships. *International Journal of Geographical Information Science*, 24, 1367-1389

649 Hamm, N.A.S., Atkinson, P.M., & Milton, E.J. (2012). A per-pixel, non-stationary mixed model for
650 empirical line atmospheric correction in remote sensing. *Remote sensing of environment*, 124, 666-
651 678

652 Hamm, N.A.S., Finley, A.O., Schaap, M., & Stein, A. (2015a). A spatially varying coefficient model for
653 mapping PM10 air quality at the European scale. *Atmospheric Environment*, 102, 393-405

654 Hamm, N.A.S., Soares Magalhães, R.J., & Clements, A.C.A. (2015b). Earth Observation, Spatial Data
655 Quality, and Neglected Tropical Diseases. *PLOS Neglected Tropical Diseases*, 9, e0004164

656 Haskard, K.A., & Lark, R.M. (2009). Modelling non-stationary variance of soil properties by tempering
657 an empirical spectrum. *Geoderma*, 153, 18-28

658 Heikkila, E.J., & Hu, L. (2006). Adjusting spatial-entropy measures for scale and resolution effects.
659 *Environment and Planning B: Planning and Design*, 33, 845-861

660 Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A.,
661 MacMillan, R.A., Mendes de Jesus, J., Tamene, L., & Tondoh, J.E. (2015). Mapping Soil Properties of
662 Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PloS one*, 10,
663 e0125814

664 Iyer, P.K. (1949). The first and second moments of some probability distributions arising from points
665 on a lattice and their application. *Biometrika*, 135-141

666 James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*.
667 Springer

668 Journel, A.G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International*
669 *Association for Mathematical Geology*, 15, 445-468

670 Journel, A.G., & Deutsch, C.V. (1993). ENTROPY AND SPATIAL DISORDER. *Mathematical Geology*, 25,
671 329-355

672 Karlström, A., & Ceccato, V. (2000). A new information theoretical measure of global and local
673 spatial association

674 Lark, R.M. (2009). Kriging a soil variable with a simple nonstationary variance model. *Journal of*
675 *Agricultural, Biological, and Environmental Statistics*, 14, 301-321

676 Lindley, D.V. (1956). On a Measure of the Information Provided by an Experiment. *The Annals of*
677 *Mathematical Statistics*, 27, 986-1005

678 Lloyd, C.D. (2007). *Local models for spatial analysis*. CRC/Taylor & Francis London

679 López-Ruiz, R., Mancini, H.L., & Calbet, X. (1995). A statistical measure of complexity. *Physics Letters*
680 *A*, 209, 321-326

681 Matilla-García, M., & Marín, M.R. (2011). Spatial Symbolic Entropy: A Tool for Detecting the Order of
682 Contiguity. *Geographical Analysis*, 43, 228-239

683 Matilla-García, M., Ruiz, J.R., & Marín, M.R. (2012). Detecting the order of spatial dependence via
684 symbolic analysis. *International Journal of Geographical Information Science*, 26, 1015-1029

685 Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society.*
686 *Series B (Methodological)*, 10, 243-251

687 Morrison, D.G. (1972). Regressions with Discrete Dependent Variables: The Effect on R². *Journal of*
688 *Marketing Research*, 9, 338-340

689 Naimi, B., & Araújo, M.B. (2016). sdm: a reproducible and extensible R platform for species
690 distribution modelling. *Ecography*, 39, 368-375

691 Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., & Toxopeus, A.G. (2014). Where is positional
692 uncertainty a problem for species distribution modelling? *Ecography*, 37, 191-203

693 Naimi, B., Skidmore, A.K., Groen, T.A., & Hamm, N.A.S. (2011). Spatial autocorrelation in predictors
694 reduces the impact of positional uncertainty in occurrence data on species distribution modelling.
695 *Journal of Biogeography*, 38, 1497-1509

696 Ord, J.K., & Getis, A. (1995). LOCAL SPATIAL AUTOCORRELATION STATISTICS - DISTRIBUTIONAL
697 ISSUES AND AN APPLICATION. *Geographical Analysis*, 27, 286-306

698 Ord, J.K., & Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global
699 autocorrelation. *Journal of Regional Science*, 41, 411-432

700 Paciorek, C.J., & Schervish, M.J. (2006). Spatial modelling using a new class of nonstationary
701 covariance functions. *Environmetrics*, 17, 483-506

702 Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*,
703 30, 683-691

704 Pham, T.D. (2010). GeoEntropy: A measure of complexity and similarity. *Pattern Recognition*, 43,
705 887-896

706 Quester, P., & Dion, E. (1997). Scaling Numerical Variables and Information Loss: An Appraisal of
707 Morrison's Work. *MARKETING BULLETIN-DEPARTMENT OF MARKETING MASSEY UNIVERSITY*, 8, 59-
708 65

709 R Development Core Team (2017). *R: A language and environment for statistical computing*. Vienna,
710 Austria

711 Ricotta, C., & Anand, M. (2006). Spatial complexity of ecological communities: Bridging the gap
712 between probabilistic and non-probabilistic uncertainty measures. *Ecological Modelling*, 197, 59-66

713 Robinson, M.D., Feldman, D.P., & McKay, S.R. (2011). Local entropy and structure in a two-
714 dimensional frustrated system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21, 037114-
715 037114-037111

716 Romme, W.H. (1982). Fire and Landscape Diversity in Subalpine Forests of Yellowstone National
717 Park. *Ecological Monographs*, 52, 199-221

718 Ruiz, M., López, F., & Páez, A. (2010). Testing for spatial association of qualitative data using
719 symbolic dynamics. *Journal of Geographical Systems*, 12, 281-309

720 Schlather, M. (2009). RandomFields: Simulation and analysis of random fields. In (p. R package
721 version 1.3.41)

722 Schwering, A. (2008). Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A
723 Survey. *Transactions in GIS*, 12, 5-29

724 Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Information*. Urbana, IL:
725 University of Illinois Press

726 Sokal, R.R., & Oden, N.L. (1978). Spatial autocorrelation in biology: 1. Methodology. *Biological*
727 *Journal of the Linnean Society*, 10, 199-228

728 Uuemaa, E., Roosare, J., Kanal, A., & Mander, Ü. (2008). Spatial correlograms of soil cover as an
729 indicator of landscape heterogeneity. *Ecological Indicators*, 8, 783-794

730 Yeung, R.W. (2008). *Information theory and network coding*. Springer

731

732

733 **Tables**

734 **Table 1** the level of dissimilarity between pairs of categories in an exemplified land use map with
735 four (sub)categories: 'mixed forest' (α_1), 'Coniferous forest' (α_2), 'olive groves' (α_3), 'vineyards' (α_4);
736 the first two belong to the main category of 'Forests' and the last two are related to 'Agricultural
737 areas'

738

Categories	α_1	α_2	α_3	α_4
α_1	0	1	2	2
α_2		0	2	2
α_3			0	1
α_4				0

739

740

741

742 **Table 2** Size and power of the ELSA test for continuous, and categorical variables with different
 743 degrees of spatial autocorrelation, ranging from no autocorrelation ($\gamma = 0$) to high autocorrelation (
 744 $\gamma = 1$); K specifies the number of categories in the categorical variables, and Ne specifies the
 745 neighbourhood size for measuring the ELSA statistic

746

747		Ne	$\gamma = 0$	$\gamma = 0.25$	$\gamma = 0.5$	$\gamma = 0.75$	$\gamma = 1$
748	Continuous	1.5	0.045	0.487	0.882	0.894	0.990
749		3	0.048	0.408	0.807	0.916	0.983
750		5	0.039	0.398	0.792	0.891	0.977
		10	0.031	0.294	0.734	0.825	0.897
		15	0.012	0.207	0.572	0.650	0.789
751	K=2	1.5	0.028	0.406	0.787	0.801	0.941
752		3	0.017	0.423	0.697	0.787	0.866
753		5	0.012	0.412	0.701	0.756	0.840
		10	0.010	0.391	0.639	0.707	0.718
		15	0.002	0.322	0.596	0.650	0.659
754	K=3	1.5	0.046	0.565	0.705	0.846	0.944
755		3	0.005	0.511	0.717	0.827	0.903
756		5	0.003	0.495	0.713	0.812	0.892
		10	0.002	0.425	0.601	0.729	0.856
757		15	0.000	0.330	0.593	0.698	0.765
758	K=4	1.5	0.013	0.515	0.741	0.885	0.927
		3	0.011	0.498	0.705	0.822	0.955
		5	0.009	0.413	0.702	0.804	0.926
		10	0.000	0.354	0.619	0.781	0.876
		15	0.000	0.283	0.532	0.773	0.792

760 **Table A1** the CORINE land cover class definitions

CORINE Code	Level 1	Level 2	Level 3
111	Artificial surfaces	Urban fabric	Continuous urban fabric
112	Artificial surfaces	Urban fabric	Discontinuous urban fabric
121	Artificial surfaces	Industrial, commercial and transport	Industrial or commercial
122	Artificial surfaces	Industrial, commercial and transport	Road and rail networks
124	Artificial surfaces	Industrial, commercial and transport	Airports
131	Artificial surfaces	Mine, dump and construction sites	Mineral extraction sites
132	Artificial surfaces	Mine, dump and construction sites	Dump sites
133	Artificial surfaces	Mine, dump and construction sites	Construction sites
141	Artificial surfaces	Artificial, non-agricultural vegetated	Green urban areas
142	Artificial surfaces	Artificial, non-agricultural vegetated	Sport and leisure facilities
211	Agricultural areas	Arable land	Non-irrigated arable land
212	Agricultural areas	Arable land	Permanently irrigated land
221	Agricultural areas	Permanent crops	Vineyards
222	Agricultural areas	Permanent crops	Fruit trees and berry plantations
223	Agricultural areas	Permanent crops	Olive groves
231	Agricultural areas	Pastures	Pastures
241	Agricultural areas	Heterogeneous agricultural areas	Annual crops/permanent crops
242	Agricultural areas	Heterogeneous agricultural areas	Complex cultivation patterns
243	Agricultural areas	Heterogeneous agricultural areas	Land occupied by agriculture
244	Agricultural areas	Heterogeneous agricultural areas	Agro-forestry areas
311	Forest and semi natural	Forests	Broad-leaved forest
312	Forest and semi natural	Forests	Coniferous forest
313	Forest and semi natural	Forests	Mixed forest
321	Forest and semi natural	Scrub and/or herbaceous vegetation	Natural grasslands
323	Forest and semi natural	Scrub and/or herbaceous vegetation	Sclerophyllous vegetation
324	Forest and semi natural	Scrub and/or herbaceous vegetation	Transitional woodland-shrub
331	Forest and semi natural	Open spaces with little/no vegetation	Beaches, dunes, sands
332	Forest and semi natural	Open spaces with little/no vegetation	Bare rocks
333	Forest and semi natural	Open spaces with little/no vegetation	Sparsely vegetated areas
334	Forest and semi natural	Open spaces with little/no vegetation	Burnt areas
411	Wetlands	Inland wetlands	Inland marshes
511	Water bodies	Inland waters	Water courses
512	Water bodies	Inland waters	Water bodies

761

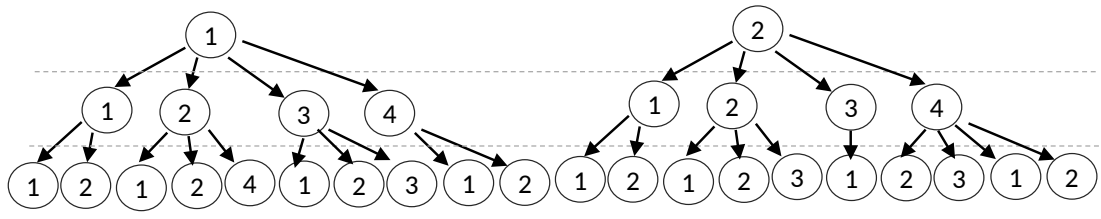
762

763

764

765

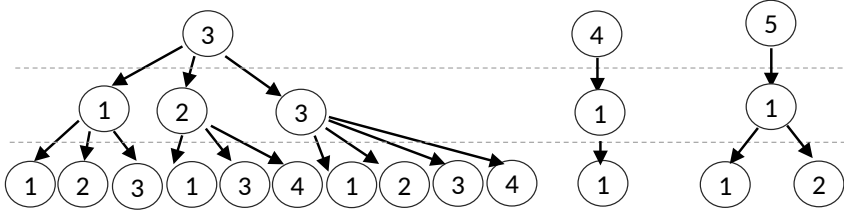
766



Level 3
Level 2
Level 1

767

768



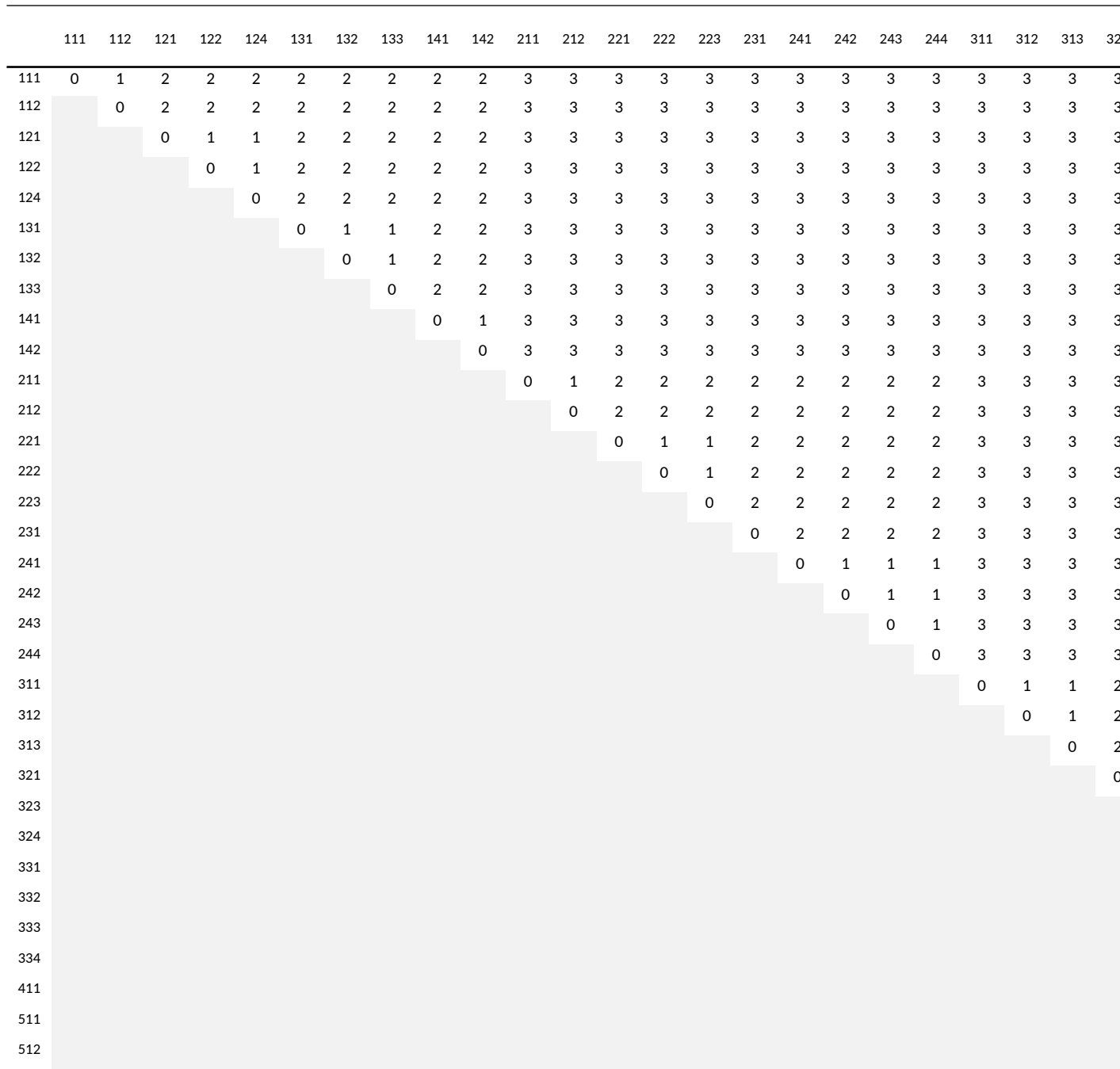
Level 3
Level 2
Level 1

769 **Fig. A1.** Hierarchical scheme of different classes in the CORINE land cover map (Fig. 7 in the main
770 text)

771

772

773



775 **Fig. A2.** The level of dissimilarity between pairs of classes in the CORINE land cover map

1 **Figure Captions**

2 **Figure 1.** The behaviour of ELSA statistic (E) and its two terms at site i (the centre of the window):

3 E_{ai} (first term in ELSA refers to attribute distance or dissimilarity between the site i and the its
4 neighbour locations), E_{ci} (second term in ELSA, refers to composition of categories within the local
5 area); (a) is a map with one category, (b, and c) are binary maps, and (d, and e) are maps with three
6 categories.

7 **Figure 2.** A hierarchical way of presenting the classes in an exemplified categorical map with 3 levels
8 of categories; the numbers in the boxes indicate d (dissimilarity) of the relevant pairs of classes

9 **Figure 3.** A flow diagram showing the procedure of finding the optimum number of categories for
10 categorizing continuous spatial data; (a) an iterative categorization procedure taking different
11 number of categories, starts from 2 categories and continues by considering one more category at
12 each iteration until a convergence threshold (e.g., 0.005) is reached; (b) calculating the ρ correlation
13 coefficient between the continuous values and each categorical variable; the convergence is defined
14 as the difference between the ρ coefficients of the current and previous iterations; (c) taking the
15 standard error of the ρ coefficients is calculated; (d) the optimum number of categories would be
16 the lowest number for which the ρ coefficient is within one standard error of the highest ρ
17 coefficient; (e) the graph representing the ρ coefficients over different iterations and the optimum
18 number of categories (dashed red line); the grey colour represents the standard error of the ρ
19 coefficients

20 **Figure 4.** Four variogram models with different nugget effects ranging between 0 and 10, used to
21 simulate the spatial surfaces with different degrees of spatial autocorrelation

22 **Figure 5.** One realization (out of 999) of the simulated continuous and categorical maps with various
23 levels of spatial autocorrelation to calculate the power and size of the ELSA statistic; γ is the ratio

24 between the partial sill to the sill in the variogram model used to simulate the maps (Figure 4) that
25 controls the level of spatial autocorrelation ranging from no autocorrelation ($\gamma=0$), and high
26 autocorrelation ($\gamma=1$); K specifies the number of classes in the categorical maps.

27 **Figure 6.** ELSA for a land cover map in the south of Spain; (a) the land cover map including
28 six classes with the same level of dissimilarity between different categories; three cells
29 within their five km neighbourhoods are specified as A, B, C; (b) the ELSA statistic for the
30 land cover map; the values of ELSA for the three cells are represented on top

31

32 **Figure 7.** Synthetic land cover map including four classes which distributions are controlled into four
33 equal zones (a); levels of dissimilarity between pairs of classes in a hierarchical view (b)

34 **Figure 8.** The ELSA map (a) and the Mean ELSA statistic in four zones of the region (b); the
35 region is divided into four zones including Z1 (upper-left), Z2 (upper-right), Z3 (lower-right)
36 and Z4 (lower-left); the boxplot (c) represents the distribution of ELSA values at grid cells
37 over different zones

38 **Figure 9.** CORINE Land cover map from the central Spain; a three-digit code is used to define each
39 land cover class (a); three randomly selected points around which the circle specifies a 5 km of their
40 neighbors (b)

41 **Figure 10.** ELSA map calculated based on the CORINE land cover map in Fig. 7 (a); the ELSA value at
42 the three specified locations (b)

43 **Figure 11.** Comparing local indicators of spatial association for a continuous raster map with positive
44 global spatial autocorrelation; the panel contains 12 map/graph/table including the simulated raster
45 map, a Moran scatterplot showing the degree of global spatial autocorrelation, four maps of local
46 Moran's I , local Geary's c , local G_i^* , and ELSA, 2 maps representing the p-values that indicate the

47 level of significance for local spatial association based on local Moran's I, and ELSA, respectively, and
48 3 graphs representing the relationship between ELSA on y axis and local Geary's c, local Moran's I,
49 and G_i^* statistics on x axis, and finally a table representing the pairwise spearman correlation
50 coefficients between different local indicators of spatial association; the symbol '***' indicates that
51 the correlation coefficient is significant at the level of 99.9%

52 **Figure 12.** Comparing local indicators of spatial association for a continuous raster maps with no
53 global spatial autocorrelation; the different components of the figure are declared in Fig. 11.
54

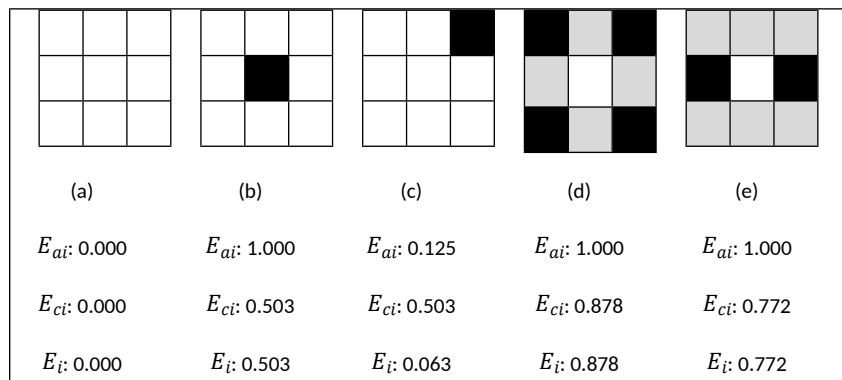
55 **Figure 13.** Comparing variogram and entrogram for three simulated binary categorical maps with
56 different levels of spatial autocorrelation ($\phi = 0, 3, 8$ from left to right); the second and third rows
57 display the corresponding theoretical and empirical variograms, respectively, and the last row
58 displays the entrograms
59

60 **Figure 14.** Two categorical maps including four classes (first row) and their corresponding
61 entrograms (second row)

62 **Figure 15.** Comparing variogram and entrogram; the first row displays the 5 simulated continuous
63 fields with different levels of spatial autocorrelation ($\phi = 0, 10, 30, 10, \text{ and } 10$ from left to right), and
64 varying parameterisation in the variogram model, nugget (C_0) and partial sill (C) (as specified in the
65 figure); the second row displays the corresponding variograms and the third row displays the
66 entrograms
67

68 **Figures**

69



70

71

Figure 1

72

73

74

75

76

77

78

79

80

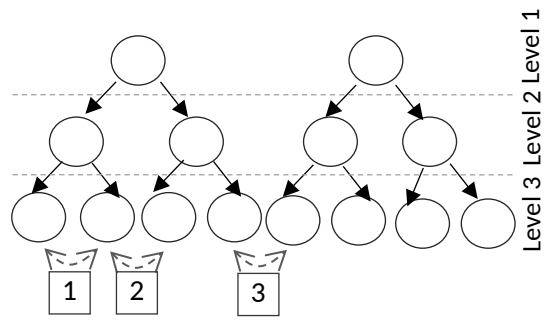


Figure 2.

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

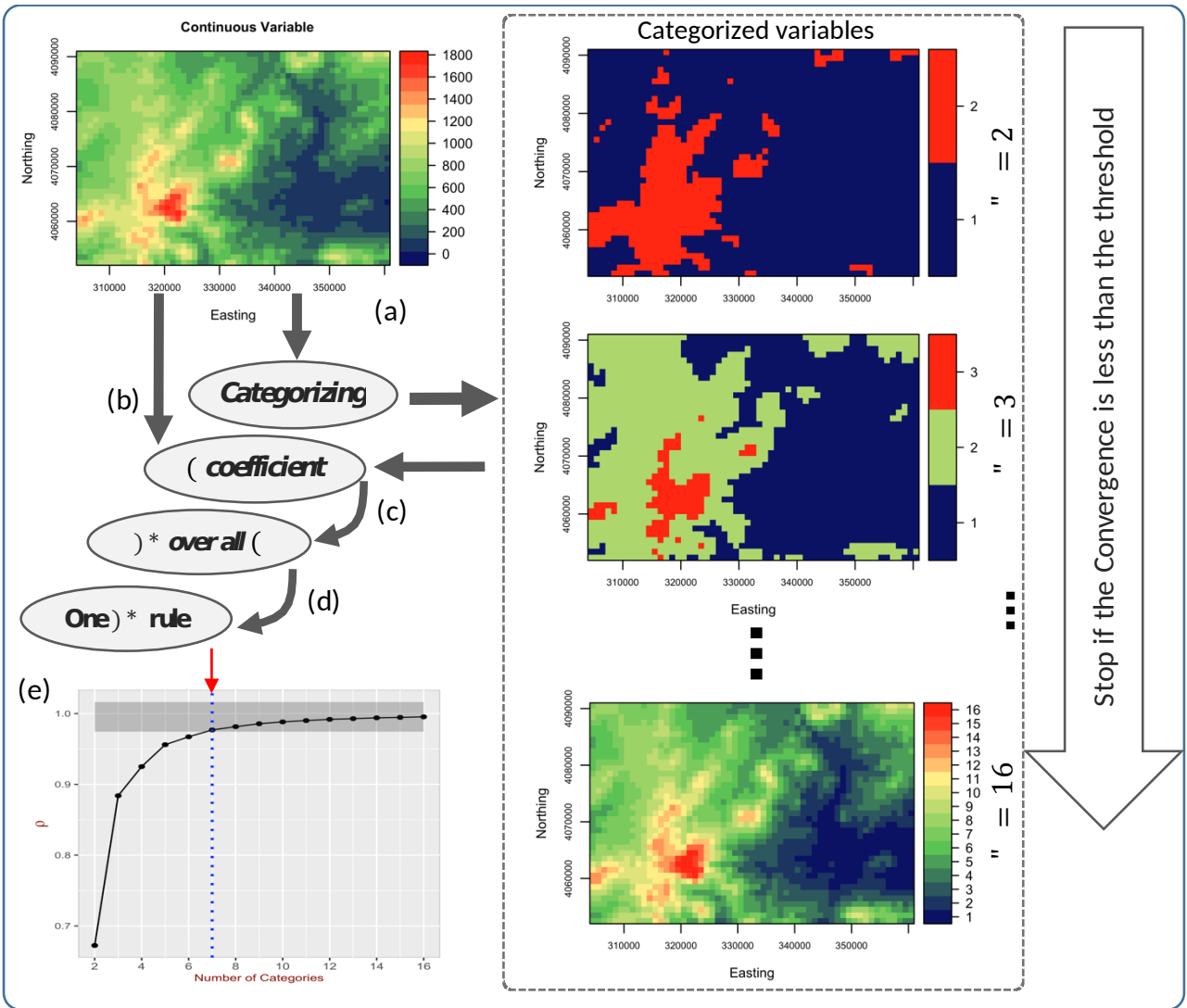


Figure 3.

105

106

107

108

109

110

111

112

113

114

115

116

117

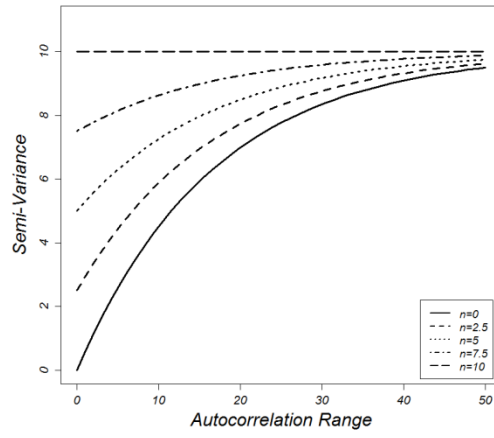
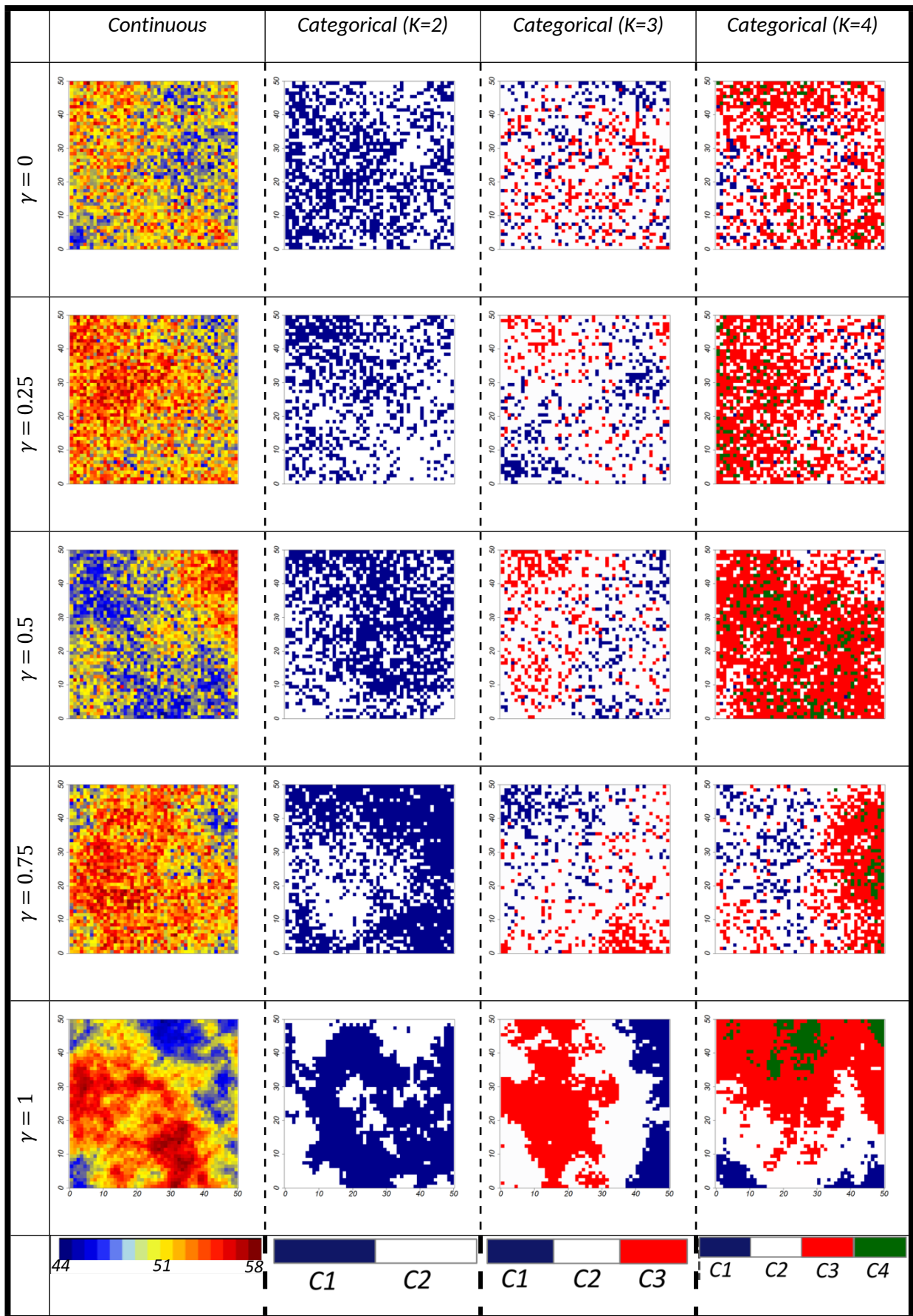


Figure 4.



118 Figure 5.

119

120

121

122

123

124

125

126

127

128

129

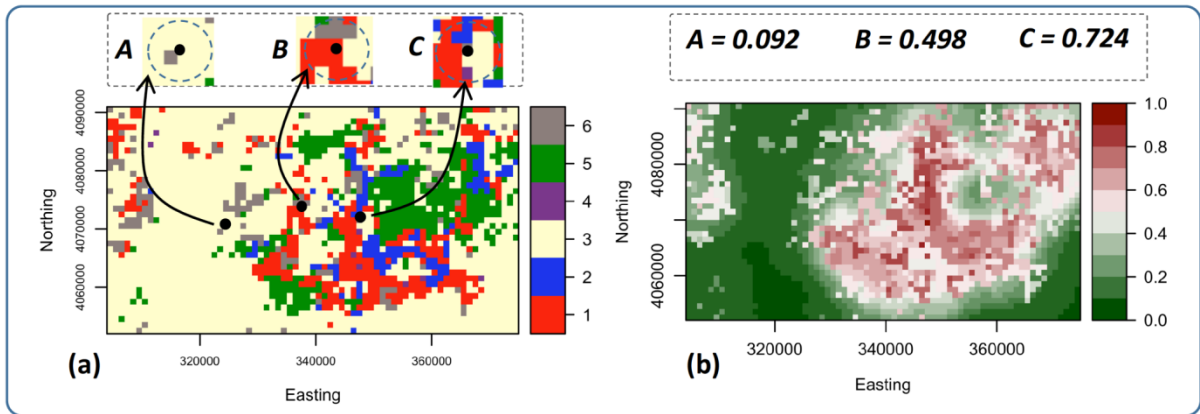


Figure 6

130
131
132
133
134
135
136
137
138
139

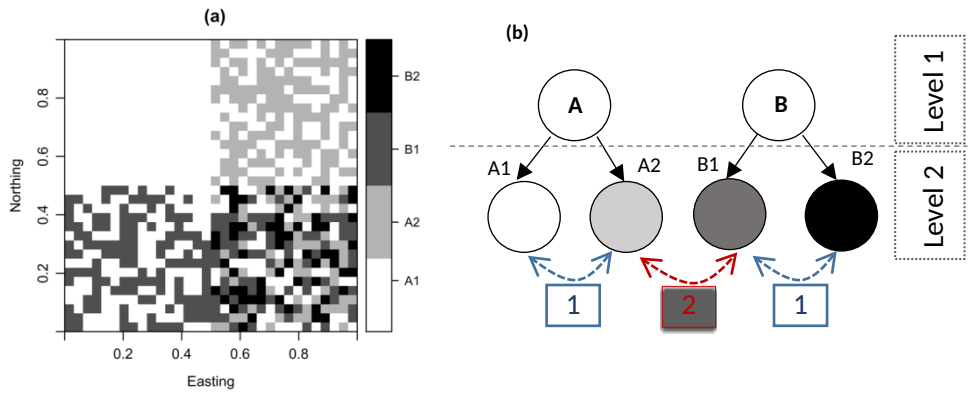


Figure 7.

140
141
142
143
144
145
146
147
148
149
150
151
152

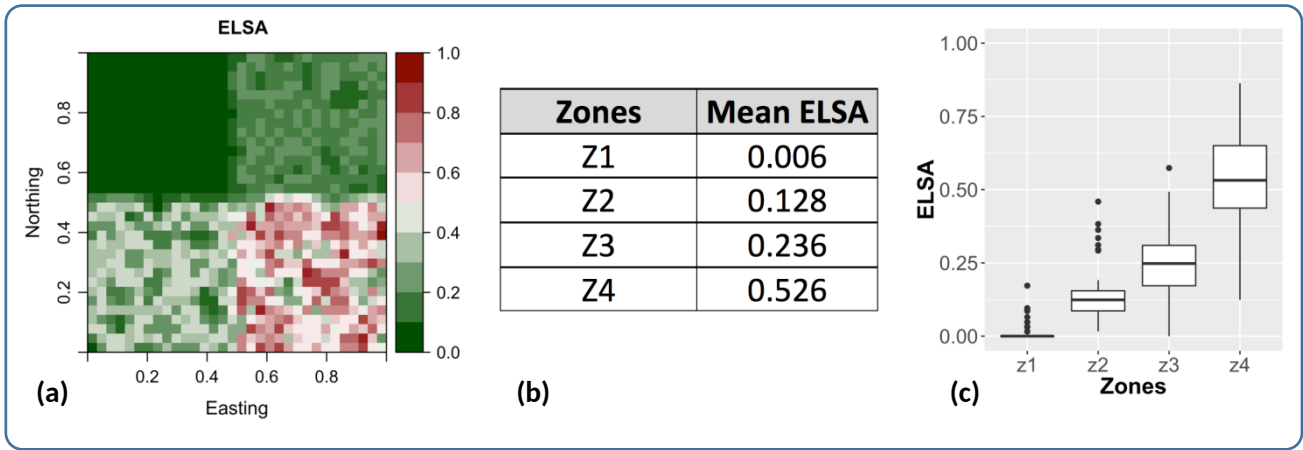


Figure 8.

153
154
155
156
157
158
159
160
161
162
163
164
165
166

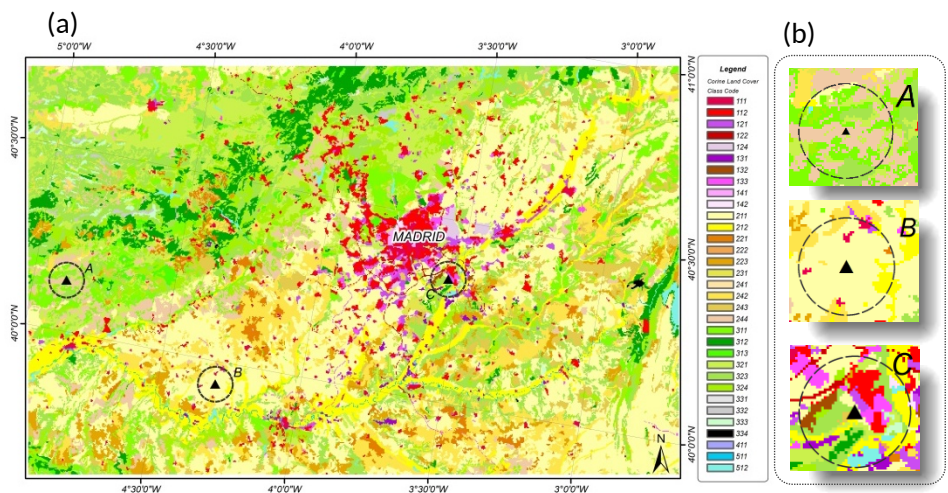


Figure 9.

167

168

169

170

171

172

173

174

175

176

177

178

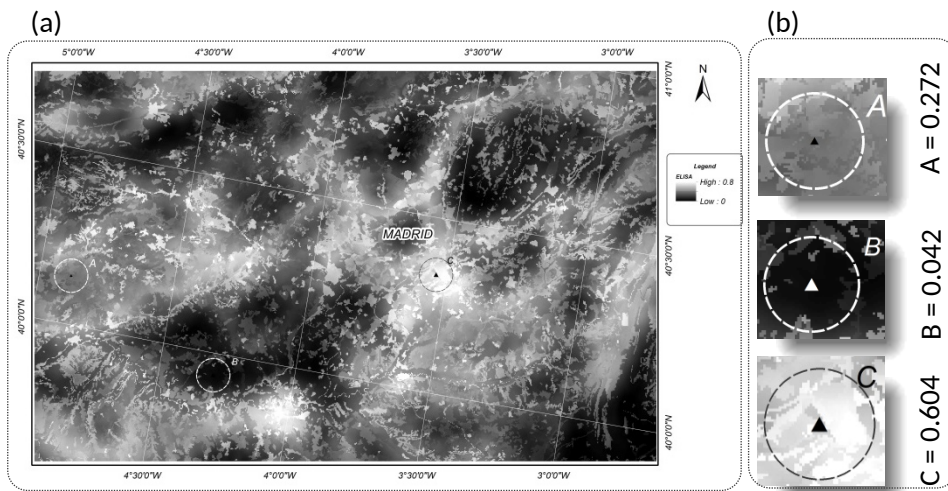


Figure 10.

179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201

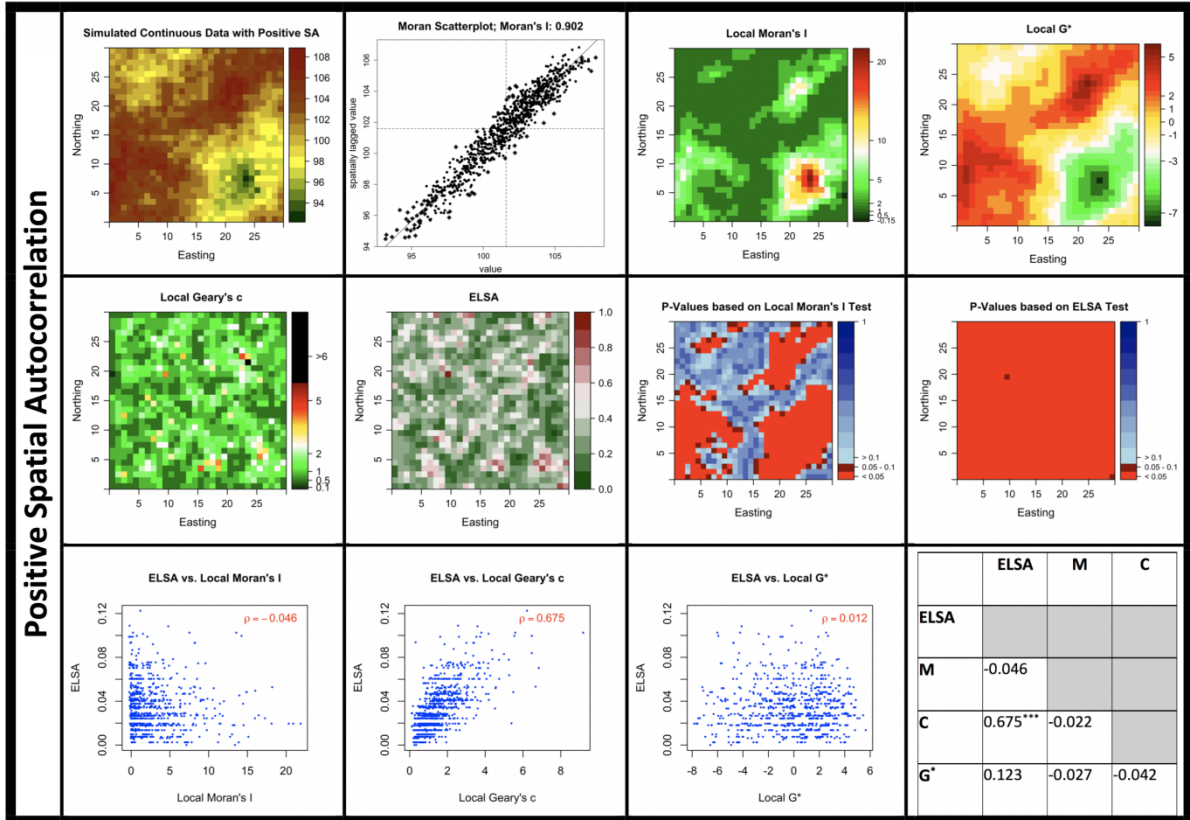


Figure 11

202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227

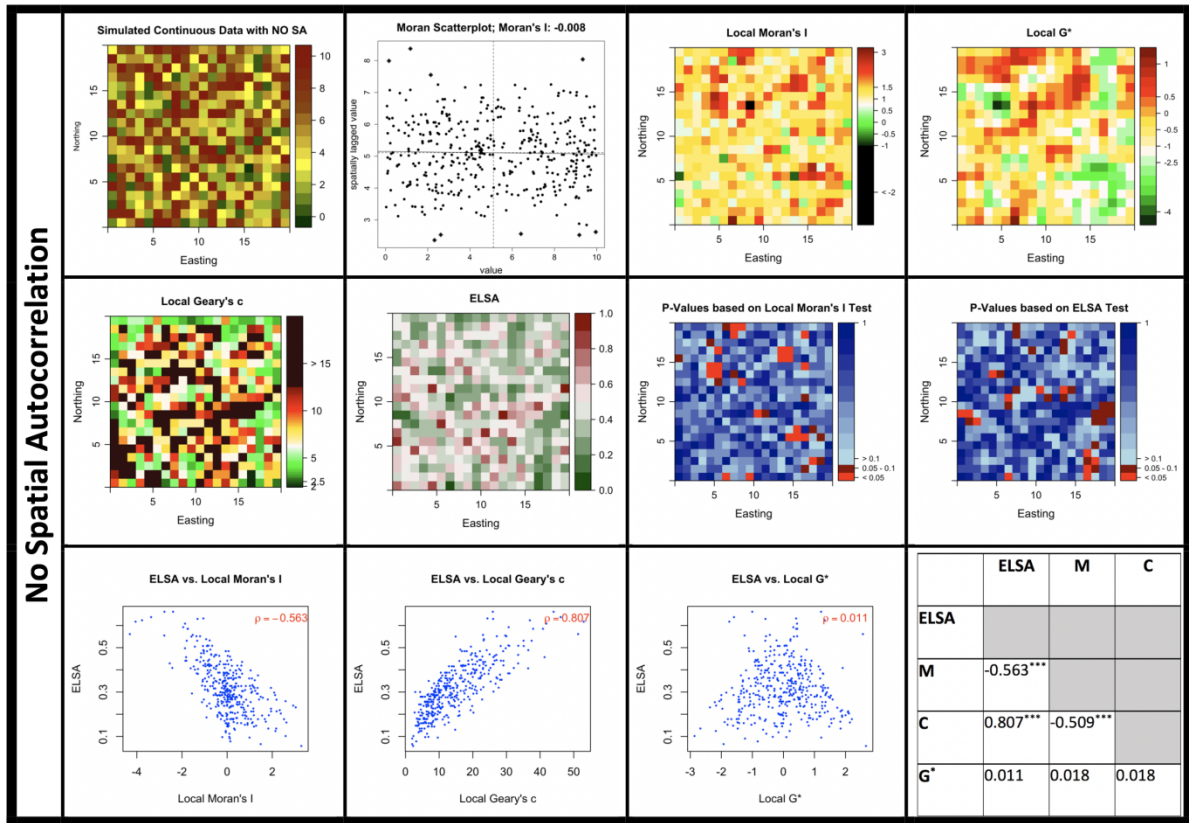


Figure 12.

228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245

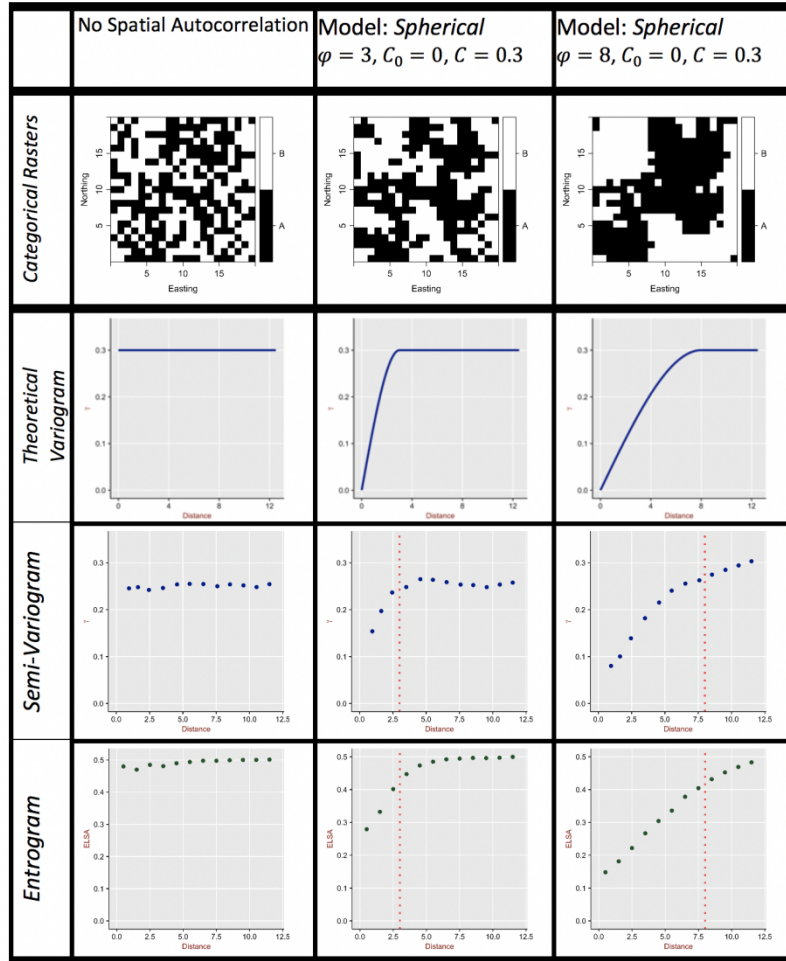


Figure 13.

246
247
248
249
250
251
252
253
254
255
256
257
258
259

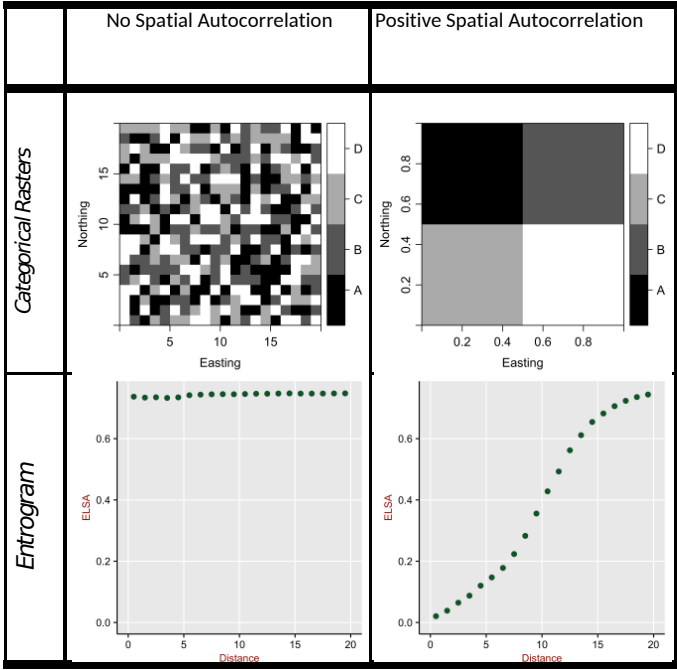


Figure 14.

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

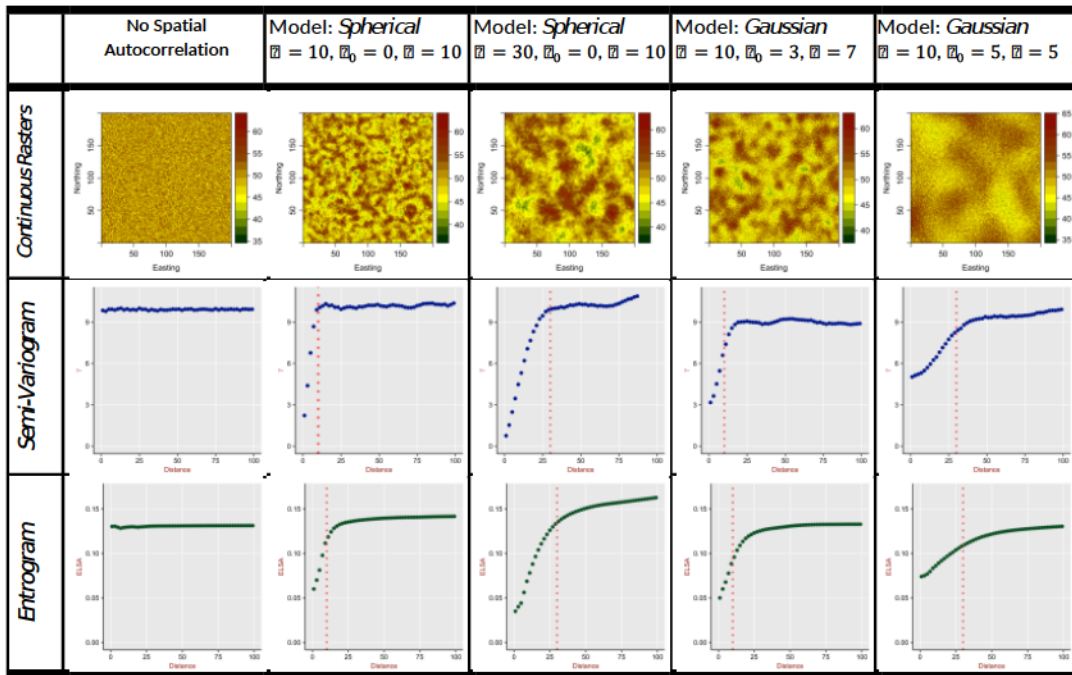


Figure 15.