Spatial association from the perspective of mutual information

Wen-Bin Zhang[a,b,c], Yong Ge[a,b*], Hexiang Bai[d], Yan Jin[e,f] and Alfred Stein[g], Peter M Atkinson[c*]

[a]*State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China*

[b]*College of Resources and Environment, University of Academy of Sciences, Beijing, China*

[c]*Lancaster Environment Center, Faculty of Science and Technology, Lancaster University, Lancaster, UK*

[d]*School of Computer and Information Technology, Shanxi University, Taiyuan, China*

[e]*School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China*

[f]*Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province, Department, Nanjing, China*

[g]*Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, Netherlands*

*Corresponding author: Yong Ge, gey@lreis.ac.cn； Peter M Atkinson, pma@lancaster.ac.uk

Measures of spatial association are important to reveal the spatial structures and patterns in geographical phenomena. They have utility for spatial interpolation, stochastic simulation and causal inference, amongst others. Such measures are abundantly available for continuous spatial variables, while for categorical spatial variables they are less well developed. In this research, we developed a measure of spatial association for categorical spatial variables coined the "entropogram", quantifying its spatial association using mutual information. Mutual information concerns information shared by pairs of random variables at different locations as revealed by their observed joint frequency distribution and marginal frequency distributions. The developed new measure is modeled as a function of lag in analogy to the variogram. While existing measures focus mainly on inter-state relationships, the entropogram models the spatial correlation in categorical spatial variables holistically. In this way, the entropogram brings multiple advantages, for example, simplifying the representation of spatial structure for categorical variables and facilitating communication. Besides, the entropogram also reflects variation in the spatial correlation between different states. We first explored the properties of the entropogram in a simulation study. Then, we applied the entropogram to analyze the spatial association of land cover types in Qinxian, Shanxi, China. We conclude that the entropogram provides a suitable addition to existing measures of spatial association for applications in a wide range of disciplines where the categorical spatial variable is of interest.

Keywords: categorical data; entropogram; mutual information; multi-categorical random function; spatial association

1. Introduction

Spatial association is an essential property of Earth science data (Fotheringham 2009; Goodchild 2011). It describes the variation in a property or between elements as a function of the distance and direction vector between observations at different locations (Cliff and Ord 1981). Spatial association is determined by the underlying spatial and dynamic processes operating on geographic landscapes, whether they arise from natural or human activities. For example, land-cover change processes may affect the spatial pattern of the landscape, which itself may affect the space-time pattern of the local micro-climate (Pielke Sr 2005). Often, spatial association can be used to infer the parameters of models of the corresponding underlying dynamic processes that led to the observed patterns, and support subsequent decision-making (Wang, Zhang, and Fu 2016; Benedetti 2020). It is, therefore, important to measure and characterize the spatial association in geographical properties and elements over the Earth surface.

At the broadest level of classification, Earth science data as a realization of random functions (RFs) can be either continuous or categorical (Ge et al 2019). A RF is a stochastic process that can generate the same realisations as a dynamic process (i.e., a RF is a surrogate for our incomplete knowledge of the dynamic process). The main difference between continuous and categorical data is that categorical data consist of states (e.g., land cover types) while continuous data take values on an interval or ratio scale (e.g., temperature). This distinction has led to the emergence of different methods to characterize various types of spatial association for RFs, including spatial autocorrelation and spatial heterogeneity (Anselin 1995; Wang et al. 2010). Since the 1950s, various statistical measures and functions have been proposed to describe the spatial association in continuous data. Widely used statistical measures for a spatial continuous field are Moran's I (Moran 1950), Geary's c (Geary 1954), the covariance function and the variogram (Matheron 1963). Moran's I and Geary's c were developed to test for spatial correlation in a continuous variable measured at discrete units. The covariance function is rooted in time-series modelling, and was adapted to model spatial dependence, while the variogram, as its generalization, was introduced specifically for handling spatial data (Matheron, 1963; Goovaerts 1997; Garrigues et al. 2006). Both the covariance function and the variogram describe how spatial variation in a continuous variable varies as a function of separation distance and direction. These functions were developed for continuous variables. They cannot be applied directly to categorical data as the states in categorical data are qualitatively different, not numerically different.

The indicator variogram was proposed as an extension of the variogram to model categorical data with states (Journel 1986). The multiple states are reduced to a set of binary spatial variables by comparing each state against all others each time, and the resulting binary data are *de facto* discrete RFs taking only two possible values (0, 1), for example, referring to the presence (1) or absence (0) of the target state. Following this transformation from states to binary values, a more general solution is to capture the corresponding frequency information of the variable states with a probability mass distribution of states. For example, the join count statistic is a widely used frequency-derived index to characterize the global spatial autocorrelation of categorical variables (Cliff and Ord 1970). In place of the variance of binary data, it represents the degree of dispersion by relating the number of connections (corresponding to the occurrence of value pairs at neighbouring locations) to the theoretical number of connections if the points were distributed randomly. More recently, it was popularized, and the number of connections was extended to the transition probability of the states at neighbouring locations (Bai et al. 2016). To address spatial heterogeneity, the conditional version of a local join count statistic was proposed (Anselin and Li 2019), while the transiogram (Li 2006) was developed to model the transition probability between different variable states as a function of spatial lag. These spatial association measures focused mainly on state-level spatial association, especially inter-state relationships, and did not result in a comprehensive representation of the full variable state space.

Entropy characterizes the spatial association of a categorical spatial variable where the transformation from states to values is no longer needed. Measures of spatial association based on entropy include symbolic entropy (Ruiz, López, and Páez 2010), spatial entropy (Leibovici et al. 2011), spatial mutual information (Altieri, Cocchi, and Roli 2018) and the entropy-based local indicator of spatial association (Naimi et al. 2019). These global and local entropy-based indices of spatial association for categorical data fail to capture any heterogeneity in the underlying stochastic process from which the realization (spatial data) is supposed to have been drawn (Atkinson and Tate 2000). Most existing entropy-derived measures assume implicitly that all spatial random variables (RVs) share the same probability mass distribution at each location. Spatial data are then considered as mutually independent samples from that distribution. This assumption of independently and identically distributed (i.i.d.) samples taken from a spatially distributed phenomenon, however, is geographically unrealistic. In this circumstance, spatial association as a function of the distance (and direction) between locations cannot be generalized for categorical data.

In this research, we introduce the concept of mutual *information* into the variogram. We develop and apply a new function to characterize the spatial association of a categorical spatial variable based on the mutual information between pairs of points, under the assumption of second-order stationarity. The developed new function is termed the entropogram, which can model the spatial association in multi-category (i.e., multi-state) spatial data directly. Specifically, it is conceived as a function of lag, in analogy to the variogram, where the variance at each lag is replaced by the corresponding mutual information about the RV at two locations. Mutual information quantifies the total amount of information shared by the RV at two locations. It reveals the spatial dependence between any two spatial locations in terms of the full variable state space instead of the individual variable states only. In this way, the entropogram can help to better understand the geographical processes underlying categorical properties from an information perspective.

In the remainder of this paper, we first define the entropogram and propose its estimation from sample data. Then corresponding confidence intervals are provided through an uncertainty analysis. Next, we present both numerical and real-world experiments that examine the performance of the proposed entropogram together with a discussion of the most salient issues. Finally, we provide some concluding remarks.

2. Capturing spatial association with mutual information

2.1 Conceptual framework

In this section, we give a brief introduction to the development of the variogram and entropy-based measures of spatial association for a single qualitative spatial variable, to demonstrate clearly our contribution.

2.1.1 Variogram

Geostatistics is based on regionalized variable (ReV) theory (Matheron 1963). ReV theory defines, first, a Random Function (RF) model, being the spatial equivalent of a Random Variable (RV) where each location has its own RV. The RF is parameterized by the variogram, which represents 'semivariance' as a function of lag (the distance and direction of separation). The semivariance is the spatial equivalent of (specifically half of) the variance of a RV for a pair of points. Application of the variogram is, therefore, accompanied by the decision to adopt a RF that is intrinsically stationary. This requires that the RF covering the study domain has a constant mean, and that the semivariance of the paired differences between RVs depends only on the lag between their two locations. In this way, the variogram characterizes spatial dependence and, more specifically, it specifies how the semivariance varies as a function of the lag between pairs of locations. Mathematically, given a RF Z the variogram γ is defined for spatial lag $\mathbf{h}$ (i.e., the distance and direction between any two locations in the study domain) as:

$$\gamma(\mathbf{h}) = \gamma(Z(\mathbf{s}), Z(\mathbf{s}-\mathbf{h})) = \frac{1}{2}\mathrm{E}\left[\left(Z(\mathbf{s}) - Z(\mathbf{s}-\mathbf{h})\right)^2\right], \quad (1)$$

where $\mathbf{s}$ is a location vector.

2.1.2 Entropy-derived measures

Most entropy-based measures of spatial association are derived directly from the classic Shannon entropy (Shannon and Weaver 1949). Entropy characterizes the different states of categorical variables simultaneously. Consider a categorical RV X with m finite states x_1, x_2,…, x_m, each with an occurrence probability $p(x_1)$, $p(x_2)$,…, $p(x_m)$, respectively. The Shannon entropy $H(X)$ of a categorical RV X represents the amount of information associated with each observation to identify its true state (Shannon and Weaver 1949). It equals:

$$H(X) = -\sum_{i=1}^{m} p(x_i) \ln\left(p(x_i)\right). \quad (2)$$

where x_i is the i^{th} state of X. $H(X)$ represents the expectation of the amount of information that can be obtained from each observation. For state x_i, this equals $-\ln\left(p(x_i)\right)$, indicating that states with lower occurrence probability can provide more information once observed. The Shannon entropy requires X to behave equivalently across space (i.e., the spatial data are considered as mutually independent samples drawn from single RVs; see Figure 1). In consequence, spatial associations between locations cannot be captured by the model due to the independence assumption. *Spatial* entropy was proposed to characterise the co-occurrence of states at position pairs separated by a distance smaller than a fixed threshold h, instead of the incidence of available states over space (Leibovici et al. 2014). Spatial entropy is also appliable to multivariate joint distributions. As we focus on a single geographical variable across space, co-occurrences are defined here as the simultaneous realization of two RVs at pairs of locations for illustration. Specifically, all state pairs observed at two locations less than distance h apart are assumed to be drawn from a bivariate distribution $< X_1, X_2 >$, and the entropy of $< X_1, X_2 >$ is defined as:

$$H(X_1, X_2) = -\sum_i \sum_j p_h(x_{1,i}, x_{2,j}) \ln\left(p_h(x_{1,i}, x_{2,j})\right), \tag{3}$$

where $x_{1,i}$ and $x_{2,j}$ are the i^{th} and j^{th} states for X_1 and X_2, respectively, and $p_h(x_{1,i}, x_{2,j})$ is their joint probability mass for co-occurrence closer than distance h. Then, the spatial entropy is built as a function of threshold h; that is, the set of state similarities at neighboring location pairs, where neighbors are defined by being closer than the threshold distance h. Note that X_1 and X_2 share the same set of states regardless of h, while samples of the bivariate $< X_1, X_2 >$ are nested, expanding with the threshold distance h. This means that the bivariate $< X_1, X_2 >$ are theoretically distinct at each threshold h, as a result of spatial heterogeneity. In summary, almost all spatial association measures are univariate or bivariate in their approach to describing spatial association, while locations *per se* are not accounted for. This is similar to ReV theory, where the variogram is a two-point statistic (Mariethoz and Caers 2014).

2.2 Mutual information described spatial association

We consider categorical spatial data as a realization of a categorical random field X. Mutual information can be naturally employed to describe the spatial association between its two constituent RVs $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ at a pair of locations $\mathbf{s}_1$ and $\mathbf{s}_2$ from which the realized state at that pair of locations is supposed to have been drawn. The mutual information described spatial association (MSA) $MSA(X(\mathbf{s}_1), X(\mathbf{s}_2))$ between $X(\mathbf{s}_1)$

and $X(\mathbf{s}_2)$ is defined by the information difference between the joint probability distribution of $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ and the sum of their marginal distributions. That is,

$$MSA(X(\mathbf{s}_1), X(\mathbf{s}_2)) = H(X(\mathbf{s}_1)) + H(X(\mathbf{s}_2)) - H(X(\mathbf{s}_1), X(\mathbf{s}_2)), \qquad (4)$$

where $H(X(\mathbf{s}_1))$ and $H(X(\mathbf{s}_2))$ are the Shannon entropy of categorical RVs $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ (see equation (2)), and $H(X(\mathbf{s}_1), X(\mathbf{s}_2))$ is the Shannon entropy of categorical RV $\langle X(\mathbf{s}_1), X(\mathbf{s}_2)\rangle$ (see equation (3)), respectively. Given a categorical spatial dataset, the observed states of interest at distinct locations are assumed to be drawn from a RF X. The second-order property of such a RF (i.e., the covariance matrix) can then be described by our proposed MSA across all location pairs.

3. Entropogram

3.1 Assumption of second-order stationarity

For a pair of locations $(\mathbf{s}_1, \mathbf{s}_2)$, reliable estimation of the joint probability function of the corresponding RVs $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$, as well as their marginal distributions, requires a number of sample observations. However, there are generally insufficient data (generally only one sample for each location) to estimate the probability distribution at each location. Therefore, analogous to the assumption of intrinsic stationarity in geostatistics, we propose to define the MSA by assuming that point pairs separated by the same spatial lag also share equal spatial association. Under this assumption, the MSA is defined as the *entropogram* τ, a function of lag $\mathbf{h}$:

$$\begin{aligned}\tau(\mathbf{h}) &= MSA(X(\mathbf{s}), X(\mathbf{s}-\mathbf{h})) \\ &= \sum_i \sum_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) \ln\left(\frac{p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)}{p(x_i(\mathbf{s}))\, p\left(x_j(\mathbf{s}-\mathbf{h})\right)}\right)\end{aligned} \quad , \qquad (5)$$

where $x_i(\mathbf{s})$ and $x_j(\mathbf{s}-\mathbf{h})$ are the i^{th} and j^{th} states for $X(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{h})$, respectively, and $p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)$ is their joint probability mass. The derivation can be found in Appendix A. Equivalent to the assumption of the shared constant mean, we assume that $X(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{h})$ share the same probability mass function $p(x_i)$, for the RVs at each location, where x_i is the i^{th} state of the study categorical variable. In this way, $p(x_i(\mathbf{s}))$ can be estimated by

$$\hat{p}(x_i(\mathbf{s})) = N_i/N, \tag{6}$$

where N_i is the number of observations belonging to the i^{th} state and N is the total number of observations. $x_i(\mathbf{s})$ is conceived as the i^{th} state at a randomly chosen location $\mathbf{s}$. Then, the joint probability $p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)$ is generated by

$$p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) = p(x_i(\mathbf{s}))p\left(x_j(\mathbf{s}-\mathbf{h})|x_i(\mathbf{s})\right), \tag{7}$$

where $p\left(x_j(\mathbf{s}-\mathbf{h})|x_i(\mathbf{s})\right)$ is the conditional probability that the virtual neighbouring position $\mathbf{s}-\mathbf{h}$ belongs to the j^{th} state given that location $\mathbf{s}$ belongs to the i^{th} state. By collecting the observations of point pairs at spatial lag $\mathbf{h}$ apart, $p\left(x_j(\mathbf{s}-\mathbf{h})|x_i(\mathbf{s})\right)$ is estimated by

$$\hat{p}\left(x_j(\mathbf{s}-\mathbf{h})|x_i(\mathbf{s})\right) = n_{ij}/n_i, \tag{8}$$

where n_i is the number of point pairs at a spatial lag $\mathbf{h}$ apart taking the i^{th} state for at least one point ($X(\mathbf{s}) = x_i$ or $X(\mathbf{s}-\mathbf{h}) = x_i$), and n_{ij} is the number of those point pairs having both the i^{th} and j^{th}state ($X(\mathbf{s}) = x_i$ and $X(\mathbf{s}-\mathbf{h}) = x_j$, or $X(\mathbf{s}) = x_j$ and $X(\mathbf{s}-\mathbf{h}) = x_i$). Thus, we have

$$\sum_j n_{ij} = n_i, \quad (9)$$

and, therefore,

$$\sum_i \sum_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) = \sum_i \sum_j (n_{ij}/n_i)(N_i/N) = 1, \quad (10)$$

which means that the estimated probability mass function $p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)$ is valid. In summary, the entropogram $\tau(\mathbf{h})$ has several important properties, including:

1. $\tau(\mathbf{h})$ is non-negative (i.e., $\tau(\mathbf{h}) \geq 0$) and the necessary and sufficient condition for $\tau(\mathbf{h}) = 0$ is that $X(\mathbf{s})$ is independent from $X(\mathbf{s}-\mathbf{h})$. Theoretically, two RVs $X(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{h})$ are independent of each other, meaning that there is no association between them, when $\tau(\mathbf{h}) = 0$.
2. Based on the definition of the entropogram, the spatial association $\tau(\mathbf{h})$ between $X(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{h})$ is symmetric, i.e., $MSA(X(\mathbf{s}), X(\mathbf{s}-\mathbf{h})) = MSA(X(\mathbf{s}-\mathbf{h}), X(\mathbf{s}))$.
3. The spatial association between a variable and itself is the Shannon entropy of that variable (i.e., $\tau(0) = H\big(X(\mathbf{s})\big)$). Indeed, $H\big(X(\mathbf{s})\big)$ is the expectation of $\ln(1/p(x(\mathbf{s})))$, where $1/p(x(\mathbf{s}))$ can be understood as the level of *surprise* at a specific state of $X(\mathbf{s})$ being observed. In this way, the spatial association between a variable and itself takes the maximum value (i.e., $\tau(0) = H\big(X(\mathbf{s})\big) \geq H\big(X(\mathbf{s})\big) - H\big(X(\mathbf{s})\big|X(\mathbf{s}-\mathbf{h})\big) = \tau(\boldsymbol{h})$). This is intuitive as observations of one variable provide the greatest information about that variable relative to other variables.

3.2 Uncertainty analysis

Based upon a given sampling framework, the unknown true probabilities in equation (5) are estimated by the frequencies of occurrence of different variable states in the sample. At different spatial lags, the sample size may be different which will lead to variation in the estimation accuracy across spatial lags. In this section, the relationship between sample estimates and the unknown true probabilities is explored.

Let the unknown true probabilities p_i, $i = 1,2,\dots,m$, come from the sequence of mutually independent RVs, each of which takes on the state i with probability p_i. An estimate of the amount of Shannon entropy $\widehat{H}$ is obtained by the corresponding sample estimates of the state incidence $\hat{p}_i$ according to equation (2). Then, the estimated entropy can be expanded in a Taylor series at the point $(p_1,\dots,p_m)$,

$$\widehat{H} = H - \sum_{i=1}^{m}(1 + \ln p_i)(\hat{p}_i - p_i) - \frac{1}{2}\sum_{i=1}^{m}\frac{(\hat{p}_i - p_i)^2}{p_i} + \frac{1}{6}\sum_{i=1}^{m}\frac{(\hat{p}_i - p_i)^3}{(p_i + \theta(\hat{p}_i - p_i))^2}, \quad (11)$$

where $0 < \theta < 1$. A detailed derivation can be found in Appendix B. As the sample size increases, the estimate $\hat{p}_i$ will tend to the true probabilities p_i, thus,

$$\mathrm{E}\left(\hat{p}_i = \frac{N_i}{N}\right) \xrightarrow{N\to\infty} p_i. \quad (12)$$

where N_i is the occurrence number of a specific variable state i and N is the sample size.

Given the sample size N, the number of occurrences of a specific variable state i can be considered as a realization from the Binomial distribution $N_i \sim B(N, p_i)$. The variance of the corresponding sample estimates of the state incidence $\hat{p}_i$ is then obtained as

$$\mathrm{E}(\hat{p}_i - p_i)^2 = \mathrm{Var}\left(\frac{N_i}{N}\right) = \frac{p_i(1-p_i)}{N}. \quad (13)$$

We then have that

$$\mathrm{E}(\widehat{H}) = H - \frac{m-1}{2N}. \quad (14)$$

where m is the number of geographical variable states. And the variance of the sample estimates $\widehat{H}$ can be obtained and approximated by

$$\begin{aligned}\mathrm{Var}(\widehat{H}) &= \mathrm{E}\left(\widehat{H} - H - \frac{m-1}{2N}\right)^2 \\ &\cong \mathrm{E}\left(\sum_{i=1}^{m}(1+\ln p_i)(\hat{p}_i - p_i)\right)^2, \\ &= \frac{1}{N}\left(\sum_{i=1}^{m} p_i \ln^2 p_i - H^2\right)\end{aligned} \quad (15)$$

where terms of order of magnitude less than or equal to N^{-2} are neglected, and the RV $\widehat{H}$ is an asymptotically normal estimate of the corresponding Shannon entropy (Basharin 1959). According to equations (4) and (5), the entropogram at a specific lag is the sum of the Shannon entropy. Therefore, confidence intervals for the entropogram can be obtained by Monte Carlo methods. Specifically, it is possible to draw samples repeatedly from the asymptotically normal random variables $\widehat{H}(X(\mathbf{s}))$, $\widehat{H}(X(\mathbf{s}-\mathbf{h}))$, and $\widehat{H}(X(\mathbf{s}), X(\mathbf{s}-\mathbf{h}))$ simultaneously and calculate the corresponding entropogram values. The mean and variance of the three normal RVs are obtained using equation (14) and (15), respectively. Note that there is no restriction on the distribution of the geographical variables *per se*.

4. Results and Discussion

To evaluate the performance of the proposed entropogram against existing common measures of spatial association, we conducted a series of simulation experiments and a real-world case study. The simulation study explores the basic properties of the proposed entropogram compared to existing methods. Next, we applied the entropogram to land cover data to demonstrate its use in characterizing the second-order properties of real geographical data.

4.1 Numerical simulations

For the simplest case, three landscape maps with two variable states were simulated with different spatial patterns of black and white combinations (Figure 2a-c). The simulated spatial pattern is simple, and the study area consists of only 10 by 10 cells, providing great control over the experiments and results. The proposed entropogram is compared with the indicator variogram in Figure 2(d-f). As there are only two variable states, the indicator variogram can be used to characterize the variance information of the corresponding RF.

The proposed entropogram refers to the spatial dependence between simultaneous realizations of RVs at two locations, by indicating the variance information as the dispersion of the state co-occurrence between those locations. This contrasts with the covariance which reflects the joint variability, or say dissimilarity, of the two RVs at those locations. Specifically, the main differences between the entropogram and the variogram are illustrated in Figure 2(g-i). Given the state at one position, the entropogram depicts to what extent the state at another location is determined by the known state (i.e., it shows their dependence from the perspective of complexity). This is actually driven by the physical meaning of our used mutual information between the two RVs. For example, the entropogram value at spatial lag distance $||\mathbf{h}|| = 3$ increases towards a greater value at distance $||\mathbf{h}|| = 5$ (see Figure 2(d)). This increase is accompanied by the conditional probability transferred from a chaotic situation to the more deterministic circumstance as shown in Figure 2(g). Given the state of one location being black (white), therefore, a location at a lag distance $||\mathbf{h}|| = 5$ apart is more likely to be correctly predicted as being white (black). Hence, for this location, the conditional probabilities of states are distributed more unevenly, as compared to lag distance $||\mathbf{h}|| = 3$. However, such state entanglement cannot be revealed by the indicator variogram. Despite the dissimilarity between the states of pairs of RVs at different locations increasing with spatial lag distance $||\mathbf{h}|| = 2$ to $||\mathbf{h}|| = 5$, the correlation intensity between states is relatively stable (see Figure 3). This means that the complexity of the state co-occurrence is consistent at these two lags, just only the dominant correlation transferred from intra-state to inter-state. Figure 2(e) shows that the entropogram successfully characterized this kind of correlation intensity between states; whilst the indicator variogram can only describe the intensity of the difference between the states of pairs of RVs at different locations.

In addition, as the sample size for the calculation of the entropogram naturally varies with spatial lag distance, the 95% confidence intervals of the entropogram are provided in Figure 2 also. While the samples are abundant for small lags, Table further gives specific values of the 95% confidence intervals of the entropogram at spatial distance lags $\|\mathbf{h}\|$ =1, 4, 8, and 12, respectively, as well as the corresponding sample sizes as examples.

We calculated Moran's *I*, the join count statistic, symbolic entropy, and the conditional probability-based join count statistic (Bai et al. 2016), to measure the global spatial association of the landscape maps in Figure 2(a-c), see Appendix C. Their values are listed in Table 2. The symbolic entropy measures whether the 5-pixel surrounding pattern is significantly different from that of a random distribution or not. We applied the rook contiguity in cases where a weight matrix was needed. According to Moran's *I*, the spatial patterns for Figure 2(a-c) are negatively auto-correlated, positively auto-correlated and randomly distributed, respectively. However, these statistics fail to characterize the detailed spatial structure or variation of the spatial association.

Compared to the statistics of spatial association, the variogram can *de facto* reflect information about the co-variability of a geographical process under the spatial stationarity decision. The proposed entropogram has the potential to reflect this information directly for categorical data. Typical categorical variables such as soil types and land cover classes generally have multiple states and exhibit complex interclass relationships, as measured through the cross-correlation, neighbouring situation and directional asymmetry of class patterns. The proposed entropogram further transfers information from the conditional probability into a general measure of spatial association across spatial lags. The degree of spatial dependence at each spatial lag is positively related to the magnitude of the corresponding entropogram measurement, and consequently reflects the spatial variation of the underlying RFs. Besides, spatial association measures can be normalized by their deviation from that of the spatial data reproduced by reassigning randomly the variable states to each location, to compare the spatial patterns between different spatial datasets with different numbers of variable states or spatial extents.

We compared the entropogram with the multi-indicator variogram for categorical data with multiple states, by generating a multi-state landscape map with a known geographical process. To do so, we produced a continuous landscape map from a Gaussian RF with a covariance function $C(h) = \exp(-0.5h/1.5^2)$, see Figure 4(a). Then, we divided the range of the simulated continuous values into five equal-length intervals and transferred the continuous landscape map into a 5-state landscape map (Figure 4(b)). The corresponding entropogram is shown in Figure 4(c) as well as the multi-indicator variograms for each state. The sample multi-indicator variograms were fitted with exponential variogram models.

Compared to the multi-indicator variograms, the entropogram provides a comprehensive spatial association measure for the whole landscape instead of the inter-state spatial associations, while the variogram focuses on spatial co-occurrence data regarding each state. The resulting degree of spatial association between those data, however, has been identified as a poor proxy for ecological interactions (Blanchet, Cazelles and Gravel 2020). Besides, if the number of variable states increases, the number of indicator variograms will also increase, complicating the analysis (Atkinson, Cutler and Lewis 1997). In fact, it is not appropriate to apply the variogram simultaneously to multi-category data, as it aims to describe the dispersion of values as a function of the distance between the observation locations. In this way, the second-order property (i.e., equivalent to the covariance function) of the *categorical* RF generating Figure 4(b) can be revealed directly by the entropogram, in the same way that the variogram describes the variance information of a *continuous* RF. It is of interest that continuous data can be discretized and then analysed through various methods regarding frequency. Similarly, measures for categorical data can also be applied to continuous data.

To explore further the information captured by the entropogram, we show the detailed conditional probability distribution patterns, or transition probability matrix of states, in Figure 4(d). These transition probabilities are *de facto* the content of the transiogram which can be used to effectively generate realistic realizations of the real spatial distribution of multinomial classes and decreasing spatial uncertainty associated with the simulated results (Li, 2006). For locations at distance $\|\mathbf{h}\| = 1$ apart, once the variable state at location $\mathbf{s}$ has been observed, the variable state at location $\mathbf{s} - \mathbf{h}$ has a relatively high likelihood of being predicted correctly. This is because some of the variable states have only a small probability to exist at location $\mathbf{s} - \mathbf{h}$, given a state at location $\mathbf{s}$. In contrast, with respect to spatial lag $\|\mathbf{h}\| = 5$, the likelihood has little difference among the possible states at location $\mathbf{s} - \mathbf{h}$ given the variable state at location $\mathbf{s}$. In this circumstance, the observation of geographical variable at one location is of little use in predicting the variable state at another location, see Figure 4(d). That is, for this environment setting, variable states at locations at distance $\|\mathbf{h}\| = 1$ apart can provide more information on the potential variable state for each other compared to those at distance $\|\mathbf{h}\| = 5$ apart. This is reflected in the entropogram by the larger value at $\|\mathbf{h}\| =$

1 than $\|\mathbf{h}\| = 5$, see Figure 4(c). Then, at spatial lags $\|\mathbf{h}\| = 10$ and $\|\mathbf{h}\| = 15$, the corresponding likelihood gradually becomes stable across the variable states such that the values of the entropogram are almost unchanged.

A key property of the entropogram is that it can deal with different numbers of states from the perspective of complexity. To examine the impact of probability mass distribution patterns and numbers of states on the entropogram, we regrouped the continuous values in Figure 4(a) into three, five and seven categories with three different probability mass distribution patterns (i.e., Uniform, Pareto and Gaussian), respectively. The histograms of the nine generated landscape maps are shown in Figure 5(a). With the expansion of the virtual variable state space, the spatial association increases at small spatial lags (see Figure 5(b)) under a fixed probability mass distribution pattern. The change in the numbers of categories here is similar to the change of support as in the variogram; but the variation described by the variogram decreases with the expansion of the support, while the dependence described by entropogram increases with the expansion of the variable state space. At large spatial lags, the values of the entropogram are stable because there is weak spatial dependence, and this is independent of the richness of the variable state space. In addition to the number of states, the proposed entropogram tends to increase with the degree of randomness of the probability mass distribution patterns with a fixed variable state space. A likely explanation is that the entropogram measures the difference between the complexity of the point pattern and the conditional probability pattern. Variation in the probability mass distribution pattern changes both the realizations of two RVs, but keeps their conditional probability pattern relatively stable. Therefore, the complexity of the point pattern tends to increase with the randomness of the probability mass distribution patterns, and results in an increase in the values of the entropogram.

4.2 Real-world application

We now turn towards a real-world application, recognizing that categorical variables are important in a range of crucial domains such as climate change (Pielke Sr 2005) and carbon emission studies (Lai et al. 2016). They are used, for example, to express a rapidly growing demand for measurement and monitoring of the corresponding landscape-level patterns and processes. In this section, we applied the entropogram to analyze the spatial association of land cover types in Qinxian, Shanxi, China. The land cover data were collected from the Global Land Cover 2000 Project (Bartholome and Belward 2005) over a rectangular area between (111°47'53.87"E, 37°6'26.28"N) and (112°48'26.31"E, 36°12'22.66"N). Figure 6 shows the landscape map of the study area with six land cover types, including 1) broadleaved, deciduous and closed tree cover; 2) needle-leaved and evergreen tree cover; 3) burnt tree cover; 4) closed-open herbaceous cover; 5) cultivated and managed areas; and 6) water bodies.

Figure 7(a) shows the results of the entropogram for the smallest spatial lags. We found that the spatial association decreased with an increase in spatial lag for neighboring positions (small lags). Given the majority of existing spatial association measures focus on interclass relationships (e.g., the cross-correlation between any two variable states), the proposed entropogram integrates such interclass relationships into a comprehensive measure of the spatial association between two locations. In fact, the conditional probability between different land cover types (i.e., the probability transition matrix of states) exhibits different distribution patterns across spatial lags, which determine the magnitude of the entropogram at each spatial lag. When the spatial lag is 1 (i.e., for the adjacent land cover), given the land cover information about one location, the land cover type at another location is concentrated on one or two states only, making it easier to predict the corresponding land cover information. If we use this transition matrix to simulate a Markov process, the mixing time of all the states is longer than that at the other lags. Figure 7(b) shows the spectral gap of the probability transition matrix of states, where thin spectral gaps indicate slower mixing as there tends to be a singular transition between states, while large gaps indicate faster mixing representing a regular transition between states. Therefore, as the spatial lag distance increases, the spectral gaps also increase (Figure 7(b)) making it relatively difficult to acquire information on one location given information on another location separated by that spatial lag. In summary, the proposed entropogram can provide a general quantitative understanding of the state correlation across spatial lags.

The proposed entropogram can be applied for the spatial prediction and simulation of multi-categorical RFs, akin to the utility of the variogram for continuous RFs (Yao et al. 2021; Shakiba, and Doulati Ardejani 2022). Figure 8 provides an example of how the entropogram can be utilized potentially to predict the variable state on unknown locations with sample data. For a given location, on which the state was assumed as unknown, its state was estimated first from the 1-pixel neighbouring states, assumed to have been observed. For each observed 1-pixel neighbouring state, the corresponding conditional probability mass distribution of the given location was obtained by equation (8) where $||\mathbf{h}|| = 1$. The probability mass distribution of all land cover types was calculated by the mean conditional probability across all the observed 1-pixel neighbouring states. The state of the encircled pixel was then estimated by maximum likelihood based on its probability mass distribution. Similarly, we used the 2-pixel neighbouring states, excluding the 1-pixel neighbouring states, to predict the state of the given location. We found that the 2-pixel neighbouring states behaved better than the 1-pixel neighbouring states in the prediction of the states at the selected locations (Figure 8(d-e)). This suggests the necessity of a variogram-analogy model for categorical RFs, as the spatial association between RVs across different spatial lags may provide different information about the underlying categorical RFs. In this way, the entropogram helps to address the key issue of how to account for variation across lags. For example, the probability mass distribution based on both the 1- and 2-pixel neighbouring states can be calculated by the weighted average of the conditional probability against each observed state, where the weights are proportionally determined by their entropogram values (i.e., the corresponding entropogram value divided by the sum of all the entropogram values for all the involved observations). Figure 8(f) demonstrates that information from the 1-pixel spatial lag coupled with that from the 2-pixel spatial lag can provide efficient information on the pattern of the data. Nonetheless, as described above, a spatially stationary stochastic process commonly needs to be assumed due to the limited availability of repeatable spatial data. In this situation, the spatial data may be over-smoothed through modelling, and fractal characteristics may be neglected, because the entropogram, just like the variogram, is essentially a two-point statistic. Different RFs may, thus, possess the same entropogram, as for the variogram.

In this research, we focused on the empirical entropogram at different lags and, specifically, the transition probability matrix of geospatial categorical data, where the interpretation of the entropogram relies on its estimated values rather than the parameters of any model that might be fitted. Future research should investigate the relationship between covariance functions of RFs and the proposed entropogram to determine whether there exists a standard model, or set of models, that might be usefully fitted to the entropogram, akin to the fitting of a model to the sample variogram. The proposed mutual information described spatial association between two RVs could also be extended to more variables in future research (Li, Ren and Han 2022), to describe higher-order properties or more complex patterns, as in multiple-point geostatistics (Mariethoz and Caers 2014).

5. Conclusion

Measures of spatial association are important tools with which to analyse Earth science and other spatial data. Categorical spatial variables represent an important class of Earth science data, but measures of spatial association are less developed for categorical spatial data than those for continuous spatial variables. In this research, we introduce the entropogram as an entropy-based measure of spatial association for categorical variables, building on concepts underlying the variogram. Specifically, the entropogram quantifies the amount of shared information as a function of the separation lag vector, allowing prediction of the outcome of a spatial stochastic process at one location given its known variable state at another location. Compared to existing measures and models of spatial association for categorical variables, which focus mainly on inter-state relationships, the entropogram simultaneously characterizes the whole state space. As such, the entropogram is complementary to existing two-point statistics applied to categorical data, and can be extended to include other variables, for example, the spatial association between different geographical properties.

6. Data and Code availability

The datasets analysed during the current study are available in the Figshare repository with the following link: https://doi.org/10.6084/m9.figshare.21687905.v2

7. Acknowledgement

The authors would like to thank Prof. Gerard Heuvelink for his valuable comments, and Managing Editor Jennifer Cassidento for her kind help. This study was supported by the

National Natural Science Foundation of China (No. 41725006, 42230110, 41871286).

8. Author Contributions Statement

WBZ conceived and designed the study, built the model, collected data, finalised the analysis, interpreted the findings, and wrote the original manuscript. YG designed the study and interpreted the findings. HB and YJ collected data and reviewed the manuscript. AS reviewed and edited the manuscript. PMA conceived and designed the study, interpreted the findings, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

Anselin, L., 1995. Local indicators of spatial association—LISA. Geographical analysis, 27(2), pp.93-115.

Anselin, L. and Li, X., 2019. Operational local join count statistics for cluster detection. Journal of Geographical Systems, 21(2), pp.189-210.

Atkinson, P.M., Cutler, M.E.J. and Lewis, H., 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. International Journal of Remote Sensing, 18(4), pp.917-935.

Atkinson, P.M. and Tate, N.J., 2000. Spatial scale problems and geostatistical solutions: a review. The Professional Geographer, 52(4), pp.607-623.

Altieri, L., Cocchi, D. and Roli, G., 2018. A new approach to spatial entropy measures. Environmental and ecological statistics, 25(1), pp.95-110.

Basharin, G.P., 1959. On a statistical estimate for the entropy of a sequence of independent random variables. Theory of Probability & Its Applications, 4(3), pp.333-336.

Bartholome, E. and Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. International Journal of Remote Sensing, 26(9), pp.1959-1977.

Bai, H., Li, D., Ge, Y. and Wang, J., 2016. Detecting nominal variables' spatial associations using conditional probabilities of neighboring surface objects' categories. Information Sciences, 329, pp.701-718.

Benedetti, Y., Morelli, F., Munafò, M., Assennato, F., Strollo, A. and Santolini, R., 2020. Spatial associations among avian diversity, regulating and provisioning ecosystem services in Italy. Ecological Indicators, 108, p.105742.

Blanchet, F.G., Cazelles, K. and Gravel, D., 2020. Co‐occurrence is not evidence of ecological interactions. Ecology Letters, 23(7), pp.1050-1063.

Cliff, A.D. and Ord, K., 1970. Spatial autocorrelation: a review of existing and new measures with applications. Economic Geography, 46(sup1), pp.269-292.

Cliff, A.D. and Ord, J.K., 1981. Spatial processes: models & applications. Taylor & Francis.

Chiles, J.P. and Delfiner, P., 1999. Geostatistics: modeling spatial uncertainty. John Wiley & Sons, New York.

Cao, F., Ge, Y. and Wang, J.F., 2013. Optimal discretization for geographical detectors-based risk assessment. GIScience & Remote Sensing, 50(1), pp.78-92.

De Jong, P., Sprenger, C. and Van Veen, F., 1984. On extreme values of Moran's I and Geary's c. Geographical Analysis, 16(1), pp.17-24.

Fotheringham, A.S., 2009. "The problem of spatial autocorrelation" and local spatial statistics. Geographical analysis, 41(4), pp.398-403.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. The incorporated statistician, 5(3), pp.115-146.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press on Demand.

Garrigues, S., Allard, D., Baret, F. and Weiss, M., 2006. Quantifying spatial heterogeneity at the landscape scale using variogram models. Remote sensing of environment, 103(1), pp.81-96.

Goodchild, M.F., 2011. Scale in GIS: An overview. Geomorphology, 130(1-2), pp.5-9.

Ge, Y., Jin, Y., Stein, A., Chen, Y., Wang, J., Wang, J., Cheng, Q., Bai, H., Liu, M. and Atkinson, P.M., 2019. Principles and methods of scaling geospatial Earth science data. Earth-science reviews, 197, p.102897.

Journel, A.G., 1986. Constrained interpolation and qualitative information—the soft kriging approach. Mathematical Geology, 18(3), pp.269-286.

Koch-Janusz, M. and Ringel, Z., 2018. Mutual information, neural networks and the renormalization group. Nature Physics, 14(6), pp.578-582.

Li, W., 2006. Transiogram: a spatial relationship measure for categorical data. International Journal of Geographical Information Science, 20(6), pp.693-699.

Leibovici, D.G., Bastin, L., Anand, S., Hobona, G. and Jackson, M., 2011. Spatially clustered associations in health related geospatial data. Transactions in GIS, 15(3), pp.347-364.

Leibovici, D.G., Claramunt, C., Le Guyader, D. and Brosset, D., 2014. Local and global spatio-temporal entropy indices based on distance-ratios and co-occurrences distributions. International Journal of Geographical Information Science, 28(5), pp.1061-1084.

Lai, L., Huang, X., Yang, H., Chuai, X., Zhang, M., Zhong, T., Chen, Z., Chen, Y., Wang, X. and Thompson, J.R., 2016. Carbon emissions from land-use change and management in China between 1990 and 2010. Science Advances, 2(11), p.e1601063.

Li, J., Ren, W. and Han, M., 2022. Mutual Information Variational Autoencoders and Its Application to Feature Extraction of Multivariate Time Series. International Journal of Pattern Recognition and Artificial Intelligence, 36(06), p.2255005.

Moran, P.A., 1950. Notes on continuous stochastic phenomena. Biometrika, 37(1/2), pp.17-23.

Matheron, G., 1963. Principles of geostatistics. Economic geology, 58(8), pp.1246-1266.

Mariethoz, G. and Caers, J., 2014. Multiple-point geostatistics: stochastic modeling with training images. John Wiley & Sons.

Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K., Toxopeus, A.G. and Alibakhshi, S., 2019. ELSA: Entropy-based local indicator of spatial association. Spatial statistics, 29, pp.66-88.

Pielke Sr, R.A., 2005. Land use and climate change. Science, 310(5754), pp.1625-1626.

Ruiz, M., López, F. and Páez, A., 2010. Testing for spatial association of qualitative data using symbolic dynamics. Journal of Geographical Systems, 12(3), pp.281-309.

Shannon, C.E. and Weaver, W. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois.

Shakiba, S. and Doulati Ardejani, F., 2022. A comparative study of novel object-based geostatistical algorithm and direct sampling method on fracture network modeling. Stochastic Environmental Research and Risk Assessment, pp.1-17.

Wang, J.F., Li, X.H., Christakos, G., Liao, Y.L., Zhang, T., Gu, X. and Zheng, X.Y., 2010. Geographical detectors - based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. International Journal of Geographical Information Science, 24(1), pp.107-127.

Wang, J.F., Zhang, T.L. and Fu, B.J., 2016. A measure of spatial stratified heterogeneity. Ecological Indicators, 67, pp.250-256.

Yao, J., Liu, W., Liu, Q., Liu, Y., Chen, X. and Pan, M., 2021. Optimized algorithm for multipoint geostatistical facies modeling based on a deep feedforward neural network. Plos one, 16(6), p.e0253174.

Author biography

Wen-Bin Zhang is a PhD candidate in the University of Chinese academy of Science and a research assistant in the Institute of Geographical Science and Natural Resources Research, Chinese Academy of Science. E-mail: Zhangwb@lreis.ac.cn. His research interests include GIScience and complexity. Applications concern health geography, population dynamics, and discovering the fact of the world.

Yong Ge is a full professor in the Institute of Geographical Science and Natural Resources Research, Chinese Academy of Science. E-mail: Gey@lreis.ac.cn. Her research interests include Spatial statistics, Spatial data science including machine learning. Applications concern Poverty, land use/land cover change detection, and scaling Earth science data.

Hexiang Bai is a Professor in the School of Computer and Information Technology at Shanxi University, Taiyuan, Shanxi, China, 030006. E-mail: baihx@sxu.edu.cn. His research interests include spatial statistics and rough sets theory-based spatial datamining.

Yan Jin is a lecturer in the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China. E-mail: jinyan@njupt.edu.cn. Her current research interests include spatial statistics, data fusion, and remote sensing applications.

Alfred Stein is a full Professor in the Department of Earth Observation and Geo-information science (ITC) at the University of Twente, Enschede, The Netherlands. E-mail: a.stein@utwente.nl. His research interests include spatial statistics and image analysis and has a specific attention on satellite image analysis. Applications concern

geohealth, climate, environment, agriculture, natural vegetation and urban development.

Peter Atkinson is a Distinguished Professor in the Lancaster Environment Centre at Lancaster University, Lancaster, UK. E-mail: pma@lancaster.ac.uk. His research interests are in spatial statistics and spatial data science including machine learning and AI, with specific attention on issues relating to spatial scale and the change-of-support problem. Applications include land cover-land use change, climate induced vegetation phenology change and disease transmission systems.

List of figure captions

Figure 1. Conceptual framework based on Chiles and Delfiner (1999). Spatial data for a qualitative geographical variable are samples from a realization of an underlying geographical process, which can be described by a Random Function. The bottom row shows the modelling of such spatial data with concepts derived from Shannon entropy. Our proposed mutual information described spatial association measure is a combination of mutual information entropy and the variogram (in red) that can better characterise the properties of a geographical process.

Figure 2. Simulated landscape maps produced with two states representing three spatial patterns: (a) negatively auto-correlated, (b) positively auto-correlated and (c) randomly distributed. (d-f) Comparisons between the entropogram (in blue) and indicator variogram (in red) for landscape maps (a-c), respectively. (g-i) The conditional probabilities ($p\big(X(\mathbf{s}-\mathbf{h})|X(\mathbf{s})\big)$, see equation (8)) of states black (in blue) and white (in red) at different spatial lags. The left two bars are respective conditional probabilities of states for locations across different spatial lags, given that the true state of one location is black. The right two bars are corresponding cases given that the true state of one location is white.

Figure 3. Comparison between the information characterized by the entropogram and the indicator variogram. (a-b) Histograms of the state co-occurrence for Figure 2(b) at lags of 1 and 5, respectively. BB: black-black; BW: black-white; WB: white-black; WW: white-white. (c-d) The conditional probabilities of the states black and white.

Figure 4. (a) A realization of a Gaussian random function. (b) The corresponding 5-state landscape map produced from (a). (c) The entropogram and multi-indicator variograms of (b). (d) The transition probability matrix between states on $X(\mathbf{s})$ and $X(\mathbf{s}-\mathbf{h})$ with $||\mathbf{h}|| =$1, 5, 10 and 15.

Figure 5. (a) Histograms of discretized landscape maps, dividing Figure 3(a) into three, five and seven categories, and each with three different probability mass distribution patterns. (b) The corresponding estimated entropograms.

Figure 6. Landscape map of vegetation types in Qinxian, Shanxi, China.

Figure 7. (a) The entropogram and (b) the corresponding eigenvalue plot of transition probability matrix of states for spatial lags from 1 to 8. The unit of spatial lag is the pixel. An eigenvalue plot shows eigenvalues of the transition matrix of states on the complex plane. The spectral gap is the area between the radius with length equal to the second largest eigenvalue magnitude and the radius with a length of 1.

Figure 8. The land cover type of the randomly selected 5 pixels (marked by asterisk in yellow) were estimated by their adjacent pixels (marked by the square in yellow) for (a) 1-pixel contiguity, (b) 2-pixel contiguity and (c) both. (d-f) Each column depicts the estimated probability mass distribution of land cover types for each of the five selected pixels, where the true land cover type is given at the bottom. The land cover type with the greatest probability mass is labelled with the corresponding land cover type. CT: Tree Cover, broadleaved, deciduous, closed; ET: Tree Cover, needle-leaved, evergreen; BT: Tree Cover, burnt; HC: Herbaceous Cover, closed-open; CM: Cultivated and managed areas; and WB: Water Bodies.

Appendix

A.

$$
\begin{aligned}
\tau(\mathbf{h}) &= H(X(\mathbf{s})) + H(X(\mathbf{s}-\mathbf{h})) - H\big(X(\mathbf{s}), X(\mathbf{s}-\mathbf{h})\big) \\
&= -\sum\nolimits_i p\big(x_i(\mathbf{s})\big) \ln\Big(p\big(x_i(\mathbf{s})\big)\Big) \\
&\quad -\sum\nolimits_j p\left(x_j(\mathbf{s}-\mathbf{h})\right) \ln\left(p\left(x_j(\mathbf{s}-\mathbf{h})\right)\right) \\
&\quad +\sum\nolimits_i \sum\nolimits_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) \ln\left(p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)\right) \\
&= -\sum\nolimits_i \left(\sum\nolimits_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)\right) \ln\Big(p\big(x_i(\mathbf{s})\big)\Big) \\
&\quad -\sum\nolimits_j \left(\sum\nolimits_i p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)\right) \ln\left(p\left(x_j(\mathbf{s}-\mathbf{h})\right)\right) \\
&\quad +\sum\nolimits_i \sum\nolimits_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) \ln\left(p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)\right) \\
&= \sum\nolimits_i \sum\nolimits_j p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right) \ln\left(\frac{p\left(x_i(\mathbf{s}), x_j(\mathbf{s}-\mathbf{h})\right)}{p\big(x_i(\mathbf{s})\big) p\left(x_j(\mathbf{s}-\mathbf{h})\right)}\right)
\end{aligned}
$$

B.

The Shannon entropy estimated from samples, $\widehat{H}$, is built upon the estimated probability of the available states ($\hat{p} = (\hat{p}_1, \dots, \hat{p}_m)$) of the variable of interest. That is,

$$\widehat{H}(\hat{p}) = -\sum_{i=1}^{m} \hat{p}_i \ln(\hat{p}_i).$$

Here $\hat{p}_i$ is a variable, the value of which depends on samples. Given the true probabilities of the variable states $p = (p_1, \dots, p_m)$, the second-order Taylor polynomial of the above function $\widehat{H}(\hat{p})$ at the point p is

$$
\begin{aligned}
\widehat{H}(\hat{p}) = \widehat{H}(p) &+ \sum\nolimits_{i=1}^{m} \frac{\partial \widehat{H}}{\partial \hat{p}_i}(p)(\hat{p}_i - p_i) \\
&+ \frac{1}{2!} \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{m} \frac{\partial \widehat{H}}{\partial \hat{p}_i \partial \hat{p}_j}(p)(\hat{p}_i - p_i)(\hat{p}_j - p_j) \\
&+ \frac{1}{3!} \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1}^{m} \sum\nolimits_{k=1}^{m} \frac{\partial \widehat{H}}{\partial \hat{p}_i \partial \hat{p}_j \partial \hat{p}_k}(\xi_L)(\hat{p}_i - p_i)(\hat{p}_j - p_j)(\hat{p}_k - p_k).
\end{aligned}
$$

where ξ_L is some real vector between $\hat{p}$ and p. Given

$$\frac{\partial \hat{H}}{\partial \hat{p}_i} = -(1 + \ln(\hat{p}_i)), \quad \frac{\partial \hat{H}}{\partial \hat{p}_i \partial \hat{p}_j} = \begin{cases} -\frac{1}{\hat{p}_i} & j = i \\ 0 & j \neq i \end{cases}, \quad \frac{\partial \hat{H}}{\partial \hat{p}_i \partial \hat{p}_j \partial \hat{p}_k} = \begin{cases} \frac{1}{\hat{p}_i^2} & j = k = i \\ 0 & else \end{cases}.$$

We have

$$\hat{H}(\hat{p}) = H - \sum_{i=1}^{m} (1 + \ln p_i)(\hat{p}_i - p_i) - \frac{1}{2} \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i)^2}{p_i}$$

$$+ \frac{1}{6} \sum_{i=1}^{m} \frac{(\hat{p}_i - p_i)^3}{(p_i + \theta(\hat{p}_i - p_i))^2}$$

where H is the true Shannon entropy of the variable of interest, p represents the true probabilities of the variable states, and $0 < \theta < 1$.

C.

The Moran's I coefficient is defined as

$$I = \frac{N}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} z_i w_{ij} z_j}{\sum_{i=1}^{N} z_i^2}.$$

where the $z_i = x_i - \bar{x} = x_i - \sum_{i=1}^{N} x_i / N$ are the centred observations based on the original observations x_i, and w_{ij} is the element of a spatial weight matrix representing the hidden subjective relations between pairs of points. N is the total number of observations.

The join count statistic (JCS) is defined as

$$JCS(X) = \frac{1}{2} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} f(i,j) - W \left(1 - \frac{n_r n_s}{N(N-1)} \right) \right)$$

where $f(i,j)$ equals 1 if points i and j are the same category, W is the sum of values in the weight matrix, n_r and n_s are the number of observations for the presence and absence of the state of interest, respectively.

The conditional probability-based join count statistic (NCP) is defined as

$$NCP(X) = \begin{cases} \frac{CP(X)}{1-P_E} & CP(X) \geq 0 \\ \frac{CP(X)}{P_E} & \text{otherwise} \end{cases},$$

$$CP(X) = P\{X(\mathbf{s}) == X(\mathbf{s}+1)\}, \quad P_E = \sum_{i=1}^{m} p^2(X(\mathbf{s}) = i),$$

where X is the dataset with states $i = 1, \dots, m$, $CP(X)$ is the probability that pairs of locations with 1 pixel lag have the same category, and P_E is the theorical value of $CP(X)$ under the assumption of no spatial association.

The symbolic entropy (S) is defined using a symbolization procedure. The surrounding five spatial neighbours of $\mathbf{S}_0$ are defined by $N_{\mathbf{S}_0} = \{\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4\}$.

$\mathbf{S}_7$	$\mathbf{S}_4$	$\mathbf{S}_8$
$\mathbf{S}_3$	$\mathbf{S}_0$	$\mathbf{S}_1$
$\mathbf{S}_6$	$\mathbf{S}_2$	$\mathbf{S}_5$

Then, the surrounding five spatial neighbours of $\mathbf{S}_0$ are transformed by the indicator function

$$I_{\mathbf{S}_1,\mathbf{S}_2} = \begin{cases} 0 & X(\mathbf{S}_1) \neq X(\mathbf{S}_2) \\ 1 & \text{otherwise} \end{cases},$$

into $\sigma_{\mathbf{S}_0} = \{I_{\mathbf{S}_0,\mathbf{S}_1}, I_{\mathbf{S}_0,\mathbf{S}_2}, I_{\mathbf{S}_0,\mathbf{S}_3}, I_{\mathbf{S}_0,\mathbf{S}_4}\}$. Finally, the symbolic entropy is

$$S(5) = -\sum_{\sigma \in \Gamma} p(\sigma) \ln\big(p(\sigma)\big),$$

where $p(\sigma)$ is the relative frequency of a symbol σ based on all the observations, and Γ is the set of all possible symbols.

Table 1. Confidence intervals of the entropogram at spatial lag distances $\|\mathbf{h}\|$ =1, 4, 8, and 12 for Figure 2. N is the sample size used to estimate the corresponding confidence intervals.

Landscape maps	95% Confidence interval			
	$\mathbf{h} = 1$	$\mathbf{h} = 4$	$\mathbf{h} = 8$	$\mathbf{h} = 12$
	$N = 342$	$N = 850$	$N = 444$	$N = 8$
(a)	[0.073, 0.075]	[0.068, 0.068]	[0.059, 0.060]	[0.671, 0.716]
(b)	[0.266, 0.270]	[0.007, 0.007]	[0.279, 0.282]	[0.671, 0.694]
(c)	[0.081, 0.083]	[0.069, 0.070]	[0.40, 0.40]	[0.141, 0.197]

Table 2. Spatial association identified by Moran's *I* (*I*), join count statistic (JCS), symbolic entropy (S), and conditional probability-based join count statistic (NCP).

Landscape maps	I	JCS	S	NCP
(a)	-1	-45	399	-1
(b)	0.89	40	422	0.89
(c)	-0.14	-7	31	-0.14

Note: the rook contiguity was applied in calculations where a weight matrix was needed.