CrossMark

A new approach to spatial entropy measures

L. Altieri[1] · **D. Cocchi**[1] · **G. Roli**[1]

Abstract Entropy is widely employed in many applied sciences to measure the heterogeneity of observations. Recently, many attempts have been made to build entropy measures for spatial data, in order to capture the influence of space over the variable outcomes. The main limit of these developments is that all indices are computed conditional on a single distance and do not cover the whole spatial configuration of the phenomenon under study. Moreover, most of them do not satisfy the desirable additivity property between local and global spatial measures. This work reviews some recent developments, based on univariate distributions, and compares them to a new approach which considers the properties of entropy measures linked to bivariate distributions. This perspective introduces substantial innovations. Firstly, Shannon's entropy may be decomposed into two terms: spatial mutual information, accounting for the role of space in determining the variable outcome, and spatial global residual entropy, summarizing the remaining heterogeneity carried by the variable itself. Secondly, these terms both satisfy the additivity property, being sums of partial entropies measuring what happens at different distance classes. The proposed indices are used for measuring the spatial entropy of a marked point pattern on rainforest tree species. The new entropy measures are shown to be more informative and to answer a wider set of questions than the current proposals of the literature.

Handling Editor: Pierre Dutilleul.

✉ L. Altieri
linda.altieri@unibo.it

1 Department of Statistical Sciences, University of Bologna, Bologna, Italy

1 Introduction

Over the last decades, several attempts have been made to account for space in entropy measures. The main aim is twofold. On the one hand, there is a need to exploit the flexibility of entropy as a measure of heterogeneity of a variable X able to capture several aspects (information, surprise, diversity, uncertainty, contagion) and computable even for qualitative variables. On the other hand, knowledge about the spatial location of occurrences ought to be included, which may be influenced by what happens at surrounding locations, i.e. their neighbours.

The available indices are challenging attempts to include space into entropy measures, all based on the univariate approach of the classic Shannon's formula of entropy (1948). These contributions, belonging to the fields of geography, ecology and landscape studies, refer to two different approaches. The first one (Batty 1974, 1976, 2010; Karlström and Ceccato 2002) restricts the analysis to only one category of the variable of interest X at a time, and uncertainty pertains to the spatial distribution of the realizations of such category over pre-defined areas. Under this perspective, Karlström and Ceccato (2002) build a spatial entropy measure which can be decomposed in partial terms and additively recomposed following the idea of Local Indices of Spatial Association (LISA) proposed by Anselin (1995). The second approach (O'Neill et al. 1988; Li and Reynolds 1993; Riitters et al. 1996; Leibovici 2009; Leibovici et al. 2014; Parresol and Edwards 2014) transforms the study variable X into a new variable Z whose values account for the distance at which simultaneous realizations of X occur. The main advantage is to embed all categories of X and the chosen distance into one variable. Unfortunately, the resulting measures are not decomposable (therefore not additive), and conceived for one distance range at a time, since uncertainty in this case pertains to the categories of the new variable Z.

Starting from the variable Z, this work constitutes a new proposal for measuring spatial entropy which generalizes previous measures by taking advantage of the properties of entropy based on bivariate distributions defined in the information theory field (Cover and Thomas 2006; Stone 2015). This novel approach considers different distance ranges by an additional study variable W representing space; as a consequence, it allows to define entropy measures that are able to discern and quantify the role of space in determining the outcomes of the transformed variable Z, and therefore of X. Indeed, the entropy of Z can be decomposed into spatial mutual information, i.e. the entropy due to space, and spatial residual entropy, i.e. the remaining information brought by the variable itself once space has been considered. Furthermore, this proposal solves the problem of preserving additivity in constructing entropy measures, allowing for partial and global syntheses.

When spatial data are considered, outcomes are georeferenced over spatial units, which may be points or areas, and spatial entropy measures can be computed accordingly. In recent literature, heterogeneity has been measured over areal datasets, and, to our knowledge, very few works deal with point data (Leibovici et al. 2014). Areal datasets are preferred, when using entropy measures, because computations are easier with discrete space. Nevertheless, most spatial entropy measures can be employed for point data, as long as they present a finite number of categories. A dataset about the presence of tree species over Barro Colorado Island, Panama, is considered to assess
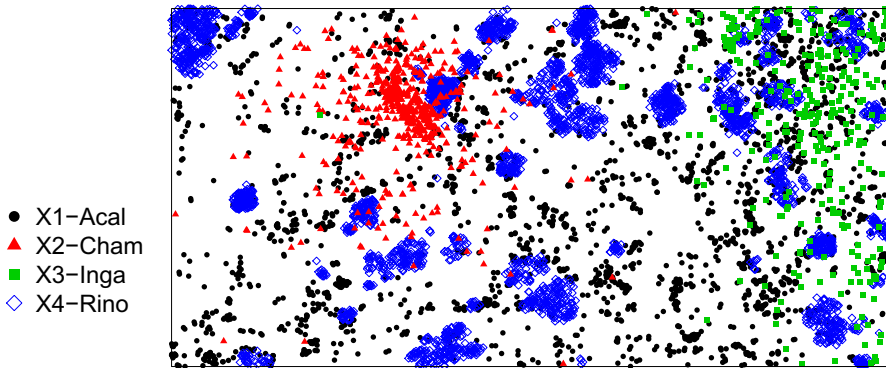
Fig. 1 Spatial distribution of 4 tree species over the Barro Colorado Island observation window

the performance of the proposed method compared to other available approaches (see, for instance, Rajala and Illian 2012). Barro Colorado Island has been the focus of intensive research on lowland tropical rainforest since 1923, and its flora is better known than in any site of comparable size throughout the world (http://www.ctfs.si. edu). Research identified several tree species over a rectangular observation window of size 1000×500 m, whose spatial distribution is of interest; the tree species constitute the point data categorical mark. This work focuses on 4 species with very different spatial configurations: Acalypha diversifolia, Chamguava schippii, Inga pezizifera and Rinorea sylvatica. The species identify the four categories of the variable of interest X, denoted as $x_i, i = 1, \ldots, 4$. Each dataset is a point pattern consisting, respectively, of $N_1 = 2678$, $N_2 = 544$, $N_3 = 311$ and $N_4 = 3718$ trees. The overall dataset has a total number of $N = 7251$ points and is reported in Fig. 1.

The theoretical aim of this paper is to illustrate a novel, rigorous methodology to exploit the properties of entropy based on bivariate distributions for taking the influence of space into account. The practical aim is to apply it to a case study for understanding to which extent space plays a role in the entropy of the four chosen tree species.

The paper is organized as follows. Section 2 introduces the basic concepts of univariate Shannon's entropy and reviews the approach based on Z. The new proposal, which exploits the properties of entropy coming from bivariate distributions, is described in Sect. 3. Then, the case study is analyzed via the different measures in Sect. 4. Finally, Sect. 5 discusses the main findings and outlines some concluding remarks.

2 Entropy and spatial entropy

The classic definition of entropy for a random variable X with I possible categories was introduced in 1948 by Shannon in his seminal work on information theory. Formally, Shannon's entropy is the expected value of a function, called information function, measuring the amount of information related to each category x_i, $i = 1, \ldots, I$. The term information is associated to the concepts of surprise and uncertainty: the

greater the surprise (and, thus, the uncertainty) in observing a value $X = x_i$, the greater the information it contains. The amount of surprise about an outcome value x_i increases as its probability decreases. In this spirit, the information function is defined as $I(p(x_i)) = \log(1/p(x_i))$, without any focus on the value of the outcome itself. The information function $I(p_X)$ is a random variable, where $p_X = (p(x_1), \ldots, p(x_I))'$ is the univariate probability mass function (pmf) of X. Shannon's entropy is defined as

$$H(X) = E[I(p_X)] = \sum_{i=1}^{I} p(x_i) \log\left(\frac{1}{p(x_i)}\right). \tag{1}$$

Being an expected value, it measures the average amount of information brought by the realizations of X as generated by the pmf p_X. When entropy is high, scarce information is available about the next realization, therefore the amount of uncertainty and heterogeneity of X is large. On the other hand, when the entropy is low, one is fairly sure about the next observation and thus uncertainty and heterogeneity of X are small. The probabilistic properties of entropy are often left apart in the applied literature, where entropy is seen as a heterogeneity index, which can be computed without the contribution of the value of the study variable for the different categories. Entropy $H(X)$ ranges in $[0, \log(I)]$, i.e. it is nonnegative and its maximum depends on the number of categories of X. The maximum value of entropy is achieved when X is uniformly distributed, while the minimum is only reached in the extreme case of certainty about the outcome.

In case of spatial data, i.e. data collected over an area, the spatial locations of occurrences are relevant, as they may influence the realizations of X over space. In particular, occurrences at certain locations may be related to what happens at surrounding locations. When heterogeneity has to be assessed, Shannon's entropy $H(X)$ is not able to discern the role of space. Hence, datasets with identical pmf p_X but different spatial distributions (e.g., strong spatial association and complete spatial randomness) yield the same $H(X)$.

An intriguing starting point to build an entropy measure accounting for space, namely a spatial entropy, consists in defining a new univariate categorical variable Z that identifies different types of couples of realizations of X over space, usually called co-occurrences, i.e. $(x_i, x_{i'})$, with $i, i' = 1, \ldots, I$. The consideration of co-occurrences underlies the idea of distance, which gives Z a promising role in the introduction of space in entropy measures. An assumption on whether to preserve the order within co-occurrences over space is needed: the realization $(x_i, x_{i'})$, with $i \neq i'$, is different from $(x_{i'}, x_i)$ if order is preserved; if order is neglected, such realizations coincide. This assumption influences the number of categories of Z, denoted by R: when order is preserved, R is $R^o = I^2$; conversely, R is $R^{no} = (I^2 + I)/2$. Under this perspective, Shannon's entropy (1) of Z, $H(Z)$, can be computed, using the pmf $p_Z = (p(z_1), \ldots, p(z_R))'$, as

$$H(Z) = E[I(p_Z)] = \sum_{r=1}^{R} p(z_r) \log\left(\frac{1}{p(z_r)}\right). \tag{2}$$

When the order is not preserved, $H(Z)$ does not depend on the spatial configuration of co-occurrences, Z maintains all the information of X, and $H(X)$ and $H(Z)$ quantify the same level of heterogeneity. Conversely, if order is preserved, the spatial direction of realizations within co-occurrences is considered; as a consequence, different $H(Z)$ can be obtained for different spatial configurations of the same set of realizations of X.

In order to properly account for space in an entropy measure based on Z, a definition of neighbourhood is needed. A way of representing a neighbourhood is via an adjacency matrix, i.e. a square matrix whose elements indicate which realizations of X are associated to form couples. In particular, for N spatial units the adjacency matrix $A = \{a_{uu'}\}_{u,u'=1,...,N}$ is a symmetric $N \times N$ matrix such that $a_{uu'} = 1$ when spatial units u and u' constitute a couple, and $a_{uu'} = 0$ otherwise; in other words, $a_{uu'} = 1$ if $u' \in \mathcal{N}(u)$, the neighbourhood of spatial unit u. The diagonal elements of A are usually all zero. Note that the spatial units may be points, defined via coordinate pairs, or areas, identified via a representative coordinate pair, such as the area centroid.

The definition of A implies that a variable $Z|A$ is constructed, whose realizations form a subset of Z outcomes that only includes co-occurrences identified by non-zero elements of A, i.e. conditioning on a fixed neighbourhood. Moreover, its conditional pmf $p_{Z|A} = (p(z_1|A), \ldots, p(z_R|A))'$ can be used to compute Shannon's entropy:

$$H(Z|A) = E[I(p_{Z|A})] = \sum_{r=1}^{R} p(z_r|A) \log \left(\frac{1}{p(z_r|A)} \right). \tag{3}$$

When an adjacency matrix A, with all off-diagonal elements equal to 1, is employed, expression (2) is a special case of (3). In all other cases, $H(Z|A)$ is indeed a spatial entropy measure. Most works which make use of the variable Z focus on the contiguity matrix, i.e. the simplest adjacency matrix A (O'Neill et al. 1988; Li and Reynolds 1993; Riitters et al. 1996; Parresol and Edwards 2014).

Leibovici (2009) and Leibovici et al. (2014) propose a spatial entropy like (3) based on $Z|A$ for a generic distance d, thus extending the idea of contiguity. The authors only develop the case of ordered co-occurrences, so that the number of categories of Z is $R^o = I^2$. The specific $Z|A$ is obtained by first fixing a distance d between occurences. Then co-occurrences are defined, for each d, as simultaneous realizations of X at any distance $d^* \leq d$, i.e. distances are taken into account according to a cumulative perspective; this way an adjacency matrix L_d is built and the variable which is constructed is $Z|L_d$. Then, Leibovici's spatial entropy is

$$H(Z|L_d) = E\left[I\left(p_{Z|L_d}\right)\right] = \sum_{r=1}^{R^o} p(z_r|L_d) \log \left(\frac{1}{p(z_r|L_d)} \right). \tag{4}$$

The probability $p(z_r|L_d)$ is, as occurs for all spatial entropy measures based on Z, the element of a univariate pmf $p_{Z|L_d}$, i.e. it is computed for a distribution conditional on L_d.

The approach just described is a first step for identifying a proper way to define a spatial entropy measure; nevertheless, two main disadvantages must be mentioned.

Firstly, all such spatial entropies are not decomposable, while additivity represents an appealing property in spatial statistics which follows the idea of Local Indices of Spatial Association (LISA) proposed by Anselin (1995). Secondly, they are based on conditional univariate distributions, so all results refer to an arbitrary choice of A.

3 Additive spatial entropy measures

This paper proposes a new class of spatial entropy measures based on the transformed variable Z, which exploits the properties of entropy based on bivariate distributions by introducing an additional study variable W representing space. The novelty of passing to a bivariate perspective has several advantages. In information theory, when two variables are jointly considered, suitable measures of entropy can be introduced (Cover and Thomas 2006; Stone 2015). In particular, mutual information is a measure of association that quantifies the information shared by two variables, while residual entropy measures the amount of information of one variable, when the effect of the other variable is removed. In our spatial context, these two measures allow to generalize previous spatial entropy measures, enjoy the desirable additivity property and help to quantify the role of space. Moreover, they receive a novel interpretation with respect to the tradition of information theory (Cover and Thomas 2006).

The random variable W defines intervals of all possible distances at which co-occurrences take place: they are denoted by $w_k =]d_{k-1}, d_k]$, with $k = 1, \ldots, K$ and d_k representing distances between points. The corresponding pmf is $p_W = (p(w_1), \ldots, p(w_K))'$. Each distance category w_k induces the definition of a specific adjacency matrix A_k which identifies co-occurrences at distance w_k and thus defines the study variable $Z|A_k$ with pmf $p_{Z|A_k}$. Therefore, $Z|A_k$ is from now on written as $Z|w_k$, as well as $p_{Z|A_k}$ is equivalent to $p_{Z|w_k}$. Under this perspective, K conditional univariate distributions cover the whole range of possible distances. The number and width of distance classes w_k is determined arbitrarily according to the specific case study.

The order within co-occurrences is not considered here. The first reason is that spatial neighbourhoods do not generally have a direction, thus co-occurrences with the same realizations in different order are undistiguishable in spatial statistics. The second reason is that neglecting the order of occurrences ensures the one-to-one correspondence between $H(X)$ and $H(Z)$, as said in Sect. 2.

3.1 Spatial residual entropy

Under the framework introduced above, K random variables $Z|w_k$ are considered, and the related conditional pmfs $p_{Z|w_k}$ are defined. Shannon's entropies $H(Z|w_k)$, computed on these variables, can be reinterpreted as spatial partial entropies:

$$H(Z|w_k) = E[I\left(p_{Z|w_k}\right)] = \sum_{r=1}^{R^{no}} p(z_r|w_k) \log\left(\frac{1}{p(z_r|w_k)}\right) \text{ for } k = 1, \ldots, K \quad (5)$$

and are the generalization of the previously described measures based on $Z|A$ in expression (3). For instance, if $w_1 = [0, d]$, then $A_1 = L_d$ and the two entropies $H(Z|w_1)$ and $H(Z|L_d)$ only differ as regards the number of terms due to order preservation. The term 'partial' refers to the fact that more than one entropy can be computed by considering different classes w_k covering the whole range of distances at which co-occurrences take place. Each partial entropy quantifies the amount of information brought by Z after the role of space has been controlled, i.e. conditional on a distance range w_k. This perspective allows to investigate the heterogeneity of the studied phenomenon by exploring the remaining entropy given the contribution of space at various levels. Indeed, spatial associations across the realizations of X can arise at different distances. Identifying these distances is a crucial topic for interpretation, and entropies $H(Z|w_k)$ properly achieve this goal. Moreover, spatial partial entropies are very flexible because distance classes w_k can be chosen by aggregating the less interesting ones and disaggregating the most informative ones.

The general meaning of the partial terms (5) is substantially different from Leibovici's entropy. The K partial entropies are computed on different distance ranges, considering K sets of couples which do not overlap with the others. Partial terms, therefore, consider different distance levels separately, while Leibovici's entropy, given d, embodies all couples within the fixed distance without distinction.

A nice characteristic of spatial partial entropies $H(Z|w_k)$, which are random variables, is that they can be summarized to obtain a well known global measure via weighting them by the probabilities $p(w_k)$ associated to each distance:

$$H(Z)_W = E[H(Z|W)] = \sum_{k=1}^{K} p(w_k) H(Z|w_k). \tag{6}$$

This measure is called residual entropy in information theory (Cover and Thomas 2006; Stone 2015). In a spatial context, it is redefined as spatial global residual entropy, and has crucial features. First, unlike usual Shannon's entropies, $H(Z)_W$ enjoys the desirable property of additivity. Second, since it exploits the joint distribution p_{ZW} of the two variables, $H(Z)_W$ is a measure of entropy based on a bivariate distribution, as can be appreciated by the following alternative formulation

$$H(Z)_W = E[E\left(I\left(p_{Z|w_k}\right)\right)] = \sum_{r=1}^{R^{no}} \sum_{k=1}^{K} p(z_r, w_k) \log\left(\frac{1}{p(z_r|w_k)}\right). \tag{7}$$

Third, spatial global residual entropy is able to quantify the residual entropy of Z by weighting the spatial partial entropies with the probabilities of the conditioning variable W. Thus, $H(Z)_W$ controls the contribution of space and measures the amount of information still brought by Z after removing the effect of the entire spatial configuration represented by W.

3.2 Spatial mutual information

In information theory, a measure of association of two variables in terms of entropy is defined as mutual information (Cover and Thomas 2006; Stone 2015). This quantity

measures the distance of the joint distribution of two variables from the condition of independence, i.e. from the product of their corresponding marginal distributions.

The mutual information for Z and W can be redefined as spatial mutual information $MI(Z, W)$, which is obtained as follows

$$MI(Z, W) = E\left[I\left(\frac{p_Z p_W}{p_{ZW}}\right)\right] = \sum_{r=1}^{R^{no}} \sum_{k=1}^{K} p(z_r, w_k) \log\left(\frac{p(z_r, w_k)}{p(z_r)p(w_k)}\right). \quad (8)$$

Spatial mutual information, like spatial residual entropy, is based on the bivariate distribution of Z and W. It is null when $p_{ZW} = p_Z p_W$ and moves away from 0 as the strength of association of Z and W increases.

The additivity property is also enjoyed by spatial mutual information, by rewriting $MI(Z, W)$ in (8) as

$$MI(Z, W) = \sum_{k=1}^{K} p(w_k) \sum_{r=1}^{R^{no}} p(z_r|w_k) \log\left(\frac{p(z_r|w_k)}{p(z_r)}\right), \quad (9)$$

where, analogously to (5), the k-th partial term, named spatial partial information, represents the contribution of space at distance range w_k:

$$PI(Z|w_k) = E\left[I\left(\frac{p_Z}{p_{Z|w_k}}\right)\right] = \sum_{r=1}^{R^{no}} p(z_r|w_k) \log\left(\frac{p(z_r|w_k)}{p(z_r)}\right). \quad (10)$$

This quantity measures the distance of each conditional distribution $p_{Z|w_k}$ from p_Z. The additivity property is thus respected the same way as in (6) for spatial residual entropy:

$$MI(Z, W) = \sum_{k=1}^{K} p(w_k) PI(Z|w_k). \quad (11)$$

The two syntheses $MI(Z, W)$ and $H(Z)_W$ can be summed to obtain the entropy of Z (Cover and Thomas 2006), that is

$$H(Z) = MI(Z, W) + H(Z)_W. \quad (12)$$

This expression shows that Shannon's entropy of Z can be decomposed into spatial mutual information and spatial global residual entropy. As a consequence, $MI(Z, W)$ receives a novel interpretation: it represents the part of entropy of Z due to the spatial configuration W.

In order to better identify the role of space, spatial mutual information can be presented in proportional terms as

$$MI_{prop}(Z, W) = \frac{MI(Z, W)}{H(Z)}, \quad (13)$$

where it can be seen that $H(Z)$ constitutes an upper limit for $MI(Z, W)$ (and also $H(Z)_W$). This is a relevant feature stating that, when order is not preserved, $H(Z)$ can always be seen as a reference value for interpretation.

Finally, by exploiting the decompositions (6) and (11) of $H(Z)_W$ and $MI(Z, W)$ respectively, together with (12), the entropy of Z can be further expressed as

$$H(Z) = \sum_{k=1}^{K} p(w_k) \left[PI(Z|w_k) + H(Z|w_k) \right]. \tag{14}$$

In other words, thanks to the bivariate perspective which introduces the additional variable W, Shannon's entropy $H(Z)$ can also be decomposed while respecting the additivity property, since the partial terms contribute to the explanation of the relationship between Z and W.

4 Measures' performance on rainforest tree data

Even if hardly any example can be found so far, all mentioned entropy measures can be used on point data. Shannon's entropy does not consider the spatial location of the data, but only the probabilities associated to the categories, which in this case are the four tree species of the Barro Colorado Island dataset. Therefore, Shannon's entropy of X and Z is only computed as a benchmark, and afterwards the measures considered in this work are computed for the tree species: Leibovici's spatial entropy, spatial residual entropy and spatial mutual information. The aim of the application is to assess the heterogeneity of the tree species over the territory, and how this is related to the spatial configuration via entropy measures. The computation of all measures contributes to highlight their properties and their ability to achieve detailed and interpretable results.

According to the observation window size of this case study, 8 distances d_k are considered in metres: $d_1 = 5$, $d_2 = 10$, $d_3 = 50$, $d_4 = 100$, $d_5 = 250$, $d_6 = 500$, $d_7 = 750$ and $d_8 = 1118$, the latter being the maximum possible distance inside the window (i.e. the rectangle diagonal). They are used both for the computation of several Leibovici's entropies (Sect. 4.2) and for the partial terms of the novel set of measures (Sect. 4.3).

For Leibovici's entropies, these distances are used to define classes of type $[0, d_k]$. Each class generates a $N \times N$ adiacency matrix L_{d_k}, where $N = 7251$, for $k = 1, \ldots, 8$. For the computation of spatial global residual entropy and spatial mutual information, distance classes are built using the same values d_1–d_8 as extremes: $w_k =]d_{k-1}, d_k]$, with $d_0 = 0$. Each distance class corresponds to an adiacency matrix A_k with the same size as L_{d_k}. Such choice of number and width of the classes w_k follows the general idea of spatial analysis that the most interesting distances are the small ones, therefore the first classes cover a smaller range than the last ones. The number of couples employed for the computation of the partial terms is considerably high for all classes, as shown in Table 1. If requested by the specific investigation, classes can be tuned accordingly. This choice of d_k allows the measures of entropy presented in Sect. 3 to be compared to Leibovici's entropies in terms of informativity and interpretability.

Table 1 Number of couples for each distance category (in thousands)

w_k	[0, 5]	]5, 10]	]10, 50]	[50, 100]	]100, 250]	]250, 500]	]500, 750]	]750, 1118]
No.	60.1	182.1	380	1132.9	5505.1	11298.9	5845.5	1880.2

Table 2 Proportions of tree species (variable X)

x_i	1	2	3	4
$\widehat{p}(x_i)$	0.369	0.075	0.043	0.513

Table 3 Proportions of unordered couples of tree species (variable Z)

z_r	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(2, 2)	(2, 3)	(2, 4)	(3, 3)	(3, 4)	(4, 4)
$\widehat{p}(z_r)$	0.136	0.055	0.032	0.379	0.006	0.006	0.077	0.002	0.044	0.263

Table 4 Shannon's entropy of X and Z

$H(X)$	$H(Z)$	$H(X)_{norm}$	$H(Z)_{norm}$
1.040	1.668	0.750	0.725

4.1 Shannon's entropy of X and Z

As regards the computation of Shannon's entropy (1) of X, the estimated probabilities $\widehat{p}(x_i)$, reported in Table 2, are the proportions of observed points of type x_i, i.e. $\widehat{p}(x_i) = N_i/N$ for $i = 1, \ldots, 4$, where N_i is the number of trees of species i described in Sect. 1.

The corresponding entropy is $H(X) = 1.04$, which is a relatively large value if compared to the maximum observable entropy, equal to $\log(4) = 1.386$. The departure from the uniform distribution is due to species 1 and 4, which are much more frequent than 2 and 3. The heterogeneity measured by $H(X)$ is only due to the frequency distribution of species, and does not account for the spatial configuration.

When the transformation Z of the study variable X is considered, the analogous non-spatial entropy measure is Shannon's entropy (2) of Z without order preservation. In such case, the number of categories of Z is $R^{no} = 10$, and $H(Z)$ is computed by means of the estimates $\widehat{p}(z_r)$ reported in Table 3. The number of couples over the observation window, considered for the computation of $H(Z)$, is over 26 millions.

The entropy value for the data is $H(Z) = 1.668$, with a maximum equal to $\log(10) = 2.303$. The departure from the uniform distribution of Z is again due to the inhomogeneous presence of the tree species.

For comparison purposes, entropy measures may be normalized, i.e. divided by their maximum value. Table 4 shows that Shannons' entropy of Z maintains the same proportion with respect to its maximum value as Shannon's entropy of X: the normalized values are very similar, as expected from the theory.
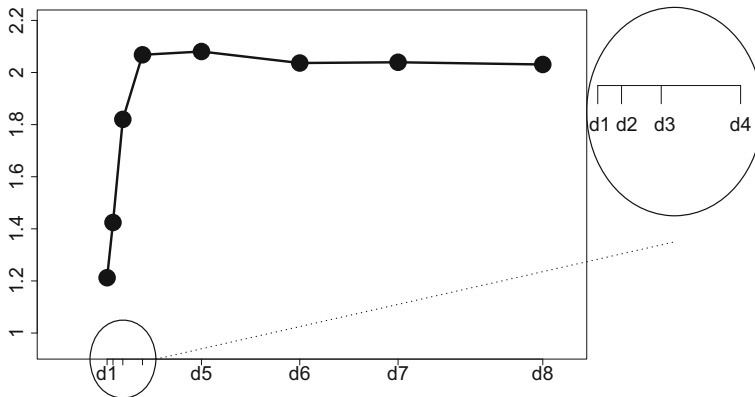
Fig. 2 Leibovici's entropy values at distances d_1–d_8

4.2 Leibovici's spatial entropy

Values for Leibovici's spatial entropy at distances d_1–d_8 are shown in Fig. 2.

The relative frequencies of the co-occurrence categories are used as estimates $\widehat{p}(z_r|L_{d_k})$ for each $H(Z|L_{d_k})$ by considering $R^o = 16$ ordered couples of tree species. Due to the way adjacency matrices L_{d_1}–L_{d_8} are built, the entropy value at each distance is computed including all the couples considered at previous distances. Values in Fig. 2 are not constant along distances, showing that the spread of the different tree species does not follow a spatially random pattern. Leibovici's entropy is able to detect some role of space, but the chained entropy values (linked by a line in Fig. 2) are less and less informative as distance increases. Therefore, the contribution of a large distance compared to a small one can only be indirectly appreciated, by considering the variation with respect to the previous values. Small distances are sensible to variation, but, as they increase, the relative weight of the previously considered couples becomes large and overwhelms innovations; therefore, after distance d_5, the entropy 'path' in Fig. 2 tends to flatness. This means that the variation in the contribution of space at the greatest distances cannot be captured any longer: a different investigation should be carried on. Another disadvantage of Leibovici's spatial entropy is that it may not be compared to a reference value in order to understand the strength of the role of space in the dataset. The actual reference value for Leibovici's entropy would be Shannon's entropy of Z, which is not a proper benchmark here due to order preservation, being influenced by the specific spatial configuration.

4.3 Spatial residual entropy and spatial mutual information

As for spatial residual entropy and mutual information, for each distance class w_k the relative frequencies of the unordered co-occurrences are computed as estimates $\widehat{p}(z_r|w_k)$.

Spatial partial entropies (5), computed at distance ranges w_1–w_8 and shown in Fig. 3, are the quantities to be compared to Leibovici's entropies of Sect. 4.2 and
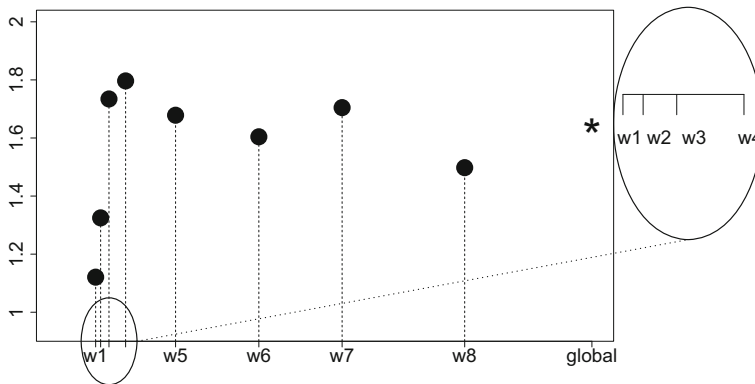
Fig. 3 Spatial partial entropy values at distance ranges w_1–w_8, plus spatial global residual entropy (*star on the right*)

Fig. 2 at distances d_1–d_8. Comparison requires some caution. First of all, spatial partial entropies consider unordered couples; as a consequence, even at the shortest distance $H(Z|w_1)$ differs from $H(Z|L_{d_1})$ for the number of categories of Z. At successive distance categories, each $H(Z|L_{d_k})$ considers all couples employed for building $H(Z|w_1)$ up to $H(Z|w_k)$, except for order preservation. This means that 'jumps' between entropy values at different distances in Fig. 2 may not be simply explained using the partial terms of Fig. 3, as they measure entropy on a different variable. Partial terms in Fig. 3 consider different distance levels separately, while Leibovici's entropies count all couples within each fixed distance without distinction. Therefore, variability in the residual information offered by (5) can be appreciated over large distance ranges as well as small ones.

Spatial partial entropies are built in a way that allows straightforward interpretation: they express the residual amount of entropy of Z (and consequently of X) after taking space into account. Therefore, they can identify the distance ranges at which residual entropy is high (or low), i.e. at which the role of space is weak (or strong). It is to remember that the partial terms are able to detect any departure from a random pattern. Results at short distances are very similar to the ones of Leibovici's spatial entropy. This accordance suggests that the main spatial pattern identified at short distances is a clustered one. This is likely due to the spatial configuration of x_4, which presents many small clusters, and of x_1, which has one main core of small size with respect to the whole observation window. At distances w_6 and w_8, the partial terms highlight a decrease in residual entropy, which is not detected by Leibovici's measures. The decrease at large distances in Fig. 3 identifies a repulsive configuration: most trees tend to have neighbours of the same species at short distances, and neighbours of different species at large distances. The increase in spatial residual entropy at distance w_7 may be imputed to a higher number of neighbours of the same species for x_4.

The 'star' point in Fig. 3 represents the spatial global residual entropy (6), which is a weighted mean of all other values and returns the general level of residual entropy of the four tree species.

The most informative quantities in this proposal are the partial terms, while spatial global residual entropy becomes crucial in order to obtain the part of entropy due
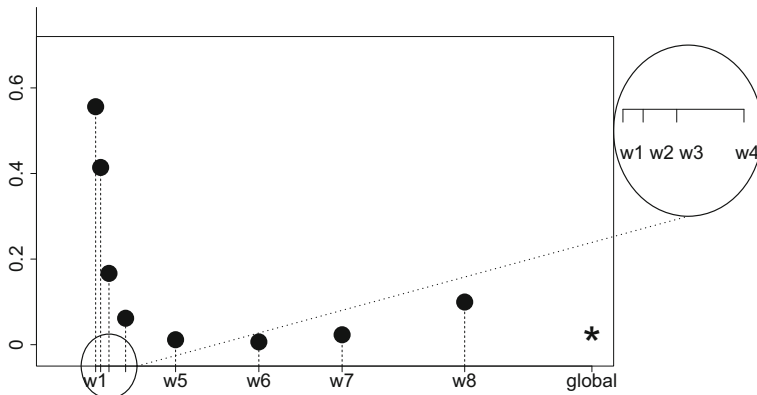
Fig. 4 Spatial partial information at distance ranges w_1–w_8, plus spatial mutual information (*star on the right*)

to space, i.e. spatial mutual information $MI(Z, W)$, which subtracts $H(Z)_W$ from $H(Z)$ as in (12). Spatial mutual information at a global level is 0.027 (star in Fig. 4); its proportional version (13) is about 2% of the entropy of Z, representing the percentage of entropy due to the spatial configuration. Thus, spatial global residual entropy approaches Shannon's entropy of Z. This shows that the global value may carry little information, and that exploring the spatial partial information terms is crucial.

Spatial partial information terms (10) in Fig. 4 allow to understand in more detail how much space affects entropy, and at which distance ranges. They are coherent with spatial partial entropies. As already said for the partial terms of spatial residual entropy, these measures also have the advantage of being easily interpretable: for instance, they say that at distance w_1 56% of entropy is due to the influence of space in tree presence.

Table 5 highlights the necessary ingredients for all computations and shows how the additivity property works for all quantities: $MI(Z, W)$ as in (11), $H(Z)_W$ as in (6) and $H(Z)$ as in (14). Being expectations, they all are computed as weighted sums of partial expectation terms at local distance ranges. The first column reports the weights, i.e. the estimate for p_W, obtained as the proportion of couples within each distance range w_k with respect to the total number of couples. A close look at the partial terms and weights $\widehat{p}(w_k)$ explains why spatial mutual information is so low at the global level: in this particular dataset there is a high number of associations of X occurring at distances w_5, w_6 and w_7 apart, which therefore receive the greatest weights at the distances with the lowest role of space. The values of Table 5 support the importance to focus on the partial terms in order to achieve the correct conclusions, without being misled by the global result alone. This is a great advantage of the spatial entropy measures here proposed.

5 Discussion and conclusions

This work illustrates a new approach to measure spatial entropy on data associated to a finite number of categories. It provides a toolbox for analyzing spatial data where distance is believed to play a role in determining the heterogeneity of the outcomes.

Table 5 Partial terms and additivity property

W	$\widehat{p}(w_k)$	$PI(Z|w_k)$	$H(Z|w_k)$	$PI(Z|w_k) + H(Z|w_k)$
w_1	0.002	0.556	1.121	1.677
w_2	0.007	0.414	1.325	1.739
w_3	0.014	0.167	1.734	1.901
w_4	0.043	0.062	1.797	1.859
w_5	0.209	0.012	1.678	1.690
w_6	0.430	0.006	1.604	1.610
w_7	0.222	0.023	1.705	1.728
w_8	0.072	0.100	1.498	1.598
	$MI(Z, W)$	$H(Z)_W$	$H(Z)$	
	0.027	1.642	1.668	

As a first step of this analysis, Shannon's entropy of the transformed variable Z is computed and used as a reference value. It represents the level of heterogeneity of the study variable X, suitably transformed to account for space. The second step defines distance classes and computes spatial mutual information, in order to identify the strength of the overall role of space in determining the data distribution. Thirdly, the partial information terms are computed in order to investigate the contribution of space to the data behaviour. In particular, they help to understand whether space plays a relevant role at each distance class. Then, spatial partial entropies are built, in order to focus on the heterogeneity of the study variable due to other sources. The comparison of partial terms across distances is also helpful to grasp the spatial behaviour of the study variable.

In particular, for the rainforest tree data of Barro Colorado Island, the proposed measures identify a very low influence of space in determining the tree species distribution at a global level. This should not make one jump to conclusions before analyzing the partial terms. Indeed, the global result is due to the high weights given to distances w_5–w_7 (Table 5), where many different associations of trees occur, yielding low spatial partial information terms. Nevertheless, two different kinds of space influence may be detected. At the shortest distance ranges (w_1–w_4), the tree species distribution is partially ruled by space. This can be ascribed to a clustering behaviour, especially due to species x_2, x_3 and x_4. Moreover, at the greatest distance (class w_8) the role of space becomes relevant again, due to the repulsive behaviour of species x_1 and the multi-cluster distribution of species x_4. The sign of association may be assessed by visual interpretation (looking at Fig. 1) and/or by looking at the $\widehat{p}(z_r|w_k)$. If couples of type (x_i, x_i) are more frequent, a clustering behaviour is detected. Conversely, repulsion takes place when the presence of couples of type $(x_i, x_{i'})$, $i \neq i'$, is higher. If needed, spatial correlation indices may be computed at the most interesting distances.

As for the theoretical properties of the approach proposed in this work, a substantial innovation with respect to all previous measures available in the literature is that this proposal successfully exploits the properties of entropy based on bivariate distributions. Therefore, it allows the entropy of the categorical variable Z to be decomposed

in two terms: the first term, spatial mutual information, accounts for the role of space in determining the data configuration, while the second term, spatial residual entropy, is a noise term summarizing the remaining information carried by the data. Another advantage of this approach comes from neglecting the order of the associations of X that build the categories of Z. First of all, this is sensible because spatial phenomena are not usually assumed to have a direction, as the interest lies in capturing the spatial heterogeneity of data over a specific area considering neighbourhood in any direction. Secondly, the proposal allows to compute a single Shannon's entropy of Z, which is a reference value for understanding the amount of spatial and of residual information.

As for the practical advantages, Sect. 4 shows how each Leibovici's measure does not allow any deeper inspection within the obtained entropy value, whereas (6) and (11) can investigate what happens at any desired distance range. Moreover, even though Leibovici's entropy may be computed at different d values, results are too chained to carry substantial information. Conversely, partial terms of the residual entropy and mutual information may assume very different values, as they reflect the data behaviour at non-overlapping distance intervals. They are very informative and allow straightforward interpretation about the phenomenon under study.

References

Anselin L (1995) Local indicators of spatial association—LISA. Geogrl Anal 27(2):94–115

Batty M (1974) Spatial entropy. Geogr Anal 6:1–31

Batty M (1976) Entropy in spatial aggregation. Geogr Anal 8:1–21

Batty M (2010) Space, scale, and scaling in entropy maximizing. Geogr Anal 42:395–421

Cover T, Thomas J (2006) Elements of information theory, 2nd edn. Wiley, Hoboken

Karlström A, Ceccato V (2002) A new information theoretical measure of global and local spatial association. Jaharb Regionalwissensc 22:13–40

Leibovici D (2009) Defining spatial entropy from multivariate distributions of co-occurrences. In: Hornsby KS et al. (eds) 9th international conference on spatial information theory 2009. Lecture Notes in Computer Science vol 5756, Springer, Berlin, pp 392–404

Leibovici D, Claramunt C, Guyader DL, Brosset D (2014) Local and global spatio-temporal entropy indices based on distance ratios and co-occurrences distributions. Int J Geogr Inf Sci 28(5):1061–1084

Li H, Reynolds J (1993) A new contagion index to quantify spatial patterns of landscapes. Landscape Ecol 8(3):155–162

O'Neill R, Krummel J, Gardner R, Sugihara G, Jackson B, DeAngelis D, Milne B, Turner M, Zygmunt B, Christensen S, Dale V, Graham R (1988) Indices of landscape pattern. Landscape Ecol 1(3):153–162

Parresol B, Edwards L (2014) An entropy-based contagion index and its sampling properties for landscape analysis. Entropy 16(4):1842–1859

Rajala T, Illian J (2012) A family of spatial biodiversity measures based on graphs. Environ Ecol Stat 19:545–572

Riitters K, O'Neill R, Wickham J, Jones K (1996) A note on contagion indices for landscape analysis. Landscape Ecol 11(4):197–202

Shannon C (1948) A mathematical theory of communication. Bell Syst Tech J 27(379–423):623–656

Stone J (2015) Information Lheory. A tutorial introduction. Sebtel Press, Sheffield

L. Altieri is Research Fellow in the Department of Statistical Sciences at the University of Bologna. She received her degree in Statistics and her Ph.D. in Statistics from the University of Bologna. Her research interests are in point processes, Bayesian statistics and environmental statistics.

D. Cocchi is Professor of Statistics in the Department of Statistical Sciences at the University of Bologna. She received her degree in Statistics from the University of Bologna and her Ph.D. in Statistics from the Université Catholique de Louvain. Her research interests are in finite population inference, Bayesian statistics and environmental statistics.

G. Roli is Assistant Professor of Statistics in the Department of Statistical Sciences at the University of Bologna. She received her degree in Statistics and her Ph.D. in Statistics from the University of Bologna. Her research interests are in causal inference, Bayesian statistics and social statistics.