Advances in spatial entropy measures

Published Version:
Altieri L., C.D. (2019). Advances in spatial entropy measures. STOCHASTIC ENVIRONMENTAL RESEARCH AND RISK ASSESSMENT, 33(4-6), 1223-1240 [10.1007/s00477-019-01686-y].

Availability:
This version is available at: https://hdl.handle.net/11585/725265 since: 2020-02-13

Published:
DOI: http://doi.org/10.1007/s00477-019-01686-y

Noname manuscript No.
(will be inserted by the editor)

Advances in spatial entropy measures

Linda Altieri · Daniela Cocchi · Giulia Roli

Received: date / Accepted: date

Abstract A very recent proposal of a set of entropy measures for spatial data, based on building pairs of realizations, allows to split the data heterogeneity that is usually assessed via Shannon's entropy into two components: spatial mutual information, identifying the role of space, and spatial residual entropy, measuring heterogeneity due to other sources. A further decomposition into partial terms deeply investigates the role of space at specific distance ranges. The present work proposes improvements to the method and adds relevant results proving that the new set of spatial entropies satisfies a list of desirable properties. We extend the methodology to sets of realizations greater than pairs. We also show that the approach is more general, better performing and more interpretable than the most popular proposals in the literature, thanks to the property of additivity and a new way of computing entropy that explicitly discards the order within sets. A novel procedure for building the necessary quantities for computations is also provided. A comparative study illustrates the superior performance of the new set of measures over representative spatial configurations. Practical questions are answered by means of a case study on land use data.

Keywords Shannon's entropy · Residual entropy · Mutual information · Additivity property · Spatial entropy · Categorical variables.

L. Altieri
University of Bologna, Department of Statistical Sciences
via Belle Arti, 41, 40126 Bologna
Tel.: +39-051-2098201
Fax: +39-051-2086242
E-mail: linda.altieri@unibo.it

D. Cocchi
University of Bologna, Department of Statistical Sciences
via Belle Arti, 41, 40126 Bologna
E-mail: daniela.cocchi@unibo.it

G. Roli
University of Bologna, Department of Statistical Sciences
via Belle Arti, 41, 40126 Bologna
E-mail: g.roli@unibo.it

1 Introduction

When a set of units can be assigned to a finite number of categories of a study variable, a standard way of assessing heterogeneity is to compute entropy. The seminal work by Shannon (1948) provided the basics to define entropy, and Shannon's formula of entropy, initially proposed in Information Theory, has rapidly become popular in many applied sciences, e.g. ecology and geography (Patil and Taillie 1982; Hoeting et al 2000; Frosini 2004; Leinster and Cobbold 2012). Hydrology is a further discipline where entropy based measures received great attention (Butera et al 2018). The reasons for the success of this index are two-fold. First of all, entropy is a measure of diversity that only considers the number of categories of the study variable and their probabilities; thus, it can be employed in a wide range of applications, even when qualitative variables are involved. In addition, entropy summarizes and captures several aspects that are differently denoted according to the specific target: heterogeneity, information, surprise, diversity, uncertainty, contagion are all concepts strongly related to entropy.

A relatively recent research field aims at accounting for space in entropy measures, as a natural generalization when the spatial location of the occurrences of the variable under study is available and relevant. Over the past decades, several works belonging to the fields of geography, ecology and landscape studies proposed measures including spatial information. These can be ascribed to two main approaches. The first starts with Batty (1974, 1976, 2010) who extends Theil's work (1972) to define a spatial entropy measure accounting for unequal space partition into sub-areas. Later, Karlström and Ceccato (2002) modified the initial proposal in order to satisfy the property of additivity in terms of decomposition of the global index into local components. A few main drawbacks of this setting should be highlighted: first, such entropy can only be computed for a binary variable, i.e. presence/absence of an attribute at each location. In addition, the local terms are not entropies and do not possess the properties of the global one. Lastly, results are heavily affected by the selected area partition. The second approach to spatial entropy introduces a different way of including space based on a suitable transformation of the study variable to account for the distance between realizations (co-occurrences); a first proposal is made by O'Neill et al (1988) for contiguous couples, extended by Leibovici (2009) and Leibovici et al (2014) to further distances and general degrees of co-occurrences. Several indices of contagion (Li and Reynolds 1993; Riitters et al 1996; Parresol and Edwards 2014) are also based on this view. Claramunt (2005) proposed a different way of including spatial information in terms of relative positioning of areas, exploiting the ratio of average distances between pairs of realizations of the same category to those between pairs of different categories. Unfortunately, the measures resulting from these distance-based approaches do not enjoy the additivity property. Moreover, results rely on an exogenous choice: they are computed conditional on a single distance and cannot capture the overall spatial behaviour of the variable of interest.

A set of spatial entropy measures has been recently presented by Altieri et al (2017), which fulfils many desirable properties. It also has the potential to add innovative elements to the Bayesian Maximum entropy approach (He and Kolovos 2018). The proposal starts from the co-occurrence approach and focuses on pairs of realizations, but overcomes the lack of relevant features of standard measures in the literature (Batty 1974; O'Neill et al 1988; Li and Reynolds 1993; Karlström and Ceccato 2002; Leibovici 2009). Shannon's entropy of the transformed variable, typical of the second approach, is decomposed into the information due to space and the remaining information brought by the variable itself once space is considered. The proposal solves the problem of preserving additivity and disaggregating results, allowing for partial and global syntheses.

The present work extends the approach in Altieri et al (2017) with major innovations and proofs: the key advantages are shown with regard to previously proposed measures, while answering practical questions of interest both with simulated and real data. Firstly, the methodology for transforming the study variable in order to account for space, presented for the very specific case of pairs in Altieri et al (2017), is generalized to greater sets of realizations, i.e. further degrees of co-occurrences. The idea of Leibovici (2009) is presented in an innovative way here, combined to the decomposition of Shannon's entropy in Altieri et al (2017), in order to disaggregate spatial mutual information, a crucial quantity that had never been exploited for spatial entropy before. Secondly, major advantages with regard to the spatial entropy indices in O'Neill et al (1988), Li and Reynolds (1993), Leibovici (2009), Leibovici et al (2014) and Parresol and Edwards (2014), that were briefly shown empirically for a special case in Altieri et al (2017), are now discussed in theory, through simulation and with real data. We first show that the traditional way of computing spatial entropy preserves the spatial ordering of realizations, implying a series of crucial drawbacks. We thus propose to discard the order and highlight that this choice brings substantial improvements: reductions in the computational burden and the identification of a unique maximum value for the entropy of the transformed variable, which is essential for interpretation and comparison. Moreover, a novel procedure is proposed to build the needed co-occurrences and the related quantities for the computation of the new set of measures. We show how to count co-occurrences at any desired distance range over the observation area, by using adjacency matrices. Besides, the superiority of the proposed measures and other relevant features are tested via a simulation study, which compares the performance of the novel approach with standard spatial entropy measures over the crucial spatial configurations (compact, repulsive, multicluster and random). The study is a substantial complement to the theoretical advantages, and enhances the interpretability issues of O'Neill et al (1988), Li and Reynolds (1993), Leibovici (2009), Leibovici et al (2014) and Parresol and Edwards (2014), due to the fact that they can only be computed on a single distance and do not consider the overall spatial configuration, nor can be decomposed. Under the crucial representative scenarios, the answers given by stardard entropy measures are ambiguous and possibly leading to misinterpretation. The extended simulation study stresses the added value of the new set of measures. Lastly, an environmental application is carried out to show the flexibility and informativity of the proposed measures. Raster data are used for the Italian Region Emilia Romagna. The pixel size (originally 1250×1250 metres) is set as 1, so that the observation area, i.e. the rectangle enclosing Emilia Romagna, is 228 units wide along the latitude, and 121 units wide along the longitude. Within the rectangular grid, the Region territory is made of 14173 cells. The variable of interest is $X =$ 'land use' and has $I = 4$ categories: x_1 'urban areas', x_2 'agricultural/artificially vegetated areas', x_3 'forests/seminatural areas' and x_4 'water areas'. The resulting map is in Figure 1. The case study shows the power and potential of the proposed set of spatial entropy measures.

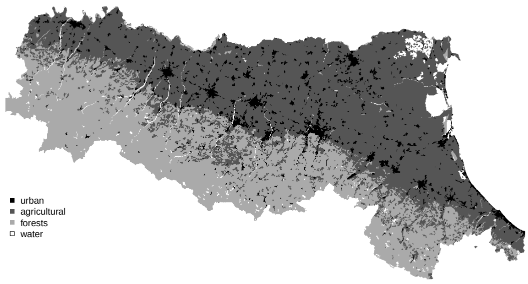

Fig. 1 Emilia Romagna Region, with 4 land cover categories

The paper is organized as follows. In Section 2, some necessary background notions are introduced. Section 3 extends the innovative way to deal with space in entropy measures and focuses on the theoretical advantages of the new approach. The proposed measures are evaluated on simulated data in Section 4 and applied to land use data in Section 5. Section 6 discusses the main findings.

2 Basics for building spatial entropy measures

Let X be a discrete random variable which takes values x_i in a set of I outcomes, $i = 1, \ldots, I$. Let $I(p_X)$ be the information function, where $p_X = (p(x_1), \ldots, p(x_I))'$ is the univariate probability mass function (pmf) of X: $I(p(x_i)) = \log(1/p(x_i))$, so that the amount of information about an outcome x_i increases as its probability decreases. Shannon's entropy of X is defined as the expected value of the information function:

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

Entropy quantifies the average amount of information brought by X according to the pmf p_X; it ranges in $[0, \log(I)]$ and its maximum value is achieved when X is uniformly distributed. A major drawback of such entropy is that it does not account for the spatial location of occurrences, so that datasets with identical pmf but very different spatial configurations share the same entropy.

An entropy measure that accounts for space, namely a spatial entropy, implies the formal definition of a neighbourhood. The concept of neighbourhood (Cressie 1993) means that occurrences at certain spatial units are influenced, in a positive or negative sense, by what happens at surrounding units, i.e. their neighbours. Spatial units may be points, defined via coordinate pairs, or areas, identified via representative coordinate pairs, such as the area centroids. Spatial units occur over the 'observation window': a fixed, limited spatial region with known size and shape; the spatial phenomenon under study potentially exists everywhere, but is only detected over the observation window. 'Distances' are always intended as Euclidean distances between coordinate pairs on the two-dimensional space. The neighbourhood system over a set of spatial units may be identified according to any chosen criterion; usually, two units are considered neighbours if they fall within a fixed distance. The spatial extent of the influence among units, i.e. the choice of the neighbourhood system, is commonly fixed exogenously. The system can be represented by a graph (Bondy and Murty 2008), where each location is a vertex and neighbouring locations are connected by edges. The simplest way of synthesizing a neighbourhood system over N spatial units is via an adjacency matrix (Anselin 1995; Bondy and Murty 2008), i.e. a square matrix whose elements indicate whether pairs of vertices are adjacent or not in the graph: $a_{uu'} = 1$ if $u' \in \mathcal{N}(u)$, that is the neighbourhood of area u, with $u = 1, \dots, N$; $a_{uu} = 0$ by definition. Note that u identifies the spatial unit, while i identifies the category of the variable X: each location u carries a spatial realization x_u, which presents one of the I categories: $x_u \in \{x_1, \dots, x_i, \dots, x_I\}$ for all u.

When working with spatial data, one should use the finest available resolution, i.e. points if data are a point pattern, or the finest grid provided if data are lattice; this is the case in the remainder of the paper.

3 A range-occurrence approach for spatial entropy measures

In order to suitably define spatial entropy measures, a series of desirable properties needs to be satisfied. Firstly, such measures should be able to split the part of entropy due to the spatial effect and the one due to other sources of heterogeneity. In particular, the two global components should account for the overall spatial configuration, not only for a single distance to define the neighbourhood, as currently proposed in the literature. Secondly, in order to allow as deep an investigation as wished, each global measure should be decomposable into partial terms, reflecting the information linked to different distance ranges. Besides, the role of space should be detected, irrespective of the spatial association being positive (clustering behaviour) or negative (repulsive behaviour). Furthermore, it should be applicable to variables with any number of categories. Lastly, it should allow straightforward interpretation in order to disseminate the results.

A set of spatial entropy measures is here considered, improving Altieri et al (2017). The approach consists in defining two new variables: Z, which transforms the information of X according to an idea of neighbourhood (by extending the proposal by O'Neill et al, 1988), and W, which accounts for the spatial configuration. In the present work, the variable Z is a categorical variable identifying 'co-occurrences' of X. This term first appears in Leibovici (2009): a co-occurrence is a set of realizations of X over the spatial domain, and is defined by fixing a degree of co-occurrence m, i.e. the cardinality of each set of co-occurrences. For simplicity, the resulting variable is denoted as Z instead of $Z^{(m)}$. In the simplest case, $m = 2$ and the categories of Z, denoted by z_r, identify pairs of categories of X, $\{x_i, x_{i'}\}$, with $i, i' = 1, \dots, I$. If $m = 3$, sets of three categories of X are considered and each category z_r corresponds to $\{x_i, x_{i'}, x_{i''}\}$, with $i, i', i'' = 1, \dots, I$; the same holds for further degrees m. The choice of m is exogenous, driven by the researcher's experience and by the purposes of each specific case study. The new variable Z has R_m categories, and its pmf is $p_Z = (p(z_1), \dots, p(z_{R_m}))'$, where $p(z_r)$, with $r = 1, \dots, R_m$, is the probability of observing the rth category of Z on any mth degree co-occurrence of X over the observation window. The pmf may be known, which requires knowledge of the pmf of X and also of its spatial structure at any distance; usually, the pmfs of both X and Z are estimated. In the current literature and in the present work, spatial entropy measures are estimated by substituting the unknown probabilities with the observed relative frequencies, obtaining the well known non parametric and also maximum likelihood entropy estimator (Paninski 2003).

A novelty of the proposed measures in Altieri et al (2017) lies in the introduction of a second discrete variable; properties of spatial entropy related to a bivariate distribution are highlighted, with a different perspective, in Leibovici and Birkin (2015). In the present work, the second variable W classifies the Euclidean distances within the observation window according to a set of distance classes, so that co-occurrences take place at different distance ranges. Intervals w_k, with $k = 1, \ldots, K$, cover all distances within the observation window: a set of distance breaks $d_0, \ldots, d_K$ is fixed, with $d_0 = 0$ and d_K being the maximum possible distance inside the window. Then, each class is $w_k =]d_{k-1}, d_k]$, and a mth degree co-occurrence at range w_k takes place if the distance between any two units of the co-occurrence set is included in the interval $]d_{k-1}, d_k]$. Co-occurrences where the minimum and the maximum distances between pairs of realizations fall in different w_k intervals cannot be assigned to a distance range and are not considered. The resulting co-occurrences are, from now on, named range-occurrences. The number K and the breaks d_k are fixed according to the context and can be modified as wished. When space is discrete, e.g. for lattice data, some geometrical restrictions should be added, in order to avoid w_k intervals with zero range-occurrences: for instance, if $m = 2$ the difference $d_k - d_{k-1}$ should be no smaller than the pixel width. In addition, increasing the value of m imposes restrictions on the minimum value allowed for the ratio $d_k/(d_k - 1)$, below which the class of range-occurrences falling within the kth distance range would be empty. This rarely happens in practical situations as, when working on a plane, usually $m = 2$ or $m = 3$ is chosen. As regards interpretation, based on the value of m one should choose sensible breaks d_ks for the distance ranges for avoiding empty classes; nevertheless, this is not a problem, since empty classes do not affect computations as they become 0 terms in the additive entropy formula and are discarded. The variable W has a pmf $p_W = (p(w_1), \ldots, p(w_K))'$, where $p(w_k)$ is the probability of a co-occurrence to fall within the kth distance range. Probabilities $p(w_k)$ depend on the degree of range-occurrence m; we follow the same choice as for Z and do not write $p_W^{(m)}$ for simplicity of notation. Such probabilities are estimated by relative frequencies, as for the pmfs of X and Z, i.e., by the frequencies of range-occurrences of degree m that lie within each distance range, irrespective of their category.

The introduction of distances combined with the use of the variable Z allows to define a neighbourhood via the construction of adjacency matrices, which, for a generic degree m, generalize to hypermatrices in the m-dimensional space. Indeed, once the degree m of co-occurrences is fixed, each distance category w_k induces the choice of a corresponding adjacency matrix (or hypermatrix) A_k that is from now on called range-adjacency matrix. Each w_k defines a ring around each areal unit. If $m = 2$, the matrix elements are $a_{uu',k} = 1$ if unit u' (its centroid, if data are lattice) falls within the ring defined by w_k and centered at unit u; they are $a_{uu',k} = 0$ otherwise. If the units are areas/pixels, the same holds for the area centroids. This allows to focus on range-occurrences, i.e. co-occurrences identified by non-zero elements of each A_k, a subset of Z conditional on a fixed distance range, denoted by $Z|w_k$. Since neighbourhood is a symmetric concept, each A_k is a symmetric matrix (or hypermatrix), therefore only one out-of-(hyper)diagonal (hyper)triangle of A_k may be considered. This way, K conditional pmfs are constructed $p_{Z|w_k} = (p(z_1|w_k), \ldots, p(z_{R_m}|w_k))'$. This conditioning is relevant to stress the logical relationship between the two random variables jointly considered: Z, pertaining to the variable under study, is influenced by W, which relates to space.

3.1 A set of spatial entropy measures

The introduction of two variables, Z and W, allows to exploit a well known relationship of the theory of entropy (Cover and Thomas 2006) which has not been considered in the traditional literature of spatial entropy measures so far:

$$H(Z) = MI(Z,W) + H(Z)_W. \tag{2}$$

In Information Theory, equation (2) states that the entropy of a variable may be split into the information brought by its relationship with another variable and the residual entropy due to other sources of heterogeneity. For spatial entropy measures and under the framework introduced above, relationship (2) represents a meaningful decomposition of Shannon's entropy of the variable Z; this shows the additional value of using Z instead of X. Spatial mutual information $MI(Z,W)$ is a Kullback-Leibler divergence which represents the component of the entropy of Z due to its relationship with the spatial configuration. Spatial global residual entropy $H(Z)_W$ measures the remaining information brought by Z.

Generalizing Altieri et al (2017), spatial mutual information in (2) is defined as:

$$MI(Z,W) = \sum_{k=1}^{K} p(w_k) \sum_{r=1}^{R_m} p(z_r|w_k) \log\left(\frac{p(z_r|w_k)}{p(z_r)}\right). \tag{3}$$

This formulation rewrites the usual definition of mutual information (Cover and Thomas 2006), applied to the case of Z and W, that highlights the relationship direction between Z and W and the decomposition of the overall role of space with respect to the contribution of every distance range w_k. Each kth internal sum is called spatial partial information, where 'partial' corresponds to a specific distance class w_k:

$$PI(Z|w_k) = \sum_{r=1}^{R_m} p(z_r|w_k) \log\left(\frac{p(z_r|w_k)}{p(z_r)}\right). \tag{4}$$

Each partial term is a Kullback-Leibler divergence quantifying the contribution to the departure from independence at each distance class w_k.

The partial-to-global relationship is respected once the PIs are weighted by the probabilities $p(w_k)$, thus satisfying the desirable property of additivity:

$$MI(Z,W) = \sum_{k=1}^{K} p(w_k) PI(Z|w_k). \tag{5}$$

Relationship (5) guarantees that the choice of the distances d_k to build the classes w_k does not influence the global value of spatial mutual information: the global result can be obtained by any split or aggregation of classes, according to the investigation that needs to be carried over the partial terms.

Spatial global residual entropy $H(Z)_W$ in (2) can be defined, following the traditional formulation for residual (or conditional) entropy (Cover and Thomas 2006), as:

$$H(Z)_W = E[H(Z|W)] = \sum_{k=1}^{K} p(w_k) \sum_{r=1}^{R_m} p(z_r|w_k) \log\left(\frac{1}{p(z_r|w_k)}\right). \tag{6}$$

The components of (6)

$$H(Z|w_k) = E[I\left(p_{Z|w_k}\right)] = \sum_{r=1}^{R_m} p(z_r|w_k) \log\left(\frac{1}{p(z_r|w_k)}\right) \tag{7}$$

are named spatial partial residual entropies. When these measures are multiplied by the probabilities $p(w_k)$, they allow spatial global residual entropy (6) to be rewritten in additive form, analogously to (5), as:

$$H(Z)_W = \sum_{k=1}^{K} p(w_k) H(Z|w_k). \tag{8}$$

Spatial partial residual entropies (7) show how each distance range contributes to the residual entropy of Z. Once more, the choice of the w_k does not affect the value of the global measure.

By incorporating relationships (5) and (8) into equation (2), a general decomposition of $H(Z)$ is obtained:

$$H(Z) = \sum_{k=1}^{K} p(w_k) \left[PI(Z|w_k) + H(Z|w_k)\right], \tag{9}$$

where the contribution of each partial term in explaining the relationship between Z and W is isolated. This way, Shannon's entropy $H(Z)$ is written in additive form, where each term can be explored to check what categories of Z and W are farther away from independence.

The methodology for spatial entropy measures presented in this Section holds irrespective of the choice of K and of the distance breaks $d_1, \dots, d_{K-1}$. Indeed, the global values of the two components of spatial entropy, i.e. spatial mutual information and spatial residual entropy, are unaffected by such choice. The distance breaks are instead crucial in defining the partial terms, that determine the detail and depth of the investigation.

To better disseminate the results, the role of space can be quantified in proportional terms:

$$MI_{prop}(Z, W) = \frac{MI(Z, W)}{H(Z)}. \tag{10}$$

This quantity ranges in $[0, 1]$ and states the contribution of space in the entropy of Z as a proportion of the marginal entropy. Shannon's entropy of Z has to be considered a reference value for both terms. This way, datasets with the same pmf p_Z but different spatial configurations share the same $H(Z)$ with a different contribution of its two components.

3.2 Advances with regard to standard spatial entropy measures

3.2.1 Discarding order within co-occurrences

The mth degree occurrence of Z may be built either following a direction in space, usually rightward and downward, or by double-counting, i.e. counting sets of occurrences moving along all spatial directions; this choice is discussed by Riitters et al (1996). Additionally, an assumption on whether to preserve the order within co-occurrences is needed when defining Z. Ordering occurrences means considering couples, triples and so on, while discarding the order means considering pairs, sets of three and so on. Thus, order preservation regards the importance of the relative spatial location of the observations, irrespective of their distance. For example, when $m = 3$, if order is preserved the triple $(x_i, x_{i'}, x_{i''})$ implies that the observation carrying the i'th category occurs at the right or below the observation carrying the ith category, and that the one presenting the i''th category is right and below the others. Under this criterion, the triple is different from $(x_{i'}, x_{i''}, x_i)$, while the unordered set of three $\{x_i, x_{i'}, x_{i''}\}$ includes both cases. Consequently, when order is discarded, the number of categories of Z is $R_m = \binom{I+m-1}{m}$; when it is preserved, the number increases to $R^o_m = I^m$, because sets of co-occurrences containing the same categories of X in a different order are counted separately in the entropy of Z. As a further example, if $m = 2$ and X is binary, the possible pairs are $\{x_1, x_1\}$, $\{x_1, x_2\}$ and $\{x_2, x_2\}$, while couples are (x_1, x_1), (x_1, x_2), (x_2, x_1) and (x_2, x_2). In the approach of the present paper, order is discarded: if $m = 2$ pairs are considered, then $R_2 = 3$, as only three terms $p_{11} = p(\{x_1, x_1\})$, $p_{12} = p(\{x_1, x_2\})$ and $p_{22} = p(\{x_2, x_2\})$ enter the computation of Shannon's entropy of Z. The standard spatial entropy measures (O'Neill et al 1988; Li and Reynolds 1993; Leibovici 2009; Leibovici et al 2014; Parresol and Edwards 2014), instead, compute such entropy with $R^o_2 = 4$ terms: p_{11}, p_{12}, p_{21} and p_{22}. Because of the double counting method, p_{12} equals p_{21}, therefore there is very little discussion about the implications of entering the two probabilities separately; only Riitters et al (1996) points this out. Actually, the standard choice of separating quantities of type p_{ij} and p_{ji} (and analogously for further degrees of co-occurrences) in computations implies order preservation within co-occurrences.

Ordering occurrences has major consequences that have not been considered over the spatial entropy literature. Firstly, considering the order does not appear sensible in spatial statistics, since spatial configurations are not generally assumed to have a direction. The purpose of a spatial entropy measure is to understand whether any kind of spatial association contributes to decrease the entropy of a variable, irrespective of the direction. Secondly, when order is discarded, the number of categories of Z is smaller. The gap between the two options grows as I increases, and results in a substantially different computational burden for large datasets. Thirdly, since order preservation increases the number of categories of Z, the resulting entropy also increases and leads to erroneous conclusions when there is no substantive motivation for separating quantities of type p_{ij} and p_{ji}. Lastly, and above all, discarding the order ensures a one-to-one correspondence between Shannon's entropy of X and Z. Indeed, if order is considered, a change in the spatial configuration, i.e. a permutation of the realizations of X over space, results in a change in the ordered co-occurrences, yielding a different entropy $H(Z)$, while $H(X)$ is the same. In the above example, if occurrences are spatially permuted, the number of couples of type (x_1, x_2) and (x_2, x_1) changes, thus p_{12} and p_{21} change too and the entropy $H(Z)$ is different in the two cases. This is a drawback, since Z is a transformed variable used as a tool to explain the spatial entropy of X: $H(Z)$ should be a stable reference value, while its two components $MI(Z, W)$ and $H(Z)_W$ should vary in order to evaluate the role of space. This is only the case when order is discarded: the number of pairs $\{x_1, x_2\}$ is the same irrespective of the spatial location of occurrences, thus p_{12} does not change and $H(Z)$ is unique.

The present paper also recommends to discard the traditional double counting rule. Double counting, indeed, increases the computational burden substantially. In addition, even when order is preserved, conclusions may be incorrect: due to the equality of p_{ij} and p_{ji} (and analogously for further order of occurrences), the entropy value is higher than the one obtained moving right- and downward; see the discussion in Riitters et al (1996).

All the above reasons encourage the choice of considering unordered occurrences as the most appropriate, combined with the approach of moving downward and rightward along the observation window. This choice removes any consideration of possible anisotropic directional effects, in agreement with the symmetrical foundations of basic entropy measures. Further details on how to build Z are given in Section 3.3. More disadvantages of preserving order are highlighted in the comparative study of Section 4.

3.2.2 A theoretical link to popular spatial entropy measures

Spatial global residual entropy (6) represents a generalization of spatial entropies available in the literature. O'Neill's entropy (1988) may be derived as a special spatial partial residual entropy (7): for lattice data, one can fix the cell width at 1; by setting $m = 2$ and focusing only on the distance class $[0, 1]$, denoted by $w_{(01)}$, contiguous couples for the subset $Z|w_{(01)}$ are considered (i.e. pixels that are adjacent in the dataset, not only in the neighbourhood system graph). O'Neill's entropy may indeed be written as:

$$H(Z|w_{(01)}) = E\left[I\left(p_{Z|w_{(01)}}\right)\right] = \sum_{r=1}^{R_2^o} p(z_r|w_{(01)}) \log\left(\frac{1}{p(z_r|w_{(01)})}\right). \tag{11}$$

Expression (11) rewrites equation (2) in O'Neill et al (1988), and takes into account the corrections to the index discussed in Li and Reynolds (1993) and the considerations of Section 3.2.1. It sums over $R_2^o = I^2$ categories, as the authors preserve the order within couples; the distance range $w_{(01)}$ corresponds to the range-adjacency matrix $A_{(01)}$, namely the contiguity matrix.

Moreover, for a generic degree m and a particular distance range $w_{(0d)} = [0, d]$, Leibovici's entropy (2009) is obtained:

$$H(Z|w_{(0d)}) = E\left[I\left(p_{Z|w_{(0d)}}\right)\right] = \sum_{r=1}^{R_m^o} p(z_r|w_{(0d)}) \log\left(\frac{1}{p(z_r|w_{(0d)})}\right). \tag{12}$$

The distance range $w_{(0d)}$ corresponds to the range-adjacency matrix/hypermatrix $A_{(0d)}$.

Other measures proposed in the literature (Li and Reynolds 1993; Riitters et al 1996; Parresol and Edwards 2014), being deterministic functions of (11), can also be derived from (7). Therefore, all comments also hold for those indices. The limitation of these measures is that they only provide a partial result. Indeed, O'Neill's entropy only uses information about adjacent couples, and ignores the rest. Leibovici's entropy works on the same principle, extending to a general d. Thus, if d is small, a great part of the spatial information is not considered; conversely, if d is large, the result is aggregate and excludes any possibility to explore the contribution of space in detail. Measure (6) allows to exploit all the spatial information and, at the same time, to disaggregate results as wished. More limitations of the standard spatial entropy indices in comparison to the range-occurrence approach are shown in Section 4.3 and 5.2.

3.3 Range-occurrences for the variable Z

A procedure to construct range-occurrences for obtaining Z is needed and proposed in what follows. Let us first consider the case of pairs, i.e. $m = 2$. In this case, Z (unordered) has $R_2 = \binom{I+1}{2}$ categories, simply denoted by R.

When spatial units are considered, which may be areas or points, the I categories of X occur over N realizations x_u, $u = 1, \ldots, N$. Analogously, the R categories of Z occur over a number $Q >> N$ of spatial realizations z_q, $q = 1, \ldots, Q$. In particular, $Q = \sum_k Q_k$, where Q_k is the number of range-occurrences at w_k, i.e. realizations of Z at the distance category w_k.

For each distance category w_k, range-occurrences are built according to the specific $N \times N$ range-adjacency matrix A_k. The cardinality of the neighbourhood $\mathcal{N}(u)_k$ of unit u, $|\mathcal{N}(u)_k| = \sum_{u'=1}^{N} a_{uu',k}$, is the number of u' spatial units belonging to $\mathcal{N}(u)_k$, where $a_{uu',k} = 1$ if units u and u' lie within distance range w_k and 0 otherwise. The number of observable pairs is $Q_k = \sum_{u=1}^{N} |\mathcal{N}(u)_k| = \sum_{u=1}^{N} \sum_{u'=1}^{N} a_{uu',k}$. From this follows that Q_k depends on the number N of realizations of X and the number of neighbours. Moreover, Q_k is different according to the counting method. If the counting of pairs is rightward and downward, only one out-of-diagonal triangle of the matrix A_k is used, while the rest receives value 0. If the so-called double counting method is followed, both out-of-diagonal triangles are used. Note that the two counting approaches result in a different number Q_k, but do not turn pairs into couples, as Q_k only counts the number of range-occurrences and does not consider their category (see Riitters et al, 1996 and Section 3.2.1 for details).

In order to construct the range-occurrences for Z, for each w_k a $Q_k \times 2$ matrix is built (for $m = 2$), each row containing the values $\{x_u, x_{u'}\}$ identified by the corresponding A_k. The matrix is composed of N blocks, one for each spatial unit u; the uth block has $|\mathcal{N}(u)_k|$ rows, depending on the spatial location of each element with respect to other units. The first column of each block replicates the spatial unit value x_u as many times as the cardinality of its neighbourhood $|\mathcal{N}(u)_k|$: $[x_u \cdot 1_{|\mathcal{N}(u)_k|}]$.

The second column of the uth block contains the neighbouring values $x_{u'}$, $u' \in \mathcal{N}(u)_k$, selected via A_k, and is constructed as follows. Let us define $vec(X)$ as a $N \times 1$ vector stacking all realizations of X, and the $N \times N$ selection matrix $\widetilde{A}_k$, which considers zeros in A_k as missing values. An element-wise product between $vec(X)$ and the uth row of $\widetilde{A}_k$ allows to isolate a $|\mathcal{N}(u)_k|$-dimensional vector, containing the $x_{u'}$ values, $u' \in \mathcal{N}(u)_k$ (potentially presenting any of the I categories), and discarding all other realizations: $[vec(X) \cdot \widetilde{A}_{u.,k}]$.

Finally, range-occurrences for Z at each distance range w_k are obtained by stacking the N blocks of type $\left[x_u \cdot 1_{|\mathcal{N}(u)_k|}, vec(X) \cdot \widetilde{A}_{u.,k}\right]$, with $u = 1, \ldots, N$. As a result the $Q_k \times 2$ matrix is

$$\begin{bmatrix} x_1 \cdot 1_{|\mathcal{N}(1)_k|} & vec(X) \cdot \widetilde{A}_{1.,k} \\ \vdots & \vdots \\ x_N \cdot 1_{|\mathcal{N}(N)_k|} & vec(X) \cdot \widetilde{A}_{N.,k} \end{bmatrix}. \tag{13}$$

For each distance w_k, a matrix (13) is generated (K matrices in total): each row of the matrix is a range-occurrence of Z at distance w_k. The realizations present at most R categories.

When $m > 2$, the number Q of realizations of Z increases accordingly. Indeed, the distance-specific matrices A_k generalize to hypermatrices in the m-dimensional space and, thus, the cardinality of the neighbourhood grows. The matrix containing the range-occurrences of Z at each distance w_k is now $Q_k \times m$, where each row identifies the category of the spatial unit of interest and of its neighbours, $\{x_u, x_{u'}, x_{u''}, \ldots\}$.

4 A comparative study of spatial entropy measures

The set of spatial entropy measures illustrated in Section 3 needs to be further investigated in order to stress their basic properties and highlight the different contexts of application. In this Section, the behaviour of the range-occurrence approach is assessed in terms of flexibility and informativity with respect to other proposals. Several datasets are generated under different scenarios to compute spatial mutual information (3) and spatial global residual entropy (6), as well as the partial terms. A comparison of this set of entropy measures is run with respect to O'Neill's entropy (11) and its generalization Leibovici's entropy (12).

For simplicity of presentation, the method is assessed by building the variable Z to represent pairs of occurrences of X over space, i.e. by fixing $m = 2$. Discrete space and a regular grid are considered; space can be discretized as wished, as long as a distance measure between areas is suitably defined within the observation window. Additionally, areas may be replaced by points; in this case, W represents the distance between points themselves, and entropy measures are defined accordingly (for a case study with point data, see Altieri et al, 2017).

4.1 Data generation

Let us consider $N = 2500$ realizations of a binary variable X by randomly setting the pmf p_X and then generating values x_1 and x_2 from a Bernoulli distribution. Let us introduce space by considering a square window gridded by 50×50 pixels; without loss of generality, each pixel is assumed to be a 1×1 square, therefore the observation window is 50×50 units.

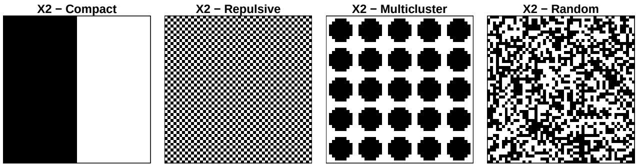

Fig. 2 Data generated for all simulation scenarios under a uniform distribution of X.

After the random generation of X outcomes, the same simulated sequence of 2500 values is organized over the window according to different spatial configurations, an example of which is presented in Figure 2, where x_1 values are represented as black pixels and x_2 values as white pixels. Configurations are chosen as they are expected to produce different entropy values:

1. compact - the most clustered spatial distribution, obtained by assigning x_1 values to the pixels located at the left part of the window and x_2 values to pixels located at the right part;
2. repulsive - the most regular spatial distribution, tending to a chessboard, obtained by assigning x_1 values to pixels adjacent to x_2-valued pixels, and vice versa. Note that a perfect chessboard can only be obtained when the number of x_1 and x_2 outcomes is the same;
3. multicluster - 25 clusters, whose centroids are generated from a Poisson point process; then, x_1 values are assigned to pixels surrounding the centroids and x_2 values to the remaining pixels. The size of the clusters is random, as it depends on the total number of generated x_1 oucomes;
4. random - a pattern with no spatial correlation whatsoever, obtained by assigning x_1 or x_2 values to pixels via simple random sampling without replacement.

Each simulated scenario is replicated 1000 times, for 1000 generations of p_X, and assigned to the pixels according to the four configurations A further dataset generated under the hypothesis of having the same number of x_1 and x_2 outcomes, yielding the maximum entropy of X, is built as a special case for each scenario, following the same criteria, and is the one displayed in Figure 2. In this special case, the 25 cluster centroids (third panel) are also forced to be located on the nodes of a regular grid over the square window.

If one focuses on the strength of spatial association and not on its type (positive or negative), a decreasing contribution of space is yielded over the four configurations. Therefore, a good measure should detect, for the compact pattern, a high level of spatial mutual information and a low spatial residual entropy. Spatial mutual information should gradually decrease across the other scenarios, reaching the lowest value for the random configuration.

Range-adjacency matrices A_k are built on pixels of size 1 and based on distances between pixel centroids; thus, the distance between contiguous pixels is 1 and the distance to farther cells along the cardinal directions belongs to the set of integers $\mathbb{Z}^+$. For the partial terms of spatial mutual information and residual entropy, W is built with categories $w_1 = [0, 1]$, $w_2 =]1, 2]$, $w_3 =]2, 5]$, $w_4 =]5, 10]$, $w_5 =]10, 20]$, $w_6 =]20, 30]$ and $w_7 =]30, 50\sqrt{2}]$ (where $50\sqrt{2}$ is the maximum distance over the observation window, a square of side 50), covering all possible distances for pairs over the dataset. This choice for the classes is motivated by the tradition of spatial statistics (Cressie 1993): class w_1 corresponds to what is

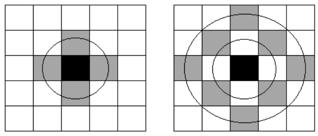

Fig. 3 Neighbourhood system for distance classes w_1 (left) and w_2 (right)

known as the 4-nearest neighbour system, while class w_2 considers the farther 8-nearest neighbours, and together they form the well-known 12-nearest neighbour system. These are the standard distance ranges when studying spatial association. In addition, at w_1 and w_2 we expect to appreciate the difference among the four scenarios. As distance increases, spatial association usually decreases, which is why the following distance ranges are gradually wider and less informative. Figure 3 illustrates the two standard neighbourhood systems for one generic pixel u in the case $m = 2$. In the left panel, the circle marking the end of the distance class has radius 1 and identifies the 4 contiguous pixels as neighbours; this forms 4 pairs of type $\{x_u, x_{u'}\}$ with u' taking 4 values. In the right panel, two circles are used as distance breaks and the resulting ring includes the farther 8 nearest pixels: 8 pairs of type $\{x_u, x_{u'}\}$ with varying u' can be built. The variable W covers all possible distances within the observation window; conclusions in this comparative study are able to highlight the different influence of space at different distance breaks. Due to the additivity property, the general validity of the proposal of the present paper holds irrespective of the chosen distance breaks.

The probabilities $p(x_i)$ for each category of X are estimated by the proportion of spatial units where x_i is observed: $\widehat{p}(x_i) = \sum_{u=1}^{N} \mathbf{1}(x_u = x_i)/N$, with $\mathbf{1}$ defining the indicator function.

The Q_k realizations of Z at distance range w_k are built by following the procedure proposed in Section 3.3 and counting right- and downward. Their relative frequencies are used to compute $\widehat{p}(z_r|w_k)$, where $r = 1, 2, 3$ given that $I = 2$ and order is discarded. Since for each distance w_k a specific range-adjacency matrix A_k is built, $K = 7$ different A_k and conditional distributions $\widehat{p}_{Z|w_k}$ are obtained. The marginal p_Z may be estimated after marginalizing out W. An estimate for p_W is also needed: for each k, $\widehat{p}(w_k) = Q_k/Q$ represents the proportion of pairs within distance range w_k.

For the computation of O'Neill's spatial entropy (11), the employed range-adjacency matrix is $A_{(01)}$; for Leibovici's entropy (12), the range-adjacency matrix $A_{(0d)}$ is used and $d = 2$ is chosen. Order is preserved for both measures.

All indices are computed for each scenario over the 1000 generated datasets, plus the special case of uniform distribution among the X categories. Results are presented via boxplots, that summarize the distribution of each index; stars highlight results achieved under the uniform distribution of X, while the dashed lines, where present, mark the indices' maxima.

4.2 Results

Shannon's entropy (1) is firstly computed for X and Z without order preservation, so that $H(Z)$ does not depend on the spatial configuration and is unique (see

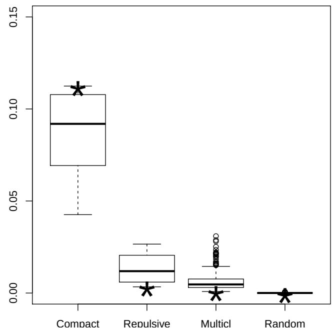

Fig. 4 Proportional spatial mutual information, 1000 simulations. Stars identify the entropy value computed on a uniformly distributed X.

Section 3.2.1). The distribution of the differences between the normalized versions of $H(X)$ and $H(Z)$ across 1000 replicates shows that Z and X may be considered as interchangeable when computing entropy: the differences range from -0.009 to 0.05, with a mean very close to 0. This highlights that Z brings the same information as X, and encourages the use of Z, without order preservation, as a starting point for explaining the spatial behaviour of realizations of X. Entropy $H(Z)$ ranges from 0.52 to 1.04 for the 1000 replicates, and is constant across spatial configurations. The two components $MI(Z,W)$ and $H(Z)_W$ vary according to the different spatial patterns; thus, the proportion of entropy due to space, i.e. the proportional version of mutual information (10), takes different values across scenarios, as shown in what follows.

Proportional spatial mutual information is displayed in Figure 4. The index effectively detects the decreasing role of space along the four spatial configurations. Focusing on the median value, in the compact pattern nearly 10% of the entropy of Z is due to the data spatial configuration (first boxplot in Figure 4); this implies that the remaining 90% of $H(Z)$ is due to residual entropy, i.e. heterogeneity due to other sources. Conversely, no influence of space emerges over the random pattern, where space does not help in explaining the data heterogeneity: $H(Z)$ and $H(Z)_W$ coincide. As expected, repulsive and multicluster configurations mirror intermediate situations. At the global level, the detected influence of space is often low and may be unsatisfactory; for this reason, spatial mutual information (3) is then disaggregated into spatial partial information terms (4) by fixing the different distance categories w_1 to w_7 introduced above.

Results for the partial terms are shown in Figure 5 (results for w_7, not reported here, are very similar to those for w_6) and highlight at what distances the spatial configuration contributes substantially to the entropy of Z. For the random pattern the role of space is detected as null at any distance range. For the compact and multicluster configurations, spatial partial information values tend to decrease as distance increases. The decline is slower for the compact pattern, where the contribution of space to entropy is appreciable up to distance w_5. For the repulsive pattern, spatial partial information takes high values for the first two distance ranges, where space plays a role, and is particularly appreciable at distance w_2 because it captures a high number of range-occurrences of type (x_1, x_1) and (x_2, x_2). Then, information drops from distance w_3 on: the abrupt decrease

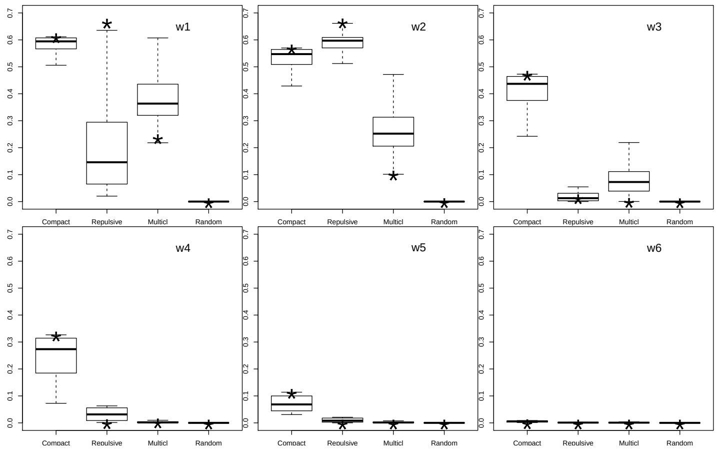

Fig. 5 Spatial partial information, 1000 simulations. Each star identifies the entropy value computed on a uniformly distributed X.

explains why the proportional global values (Figure 4, second boxplot) are sensibly lower than those coming from the compact configuration, though still different from zero.

Spatial partial residual entropies (7) at distances w_1 to w_6 (results for w_7 are not reported for the same reasons mentioned above) are summarized in Figure 6. The panels referring to short distances (i.e., w_1 and w_2) are the most relevant: when compared with other patterns, compact and repulsive configurations have lower levels of residual entropy, given the stronger contribution of space detected by the partial information terms.

Under the special case of having 50% of the outcomes of the variable X of type x_1 and 50% of type x_2, consequences can be observed on the decomposition of $H(Z)$ (stars in Figures 4, 5 and 6). The most interesting aspect concerns the repulsive pattern, where a uniform distribution corresponds to a perfect chessboard configuration. In such case, at distance w_1 all range-occurrences belong to the same category of Z, yielding the maximum spatial partial information value (second boxplot of first panel in Figure 5) and a null residual entropy (second boxplot of first panel in Figure 6). The recognition of such strong role of space supports the desirable features of the proposed set of measures, which are not possessed by standard indices (see the results of Section 4.3).

Figure 7 presents interpretable and comparable results: the median values at each distance class over the 1000 simulations are computed, then the sum $PI(Z|w_k) + H(Z|w_k)$ of (2) is set to 1 at each w_k. This enables to appreciate the relative contribution of both terms for each scenario. The detection of a different role of space among scenarios is evident up to distance w_5. In the compact configuration, the two terms are almost even at short distance, then partial information decreases slowly and remains present up to the last distance class. The repulsive scenario shows peaks at short distances, where a chessboard-type configuration may be detected, while from distance w_3 on the contribution of space

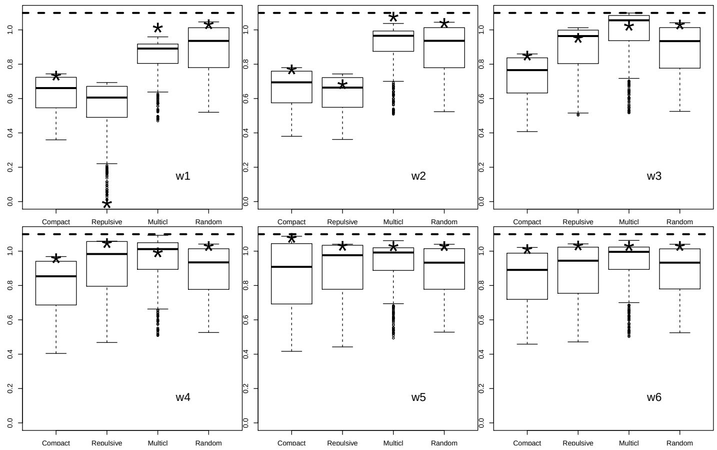

Fig. 6 Spatial partial residual entropies, 1000 simulations. Each dashed line corresponds to the index maximum; each star identifies the entropy value on a uniformly distributed X.

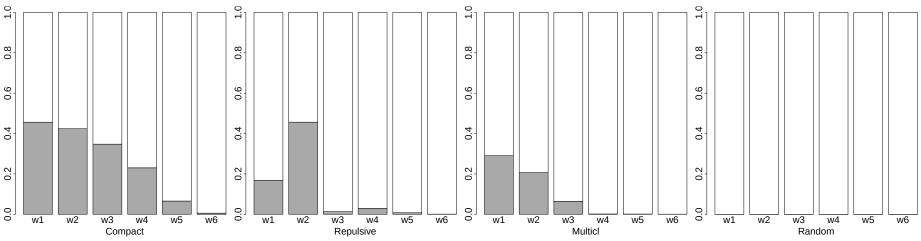

Fig. 7 Partial information (gray areas) and partial residual entropies (white areas) in proportional terms.

becomes negligible. The multicluster pattern can be seen as a reduced version of a compact one, and partial information behaves accordingly. Lastly, there is no spatial information in the random dataset, irrespective of the distance class.

4.3 A comparison to standard spatial entropy measures based on Z

Results for O'Neill's spatial entropy (11) and Leibovici's spatial entropy (12) at distance $d = 2$ are displayed in Figure 8. For the compact and the multicluster data, O'Neill's and Leibovici's indices tend to behave the same way and return the same amount of information. This states that the choice of d in Leibovici's entropy barely influences the entropy values and is not useful in discriminating, as long as d is smaller than the cluster size.

O'Neill's index (Figure 8, left panel) studies the spatial entropy at a distance equal to the pixel size, as does the first partial term of spatial mutual information (Figure 6, first panel), but preserves the order in building couples. Leibovici's entropy values (Figure 8, right panel) are aggregations of the values in the first and

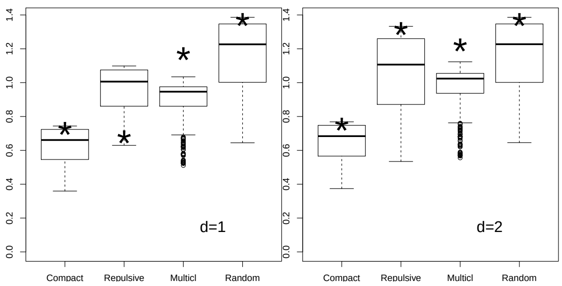

Fig. 8 O'Neill's entropy (left) and Leibovici's entropy with $d = 2$ (right), 1000 simulations. Each star identifies the entropy value computed on a uniformly distributed X.

second panels of Figure 6, with order preservation. The difference in the results, when compared to those of Section 4.2, highlights the consequences of considering partial terms and discarding the order. Partial residual entropies (7) consider different distance ranges separately, while Leibovici's entropy counts all couples within a fixed distance d without distinction nor possibility of further inspection. As a result, the measures of Figure 8 are only able to give partial knowledge; moreover, due to order preservation, their Shannon's entropy $H(Z)$ is not unique and cannot be used as a benchmark (see Section 3.2.1). This implies that a proportional contribution of the entropy due to the spatial structure cannot be quantified here, which is instead done for the decomposable set of spatial entropy measures in Figure 7.

A further major limitation of considering order within couples in (11) and (12) is that for the repulsive pattern (Figure 8, second boxplot of both panels) much greater entropy values are returned than for the compact configuration (first boxplots). For Leibovici's entropy, values for the repulsive pattern are even higher than in O'Neill's entropy, and have a distribution which is very similar to that of the random patterns. In the special case of uniform distribution for X, the entropy value for the repulsive pattern (star of the second boxplot in Figure 8, left panel) cannot reach the minimum value 0 due to order preservation. Conversely, a well performing spatial entropy measure based on co-occurrences should account for the presence of a spatial pattern without distinguishing a negative correlation from a positive one, because the very definition of entropy is based on the idea of heterogeneity and surprise, and is different from the definition of spatial association, that focuses on the type of the relationship. For the spatial measures considered in this work, heterogeneity concerns range-occurrences, not single realizations of X. Therefore, results for the compact and repulsive patterns should be more similar than they appear in Figure 8. In particular, for the uniform dataset with a repulsive pattern the lower limit 0 should be reached, since all pairs are equal: having all realizations of type $\{x_1, x_2\}$ means zero surprise and maximum homogeneity in observing pairs. It also means maximum (negative) spatial association for the variable X. This does not occur in Figure 8. Such desirable feature is met by spatial partial residual entropies (Figure 6, first and second panel), and constitutes an additional reason for discarding order in building Z.

4.4 Extension to data with more than two categories

Data are also generated for $I = 5$ and $I = 20$ categories. The small gap between $H(X)$ and $H(Z)$ found for $I = 2$ becomes even more negligible as I increases.

Only the compact and the random patterns can be distinguished and compared here. Most results are analogous to those reported for the binary case: the contribution of space is correctly quantified in both scenarios by spatial mutual information and its partial terms, while spatial residual entropy and its partial terms suitably measure the heterogeneity not due to space at different distances. It is worth higlighting that, when switching from $I = 2$ to $I = 5$ and $I = 20$, all entropy values increase while the variability across replicates decreases both for the compact and the random configuration. The divergence between the two spatial configurations due to the reduced variability is another desirable feature possessed by the range-occurrence approach.

The distributions of the proportional version (10) of spatial mutual information are centered around similar values for $I = 2$, $I = 5$ and $I = 20$, highlighting that the role of space is detected as constant across different numbers of categories. This similarity is a key advantage for interpreting proportional spatial mutual information, since it is comparable across different variables.

5 Application to land use raster data

After assessing its properties in Section 4, the set of entropy measures proposed in Section 3 is employed in a real case. This Section is a toolbox for evaluating the entropy of spatial data, which shows how to tune the choice of the distance classes.

The application deals with a subset of European CORINE data. CORINE, (COoRdination de l'INformation sur l'Environnement, i.e. Coordination of Information on the Environment) is a programme approved by the European Community Council in 1985 with the aim of gathering, coordinating and ensuring the consistency of information on the state of the environment and natural resources in the Community. CORINE Land Cover (CLC) is a project within the CORINE programme specifically created for monitoring land characteristics. It works on map imageries coming from satellite pictures, which are successively photo-interpreted and publicly supplied as data matrices. The European territory is divided into units which are classified according to CLC nomenclature in forty-four classes of land use. The last extensive dataset update was run by EEA (European Environmental Agency) in 2012, developed within GMES programme (Global Monitoring for Environment and Security). The resulting Land Cover datasets are freely available for download at http://eea.europa.eu/.

5.1 Results for land use data

The toolbox develops in three main steps. As a start, Shannon's entropy of X is computed. The proportion of pixels, which estimate $p(x_i)$, for each of the 4 land cover categories are in Table 1. The resulting entropy is $H(X) = 0.957$, with a maximum of $\log(I) = \log(4) = 1.39$. The normalized entropy is $H(X)_{norm} = 0.69$.

Table 1 Proportions of land cover categories (variable X).

	urban	*agricultural*	*forests*	*water*
x_i	x_1	x_2	x_3	x_4
$\widehat{p}(x_i)$	0.055	0.533	0.385	0.027

Table 2 Proportions of pairs of land cover categories (variable Z).

z_r	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
$\{i,i'\}$	$\{1,1\}$	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,2\}$	$\{2,3\}$	$\{2,4\}$	$\{3,3\}$	$\{3,4\}$	$\{4,4\}$
$\widehat{p}(z_r)$	0.003	0.058	0.042	0.003	0.286	0.411	0.029	0.147	0.021	0.001

Pairs ($m = 2$) of categories of X are then considered to build the variable Z. The Z categories and the corresponding proportions, estimates of $p(z_r)$, are shown in Table 2. In this case, $H(Z)$ is 1.524. The normalized entropy, divided by its maximum $\log(R_2) = \log(10) = 2.3$, is $H(Z)_{norm} = 0.662$. Following the results of the comparative study of Section 4, the normalized entropies of X and Z are similar, which further supports the choice of using Z in order to gain information about X. The normalized values are not very high, due to the departure of the estimated marginals $\widehat{p}_X$ and $\widehat{p}_Z$ from the uniform distribution, as can be seen in Tables 1 and 2. Entropy $H(Z)$ does not depend on the spatial configuration, but is fundamental in the spatial entropy perspective since it is the reference value for both spatial residual entropy and spatial mutual information.

The second step of the proposed spatial entropy toolbox consists in the computation of spatial mutual information (3) and spatial global residual entropy (6), which sum to $H(Z)$. In this application, at the global level the spatial residual entropy nearly coincides with its reference value $H(Z)$; the spatial mutual information is only $MI(Z,W) = 0.01$, and its proportional version (10) is $MI_{prop}(Z,W) = 0.007$. This would give the first erroneous idea that the entropy of the variable 'Land use' over the Emilia Romagna Region is entirely due to sources of heterogeneity other than space, but this is a global result, which does not investigate distance ranges in detail. It is, on the contrary, very important to define distance classes w_k and to compute the partial terms in order to deepen the spatial data heterogeneity understanding.

The third step consists in choosing suitable distance classes for the partial components, namely spatial partial information entropy terms (4) and spatial partial residual entropies (7). Distance is measured between centroids, and breaks are chosen, for small distances, based on the pixel size as follows: $d_0 = 0$, $d_1 = 1$ and $d_2 = 2$; the resulting neighbourhood systems for identifying range-occurrences at distances w_1 and w_2 are of common use in spatial statistics and have been shown in Figure 3.

As a first option, $K = 6$ distance classes are chosen, and for classes w_3 to w_6 the distance range after d_2 is cut into four equal parts. Table 3 shows the resulting distance classes and the corresponding partial terms, together with the distance weights $\widehat{p}(w_k)$. The results support the importance to focus on the partial terms to achieve the correct conclusions, without being misled by the global result alone. Spatial partial information terms (4) in Table 3, second line, allow to understand in detail how much space affects the entropy of land use data at each distance

Table 3 First option - distance classes, partial information and partial residual entropies.

k	1	2	3	4	5	6
w_k	[0; 1]	]1; 2]	]2; 66]	]66; 130]	]130; 194]	]194; 258]
$\widehat{p}(w_k)$	0.0003	0.0005	0.4760	0.3833	0.1332	0.0067
$PI(Z\|w_k)$	0.338	0.305	0.009	0.005	0.025	0.024
$H(Z\|w_k)$	1.421	1.438	1.534	1.496	1.495	1.501

Table 4 Second option - distance classes, partial information and partial residual entropies.

k	1	2	3	4	5	6	7
w_k	[0; 1]	]1; 2]	]2; 5]	]5; 15]	]15; 66]	]66; 130]	]130; 258]
$\widehat{p}(w_k)$	0.0003	0.0005	0.0045	0.0376	0.4339	0.3833	0.1399
$PI(Z\|w_k)$	0.338	0.305	0.260	0.172	0.005	0.005	0.028
$H(Z\|w_k)$	1.421	1.438	1.458	1.496	1.526	1.496	1.497

range w_k. The first two partial terms $PI(Z|w_1)$ and $PI(Z|w_2)$ are the greatest, confirming the well known law of geography which says that space plays a more relevant role at short distances. As Table 3 shows, however, they are weighted by the two very small values $\widehat{p}(w_1)$ and $\widehat{p}(w_2)$, since the range they cover in the observation window is much smaller than the range covered by intervals w_3 to w_6. This explains why the global spatial mutual information is so little affected by the relevant first two partial terms. The partial information terms at distance intervals w_3 to w_6 cover wide distance ranges, due to the observation window size, and are low, detecting very little influence of space on such wide ranges, away from small distances. The weights for w_3 and w_4 are the highest, while those for w_5 and w_6 are low because at great distances many couples are discarded as they lie outside the Emilia Romagna boundaries.

Spatial partial residual entropies (7) in Table 3, third line, receive a straightforward interpretation: they express the residual amount of entropy of Z (and consequently of X) after taking space into account. Complementarily to the partial information terms, the partial residual entropies are lower for distance ranges w_1 and w_2 and higher for the other four classes. A peak is present in the spatial residual entropy at w_3, which is the distance class with the highest weight; moreover, the spatial partial information for this class is higher than the one for class w_4, as Table 3 shows. Besides, there is an abrupt decrease in partial information from w_2 to w_3. This suggests that class w_3 may be internally heterogeneous as regards the role of space and may need to be further investigated.

For these reasons, the distance ranges are modified in order to better suit the suggestions coming from data: distance w_3 is further split in three sub-classes, while distance w_5 and w_6 are aggregated, since the two corresponding partial entropy values are nearly identical and classes have small weights. The results obtained according to the second option are reported in Table 4. It shows how the second option for the distance classes is more suitable for the dataset under study. Former class w_3 is divided into one small class, similar to w_1 and w_2, one intermediate class and one wide class (new classes w_3, w_4, w_5). This way, a smoother trend in the spatial partial information terms and spatial partial residual entropies may be appreciated, showing that space plays a relevant role also at

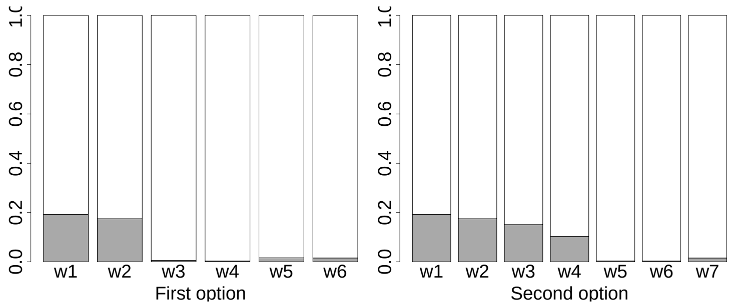

Fig. 9 Partial information (grey areas) and partial residual entropies (white areas) in proportional terms for the two distance options.

distances greater than 2. Moreover, aggregating the former very similar w_5 and w_6 into a new w_7 does not lead to a loss of information.

A comparison of the two options for the distance classes is in Figure 9: at each distance class, the sum $PI(Z|w_k)+H(Z|w_k)$ is set to 1, so that the contribution of space may be appreciated in proportional terms and is comparable. It is immediate to see that space explains one fifth of the data entropy at short distances, and that the second option is better in this context as it allows to grasp the gradual decrease of the role of space as distance increases. Therefore, it can be concluded that spatial association plays a relevant role in explaining the behaviour of land use data in Emilia Romagna up to a distance of $d = 15$ pixels.

The above considerations constitute an example of how scientists should proceed when studying the spatial entropy of specific datasets: the initial choice for the distance classes may be refined as wished, until reaching a satisfactory conclusion, without affecting the global terms.

5.2 A comparison to O'Neill's and Leibovici's entropies

The latter option for categorizing W is used for comparison to O'Neill's and Leibovici's entropy.

O'Neill's entropy is $H(Z|w_{(01)}) = 1.563$ and is comparable to $H(Z|w_1) = 1.421$ (Table 4) except for order preservation. As explained in Sections 3.2.1 and 3.2.2, the value 1.563 has major interpretation limitations. It is greater than $H(Z|w_1)$ because, due to order preservation, the number of categories of Z is higher; therefore, this does not necessarily detect a greater residual entropy. Because of the non-uniqueness of Shannon's entropy of Z when order is preserved, a quantification of the information due to space is not possible. Besides, O'Neill's entropy gives no information available about distances greater than 1. In addition, no measure of the contribution of space is provided, while the range-occurrence approach measures it as $PI(Z|w_1) = 0.338$ in Table 4.

Leibovici's entropy is computed with $d = 15$, since this value collects distances at which partial spatial information is greater than 0.1 (up to w_4 in Table 4). The entropy is $H(Z|w_{(0d)}) = 1.907$. The possibility of choosing d makes this index more flexible, thus preferable to O'Neill's entropy. However, all the above considerations concerning interpretation limitations hold here, irrespective of the chosen d. This result is hard to interpret and is aggregate, without the possibility to split it for deeper inspection.

6 Discussion and conclusions

This paper presents innovations, substantial add-ons and discussion about a set of spatial entropy measures illustrated in Section 3. The range-occurrence approach allows to exploit the full probabilistic framework provided in Information Theory and to satisfy several crucial features.

The first innovation of this paper is methodological: the approach, originally proposed for pairs of realizations of X (Altieri et al 2017), is now extended to a general degree of co-occurrences, i.e. sets of three, four and so on. The method is extremely flexible and suitable for many applications. The idea initially proposed by Leibovici (2009) is here modified to discard the spatial order within co-occurrences and is employed for decomposing Shannon's entropy of Z and interpreting spatial mutual information.

The second innovation is the illustration of both theoretical and practical advantages of discarding the spatial order within co-occurrences. We show how the standard way of computing spatial entropy based on the transformed variable Z implies order preservation within co-occurrences and has major disadvantages, in particular the non-uniqueness of Shannon's entropy of Z with consequent interpretability limitations. The choice of discarding order is reasonable in spatial analysis, where the interest lies in understanding the spatial heterogeneity of data over a specific area, while spatial phenomena are not usually assumed to have a direction. Besides, neglecting the order improves the ability to recognize spatial patterns, while standard indices, based on order preservation, are not usually able to detect the presence of a negative spatial association. Moreover, when order is discarded, Shannon's entropy of Z does not depend on any spatial configuration and is unique. Thus, it can be used as a reference value for interpreting spatial mutual information and spatial residual entropy.

Results of the present paper highlight the difference between the main approach presented here and O'Neill's and Leibovici's approach. The latter is well established in the literature and constitutes an alternative option to the theory of Section 3; based on results in our work, we recommend to choose between such indices with care and awareness about the consequences in the results.

As a third contribution, a novel detailed procedure is proposed for building the variable Z for a generic degree of co-occurrences. Traditional spatial statistical tools are exploited, such as adjacency matrices, and helpful guidelines are provided for practical work. The procedure sheds light on building Z, which was a black box in previous papers.

In addition, a comparison to standard spatial entropy measures (O'Neill et al 1988; Li and Reynolds 1993; Leibovici 2009; Leibovici et al 2014; Parresol and Edwards 2014) is carried out, both in commenting the theory and in the simulation study. All these measures are affected by the choice of the distance range and do not allow any deeper inspection. Conversely, the property of additivity of the range-occurrence approach enriches interpretation and constitutes a substantial theoretical improvement: the partial terms are very flexible in identifying the most informative distances to explain the phenomenon under study. When global spatial mutual information and spatial residual entropy are computed, probabilities of range-occurrences at different distance classes, $p(z_r|w_k)$, are weighted by $p(w_k)$, the probability of each distance class, so that the relative weight of all distances is respected. Therefore, the definition of equal-sized distance classes with constant probabilities is not required. The categories of W must be proposed according to the context, as the less interesting distances should be aggregated while the most relevant ones ought to be considered with more detail.

A further contribution, that can be appreciated from the simulation study, is to validate the set of measures via their performance. Results from Section 4 show that spatial mutual information and spatial residual entropy represent a complete and powerful statistical tool to study the heterogeneity of spatial data: they are able to correctly quantify the contribution of space to the entropy of a dataset for any spatial configuration and any number of categories of the study variable. Moreover, they allow easy interpretability and delivery of results.

Lastly, the methodology is here applied to land use data: at the global level, the contribution of space to the computed entropy is very low. The decomposition into partial terms allows to identify distances which are relevant in understanding the data behaviour, where the partial information terms are greater than zero. The role of space decreases smoothly as distance increases, where most of the data heterogeneity is due to other sources. A comparison to standard spatial entropy measures (O'Neill et al 1988; Li and Reynolds 1993; Leibovici 2009; Leibovici et al 2014; Parresol and Edwards 2014), each only producing a single number, once more shows the enrichments of the range-occurrence approach. The application presented in Section 5 provides a complete toolbox for analyzing spatial data where distance is believed to play a role in determining the heterogeneity of the outcomes. The first step consists in computing Shannon's entropy of Z as a reference value. Secondly, spatial mutual information is computed and its proportional version identifies the overall role of space. Alternative options for the distance classes can be suitably defined in order to investigate the partial information terms, which help to understand whether space plays a relevant role at each distance class, while spatial partial residual entropies 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 heterogeneity of the study variable.

Acknowledgements This work is developed under the PRIN2015 supported project 'Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT)' [grant number 20154X8K23] funded by MIUR (Italian Ministry of Education, University and Scientific Research).

References

Altieri L, Cocchi D, Roli G (2017) A new approach to spatial entropy measures. Environ Ecol Stat published online at http://link.springer.com/article/10.1007/s10651-017-0383-1, 1-16
Anselin L (1995) Local indicators of spatial association - LISA. Geogr 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
Bondy JA, Murty USR (2008) Graph Theory. Springer, New Yotk

Butera I, Vallivero L, Ridolfi L (2018) Mutual information analysis to approach nonlinearity in groundwater stochastic fields. Stochastic Environmental Research and Risk Assessment URL `https://doi.org/10.1007/s00477-018-1591-4`

Claramunt C (2005) A spatial form of diversity. In A.G. Cohn and D.M. Mark (Eds.). COSIT 2005, LNCS 3693, 218-231

Cover TM, Thomas JA (2006) Elements of Information Theory. Second Edition. John Wiley & Sons, Inc., Hoboken, New Jersey

Cressie N (1993) Statistics for spatial data. Revised Edition. John Wiley & Sons, Inc., Hoboken, New Jersey

Frosini BV (2004) Descriptive measures of ecological diversity. In Environmetrics. Edited by J. Jureckova, A. H. El-Shaarawi in Encyclopedia of Life Support Systems (EOLSS), revised edn. 2006. Developed under the Auspices of the UNESCO, Eolss Publishers, Paris, France. http://www.eolss.net

He J, Kolovos A (2018) Bayesian maximum entropy approach and its applications: a review. Stochastic Environmental Research and Risk Assessment 32:859–877

Hoeting JA, Leecaster M, Bowden D (2000) An improved model for spatially correlated binary responses. J Agric Biol Environ Stat 5:102–114

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

Leibovici DG (2009) Defining spatial entropy from multivariate distributions of co-occurrences. In K. S. Hornsby et al. (eds.): COSIT 2009, Lecture Notes in Computer Science 5756, pp 392-404, Berlin, Springer

Leibovici DG, Birkin MH (2015) On geocomputational determinants of entropic variations for urban dynamics studies. Geogr Anal 47:193–218

Leibovici DG, Claramunt C, LeGuyader D, 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

Leinster T, Cobbold CA (2012) Measuring diversity: the importance of species similarity. Ecology 93(3):477–489

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

O'Neill RV, Krummel JR, Gardner RH, Sugihara G, Jackson B, DeAngelis DL, Milne BT, Turner MG, Zygmunt B, Christensen SW, Dale VH, Graham RL (1988) Indices of landscape pattern. Landscape Ecol 1(3):153–162

Paninski L (2003) Estimation of entropy and mutual information. Journal of Neural Computation 15:1191–1253

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

Patil GP, Taillie C (1982) Diversity as a concept and its measurement. J Am Stat Assoc 77:548–561

Riitters KH, O'Neill RV, Wickham JD, Jones KB (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(3):379–423 and 27(4):623–656

Theil H (1972) Statistical Decomposition Analysis. North Holland, Amsterdam