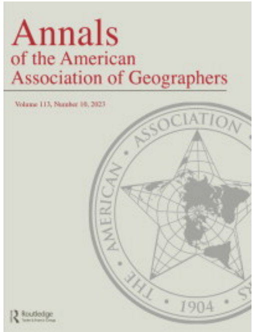

Annals of the American Association of Geographers

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/raag21

Information Consistency-Based Measures for Spatial Stratified Heterogeneity

Hexiang Bai, Hui Wang, Deyu Li & Yong Ge

Information Consistency-Based Measures for Spatial Stratified Heterogeneity

Hexiang Bai,* Hui Wang,* Deyu Li,* and Yong Ge†

*School of Computer and Information Technology, Shanxi University, China
†State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, China

As a typical form of spatial heterogeneity, spatial stratified heterogeneity is widely observed in geographical phenomena. Although the *q* statistic provides a measure of spatial stratified heterogeneity using variance differences, it is not suitable for nominal target variables and neglects information differences between strata at higher order moments. Based on the mutual information and relative entropy between variables, two spatial stratified heterogeneity measures are proposed for nominal and continuous target variables, respectively. Permutation tests are then used to determine their statistical significance. The proposed measures are suitable for either nominal or continuous target variables. They make no assumptions regarding the distribution of target variables, and return a value of zero only when the distribution of the target variable is independent of the explanatory variable. Experiments on five illustrative data sets and three publicly accessible data sets show that the proposed measures are consistent with the *q* statistic and can detect the existence of spatial stratified heterogeneity when the *q* statistic fails, so long as there are significant differences between the distributions in different strata. *Key Words: mutual information, relative entropy, spatial heterogeneity, spatial stratified heterogeneity.*

Spatial dependency and spatial heterogeneity are two closely related important properties of geographic-related phenomena and have been recognized as two laws of geographical information sciences (Cliff and Ord 1973, 1981; Goodchild 2003). They reveal patterns hidden in spatial data from two perspectives. Spatial dependency indicates that "nearby things are more related than distant things" (Tobler 1970), and spatial heterogeneity refers to the variation of geographical phenomena over space (H. Li and Reynolds 1995). Currently, lots of methods have been developed for detecting and measuring spatial dependency, such as Moran's *I*, Getis's G, and join count statistics (Getis and Ord 1992; Moran 1948; Kabos and Csillag 2002). In addition, spatial dependency has been introduced into statistical models for the accurate modeling of geographical phenomena (Anselin 1988; Goovaerts 1997; Fotheringham, Charlton, and Brunsdon 1998; Griffith and Paelinck 2011; Zolnik 2021; Zhang 2023).

Spatial heterogeneity is also an important topic in analyzing geospatial data, such as populations, communities, ecosystems, and landscapes (Shaver et al. 2005), and occurs in different forms. For example, the variation of a qualitative or quantitative value over space (Dutilleul and Legendre 1993), which can be further classified as tendency and anisotropy (Longley and Tobón 2004; Ge et al. 2019), is important in identifying surface patterns. Spatial stratified heterogeneity (SSH) refers to the pattern variations in different strata or areas (J. Wang et al. 2010; Mobley et al. 2012; J. F. Wang, Zhang, and Fu 2016). Generally, the presence of spatial dependency indicates aggregation of similar things in space, which in turn could lead to variation of values over space or different strata; that is, the presence of spatial heterogeneity. Therefore, lots of local statistics originating from spatial dependency have been used to measure the spatial heterogeneity over space, such as local indicators of spatial association (LISA; Anselin 1995), local indicators for categorical data (LICD; Boots 2003; Bai et al. 2016), variogram (Goovaerts 1997), and landscape gradient and patchiness (Sokal and Oden 2008a, 2008b). SSH is also closely related to spatial dependency. Strong spatial dependency could lead to similar pattern aggregation in each stratum. The local spatial dependency indexes, however, mainly focus on the

Annals of the American Association of Geographers, 113(10) 2023, pp. 2512–2524
Initial submission, December 2022; revised submissions, February and April 2023; final acceptance, May 2023
Published by Taylor & Francis, LLC.

tendency or anisotropy of geographical phenomena and do not take into account another important factor of SSH, the stratification of the study area.

SSH is indispensable in analyzing geographic-related phenomena. J. F. Wang, Zhang, and Fu (2016) summarized the necessity of SSH in spatial analysis. First, global measures such as the mean and variance of the entire study area might neglect the differences between subareas, which can result in biased or even incorrect explanations and predictions (Schwanghart, Beck, and Kuhn 2008) and is also an important issue that can be addressed from the perspective of local value variation over space using spatial local dependency indexes (Boots 2003; Anselin 1995). Second, SSH can be used to find possible causal association between the explanatory variable and target variable by inspecting the consistency of the patterns of the target variable in the strata formed by the explanatory variable (Gustafson 1998). Finally, neglecting SSH could result in valuable information that is important to the correct modeling and understanding of the target geographical phenomena being overlooked (Dutilleul 2011; Krzyzanowski and Manson 2022). Therefore, it is important to test for the existence of statistically significant SSH before further examining the mechanisms hidden within phenomena.

Currently, the existence of SSH can be detected using the homogeneity of variance test for multiple groups, such as Bartlett's test (Bartlett 1937) and Levene's test (Levene 1960), from the perspective of inhomogeneous variance in each stratum. To further quantify the degree of SSH besides testing its significance, J. F. Wang, Zhang, and Fu (2016) developed the q statistic based on Levene's test, which has been applied in many real-life applications (Luo et al. 2019; L. Li et al. 2022; Ye et al. 2022). The q statistic uses the ratio between the variance within each stratum and the pooled variance of an entire study area to quantify SSH (J. Wang et al. 2010); that is, $q = 1 - \frac{\sum_{i=1}^{l} N_i \sigma_i^2}{N\sigma^2}$, where l is the number of strata, N_i and N are the numbers of observations in the ith stratum and the whole study area, and σ_i and σ are the standard deviations of the ith stratum and the whole study area. A larger q indicates a larger difference between the within-strata variance and pooled variance and makes it more likely that the SSH is significant.

The q statistic has two limitations when applied in exploratory spatial data analysis, however. First, it cannot be used to deal with nominal target variables. Nominal variables are categorical variables with two or more classes lacking any intrinsic or clear ordering commonly used to represent geographical phenomena. Human concepts are often explained by nominal quantities or classifications (Womble 1951), rather than by quantities, such as land use (Ye et al. 2022), soil type (Trepanier, Pinno, and Errington 2021), and lithology type (Jiang et al. 2021). As it is difficult to define the variance for nominal target variables, the q statistic is not appropriate in such situations. Second, the q statistic only quantifies the SSH using variance differences between distributions in different strata. It cannot detect the SSH related to other characteristics of probability distributions. Later we give an example in which SSH is not detected when using the q statistic.

To address these limitations, it is necessary to develop new measures for SSH. The essence of measuring SSH involves quantifying the degree of consistency between the strata formed by explanatory variables and the target variable (J. F. Wang, Zhang, and Fu 2016). Strong consistency indicates that the stratification (a nonoverlapping division of the study area) determines the target variable's pattern to a large extent; that is, similar patterns are observed within each stratum. In such situations, the patterns between strata tend to have a large discrepancy in most cases and there exists strong SSH, or else a large discrepancy of patterns will be observed within strata, except when the target variable is homogeneous across the whole study area. Taking the q statistic as an example, a large q statistic indicates not only strong SSH but also small within-strata variance; that is, similar values in each stratum and strong consistency between the strata and the target variable. This strategy has also been adopted by the spatial tessellation stratified sampling for identifying the effective sample size (Stevens and Olsen 2004; Griffith 2005). Besides variance, the consistency between explanatory variables and target variables can be quantified from different perspectives. For example, from the perspective of the amount of information that the stratification contains about the target variable (Cover and Thomas 2006), information theory provides an effective solution for measuring consistency between variables using entropy-based measures (Foithong, Pinngern, and Attachoo 2012; Tian et al. 2013), which has also been used to quantify the uncertainty of spatial data (Batty 1974, 2010; Xiao 2021) and enhanced through introducing spatial dependency for spatial data Griffith, Chun, and Hauke (2022).

In all entropy-based measures, both the mutual information and relative entropy are effective consistency measures (Finn 1993; Zhao et al. 2015; Keuper and Brox 2016). The mutual information interprets the consistency between explanatory variables and target variables as the reduced uncertainty of the target variable given some knowledge of the explanatory variables (Cover and Thomas 2006). Some researchers have extended mutual information using spatial entropies (B. Wang, Wang, and Chen 2012). A greater reduction in uncertainty implies better consistency between the explanatory variables and target variables. This measure has been used in evaluating the consistency between thematic maps (Finn 1993), checking the temporal consistency between frames (Keuper and Brox 2016), and selecting features that preserve the degree of consistency (Foithong, Pinngern, and Attachoo 2012; Tian et al. 2013). The relative entropy measures the distance between two distributions, and has been successfully applied in measuring statistical consistency between sea clutter (Zhao et al. 2015), matching multimodal data (Chen et al. 2021), and integrating features from multiple views in image processing (H. Wang et al. 2017). In addition, Heikkila and Hu (2006) extended the relative entropy by considering scale and resolution effects in spatial data. Small values of the relative entropy indicate better consistency between two distributions. Generally, the degree of consistency can be increased by maximizing the mutual information (Yuan et al. 2021) or minimizing the relative entropy (Xue et al. 2009; H. Wang et al. 2017).

Although these two measures are essential in many machine learning theories and have been widely used in processing spatial data, further adaptation is needed for detecting and measuring SSH. An important issue is to determine whether the measured values calculated from data occur by chance, which can avoid the potential influence of latent spatial dependency (Griffith 2000, 2005; Griffith and Plant 2022) as latent spatial dependency could lead to pattern variations between strata. The values of these measures only show the degree of consistency between the strata (partitions) formed by conditions and the target variable. Hypothesis tests are also needed to clarify whether the strata significantly reduce the inconsistency of the target variable. In addition, these two measures should be normalized to [0, 1], whereas 0 represents no SSH and 1 represents full SSH, to simplify the interpretation and compare the degree of SSH.

By introducing permutation tests (Costanzo 1983; Lahiri 2003; Wasserman 2010) into information theory-based measures, this article describes the development of two new SSH measures from a consistency perspective to address the two limitations of the q statistic. The first measure uses the Shannon mutual information to measure the consistency degree between strata formed by explanatory variables and a nominal target variable. This is normalized to [0, 1] to quantify SSH for the nominal target variable. The second measure averages the normalized relative entropy between the density functions in each stratum and in the whole study area to quantify the SSH for a continuous target variable. For both measures, permutation tests are used to determine whether the measure is statistically significantly different from the value given by a random reshuffle of the target variable. Experiments on five illustrative data sets and three publicly accessible data sets show that the proposed measures are effective in quantifying SSH. In addition, the results from the proposed measures are consistent with those using the q statistic and the proposed measures have three advantages:

1. The proposed measures are suitable for both continuous and nominal variables, whereas the q statistic can only deal with continuous variables.
2. No assumptions regarding the data distribution of the target variable are required.
3. The proposed measures can detect the existence of SSH ignored by the q statistic, so long as the explanatory variable is not independent of the target variable.

In the next section, we formalize the stratification of measurements using the concept of an information table, which encodes the stratum and the target variable value for each observation in the form of a table. Two measures are then proposed based on mutual information to measure SSH for nominal and continuous variables, collectively referred to as the extended mutual information. Finally, the results of comparison experiments on five illustrative data sets and three publicly accessible data sets are presented to demonstrate the effectiveness of the proposed measures.

Method

Conceptually, when there exists a large discrepancy between the probability distribution of the target variable in each stratum and its global probability distribution, the stratification determines the target variable to a great extent and the stratification is strongly consistent with the target variable. In such a situation, it is impossible to observe homogeneous probability distribution of the target variable in all strata. As a result, the spatial distribution varies in different strata; that is, there exists SSH. For nominal target variables, the mutual information can quantify the difference of probability distribution from the perspective of information reduced when introducing the stratification (Cover and Thomas 2006). For continuous target variables, the difference between the within-strata probability distribution and global probability distribution is commonly quantified using relative entropy (Theodoridis 2020).

Through normalizing these two measures to interval [0,1], two new measures are developed for nominal and continuous variables, respectively. In the following, we first clarify and define important terms such as the target variable and stratification. Next, all the information needed in quantifying the SSH of the target variable is organized using an information table. Finally, measuring and detecting SSH for nominal and continuous target variables is introduced. In addition, a glossary of symbols is provided in the Appendix to facilitate the understanding of the formulas.

To begin with, a target variable D is the variable corresponding to the observed feature of interest in the study area. For example, the annual precipitation, the annual vegetation production, and the soil type in the study area are all possible target variables. Generally, multiple observations of the feature are collected in different locations in the study area to analyze the spatial distribution of the feature in general.

Next, to inspect the SSH of D, the study area is divided into several nonoverlapping subareas. Formally, a stratification including k strata (subareas) $S = \{s_1, ..., s_k\}$ can be defined as a nonoverlapping subdivision of the study area, and all the strata in the stratification comprise the whole study area. The stratification can be constructed using administrative subdivisions or other categorical variables according to application requirements; for example, the stratification of a country using its provincial-level administrative divisions. Then, each province is a stratum, and all the provinces constitute the country again. In some situations, the stratification can also be constructed using climate zones (J. F. Wang, Zhang, and Fu 2016) and land-cover types (McRoberts et al. 2002). Correspondingly, the climate zones or the areas with different land-cover types have no overlapping areas, and the union of these areas with different climate types or land-cover types equals to the study area again. Additionally, the variables that are used to construct a stratification are referred to as explanatory variables. For example, the land-cover types and climate zones in the previous example are both explanatory variables.

For convenience, an information table is used to assemble all the information required in detecting SSH in the study area. Supposing N observations are collected in the study area, an information table is a table with N rows and two columns. Each row in an information table represents an observation of the target variable. The first column represents the stratum where each observation is collected, and the second column is the feature value for each observation. Formally:

Definition 1. An information table is a triple $\{U, S, D\}$, where U is the set of all observations, $S = \{s_1, ..., s_k\}$ is the k strata of the study area, D is the target variable with domain V_D, and $S(u)$ and $D(u)$ are the mappings from U to the stratum and target variable value of $u \in U$, respectively.

For example, the annual precipitation from 100 meteorological observation stations was collected to inspect the SSH of the annual precipitation in a province. Here, the study area is the province, which is constituted by six prefectures, the annual precipitation is the target variable D with domain $V_D = [0, +\infty)$ and U is the set of all the 100 observations, and $D(u)$ is the annual precipitation (e.g., 530 mm) of the observation $u \in U$. Suppose the study area was divided into six strata according to the prefectures. Then, $S = \{s_1, ..., s_6\}$ is the set of these six strata where s_1 to s_6 can be the name of the six prefectures, and $S(u) = s_1$ represents that the observation $u \in U$ was collected in the stratum or prefecture s_1. These constitute an information table as is shown in Table 1.

Extended Mutual Information for SSH

Given an information table $\{U, S, D\}$, where $S = \{s_1, ..., s_k\}$, for simplicity, denote d and s as the random variables corresponding to the stratum and the target variable value for an observation, respectively. When D is a nominal target variable, the entropy of

Table 1. An example information table that has six strata and 100 observations

Stratum	Annual precipitation
s_1	530 mm
s_5	323 mm
s_3	137 mm
…	…
s_6	800 mm

d is $H(d) = -\sum_{x\in V_d} p(x)\log p(x)$, and the entropy of d conditioned to s is $H(d|s) = -\sum_{s_i\in s}\sum_{x\in V_D} p(s_i, x)\log p(x|s_i)$, where $p(x)$ is the probability of observing x in U, $p(s_i, x)$ is the probability of observing s_i and x in U, and $p(x|s_i)$ is the probability of observing x given that the stratum is s_i.

For example, suppose that D is a nominal variable with two categories x_1 and x_2 and the probabilities of these two categories are both 0.5. Then $H(d) = -\log(0.5)$. Assume there are only two strata s_1 and s_2 and each stratum contains half of the observations after a stratification. The probability of observing these two categories in s_1 is 0.7 and 0.3, respectively; that is, $p(x_1|s_1) = 0.3$ and $p(x_2 s_1) = 0.7$. The probability of observing these two categories in s_2 is 0.3 and 0.7, respectively; that is, $p(x_1 s_2) = 0.7$ and $p(x_2 s_2) = 0.3$. Then, $p(s_1, x_1) = p(s_1) \times p(x_1 s_1) = 0.15$. Similarly, it is easy to get that $p(s_1, x_2) = 0.35$, $p(s_2, x_1) = 0.35$ and $p(s_2, x_2) = 0.15$. As a result, $H(ds) = -0.15 \times \log(0.3) - 0.35 \times \log(0.7) - 0.35 \times\log(0.7) - 0.15 \times \log(0.3) = -0.3 \times \log(0.3) - 0.7 \times \log(0.7)$.

The mutual information between d and s is $I(d, s) = H(d) - H(ds)$ (Cover and Thomas 2006), measuring the amount of information (in units of nats) of the target variable reduced by the stratification (Cover and Thomas 2006). A larger mutual information gives a smaller $H(ds)$. Correspondingly, a small $H(ds)$ indicates a large difference between the probability distribution of d given $s_i \in S$ and the probability distribution of d from the perspective of information reduced, and a large extent to which S determines the probability distribution of d; that is, a high degree of consistency of d within each $s_i \in S$ and a great degree of SSH. In particular, if there exists fully stratified heterogeneity whereby S completely determines D, that is, the value of D is unique given s_i, the mutual information approaches its maximum value $H(d)$. If there is no SSH, the stratification S has no effect on D. Then, $H(ds) = H(d)$ and $I(d, s)$ approaches its minimum 0. This entropy-based measure is normalized as

$$I_N(d, s) = \frac{I(d, s)}{H(d)} = \frac{H(d) - H(ds)}{H(d)} \tag{1}$$

to fit the range and physical meaning of the q statistic. This extended mutual information for nominal target variable I_N has a range of [0, 1], where 0 means that there is no stratified heterogeneity and 1 represents a fully stratified population. Following the previous two-category example, $I_N(d, s) = \frac{\log(0.5) - [0.3\times\log(0.3) + 0.7\times\log(0.7)]}{\log(0.5)}$ in such a situation.

This measure is not suitable for a continuous d, however, because continuous random variables use differential entropy rather than entropy. The differential entropy might be less than zero, and the range of $I_N(d)$ is no longer [0, 1] (Cover and Thomas 2006). As a result, SSH must be quantified in another way for a continuous d.

Given that D is a continuous target variable, the extended mutual information between d and s is defined as

$$I_C(d, s) = \sum_{s_i\in S} p(s_i)\frac{\arctan\left(RelE\left(f_{d_i}||f\right)\right)}{\pi/2}, \# \tag{2}$$

where d_i is the random variable corresponding to the target variable in stratum s_i, and f_{d_i} and f are the density functions of d_i and d, respectively. Additionally, $RelE\left(f_{d_i}||f\right)$ is the relative entropy of f_{d_i} and f. It is easy to find that $I_C(d, s) \geq 0$ and the maximum possible $I_C(d, s)$ is 1.

For example, suppose that there exist two strata s_1 and s_2 in a stratification S and each stratum contains half of the observations, and d has a uniform distribution on the interval [0, 100] in the whole study area. Meanwhile, in s_1, d_1 has a uniform distribution on the interval [0, 50]. In s_2, d_2 has a uniform distribution on the interval [50, 100]. Then, $f(x) = 1/100$ on the interval [0, 100], $f_{d_1} = 1/50$ on the interval [0, 50], and $f_{d_2} = 1/50$ on the interval [50, 100]. Accordingly, $RelE(f_{d_1}||f) = RelE(f_{d_2}||f) = \log(2)$. Then it is easy to get that $I_C(d, s) = 2\arctan(\log(2))/\pi$.

$I_C(d, s)$ quantifies and normalizes the information difference between d and d_i. A large value of $RelE\left(f_{d_i}||f\right)$ in each stratum means that the distance from f_{d_i} to f in that stratum is large and $I_C(d, s)$ is large. This indicates that S determines d_i to a large extent, so there is strong consistency between the S and the D. Specifically, when $I_C(d, s) = 1$, the distance between f and f_{d_i} in each stratum is ∞. This corresponds to full SSH. If $RelE(f_{d_i}|f)$ is small in each stratum, then the distance from f_{d_i} to f in each stratum is small and $I_C(d,s)$ is small. This indicates that S only slightly affects the distribution of d_i in each stratum, and the distribution of the d_i in each stratum is close to f. Accordingly, the heterogeneity between strata is small. Specifically, when $I_C(d,s)=0$, f and f_{d_i} are identical almost everywhere in each stratum, which means the stratification has no effect on D, that is, there is no SSH. More precisely, $I_N(d)$ and $I_C(d)$ have the following properties, respectively.

Property 1. Given an information table $\{U,S,D\}$, and let $S=\{s_1,\ldots,s_k\}$ be a stratification on U. $I_N(d,s)=0$ if and only if d and s are independent of each other.

Proof. If d and s are independent of each other, $I(d,s)=0\Rightarrow I_N(d,s)=0$. Conversely, $I_N(d,s)=0\Rightarrow I(d,s)=0$, which indicates that d and s are independent of each other. □

The independence between d and s indicates that the probability distributions of d in all strata are identical, and there is no SSH. Conversely, no SSH indicates no pattern variation of d in different strata. In turn, the probability distribution of d is independent of the stratification. For instance, when there is no SSH, the probability of observing each of the categories is 0.5 in each stratum in the previous two-category example. Then, $p(s_i,x_i)=p(s_i)\times p(x_i s_i)=0.25$, $i=1,2$, and $H(dS)=-0.25\times\log(0.5)-0.25\times\log(0.5)-0.25\times\log(0.5)-0.25\times\log(0.5)=-\log(0.5)$. As a result $I_N(d,s)=0$.

Property 2. Given an information table $\{U,S,D\}$, and let $S=\{s_1,\ldots,s_k\}$ be a stratification on U. $I_C(d,s)=0$ if and only if $f_d(x)$ and $f_{d_i}(x)$ are identical almost everywhere in each stratum.

Proof. $I_C(d,s)=0\Longleftrightarrow RelE(f_{d_i}||f)=0$ for each $s_i\in S$ $\Longleftrightarrow$ In each stratum s_i, $f_{d_i}(x)$ and $f_d(x)$ are identical almost everywhere. □

Following the previous continuous target variable example where d has a uniform distribution, when d_1 and d_2 both have uniform distribution on the interval [0, 100], which is identical to that of d, then $RelE(d_i||d)=0$ for d_i and d. Subsequently, $I_C(d,s)=0$. In other cases, $I_C(d,s)$ is not zero and quantifies the degree to which s determines d. Conversely, when $I_C(d,s)$ is zero, d_1 and d_2 both have a distribution that is identical to the distribution of d almost everywhere. Obviously, in such situations, the distribution of d_1 and the distribution of d_2 are also identical to each other almost everywhere; that is, there exists no SSH in such situations.

Property 2 shows that $I_C(d,s)=0$ only when d and d_i are identical almost everywhere in each stratum. This property is not true, however, when using the q statistic. Assume that d has a Gaussian distribution $N(0,\ 1)$, and d_1 and d_2 in strata s_1 and s_2 have Gaussian distributions of $N(0,1/100)$ and $N(0,199/100)$. In addition, let the number of measurements in s_1 and s_2 be the same (e.g., 100 in each stratum). The q statistic is then $1-(100\times 1/100+100\times 199/100)/(200)=0$, which implies no SSH. The data in the two strata have large differences, however, because one set is much more widely dispersed than the other. As an alternative, $I_C(d,s)\approx 0.39$, which suggests a medium level of difference. As a result, $I_C(d,s)$ can detect SSH that is ignored by the q statistic.

Although the independence between d and s or the equivalence between d and d_i indicates no SSH from the information consistency perspective according to Properties 1 and 2, it is not an indicator of no spatial dependency. Spatial dependency is closely related to and influences SSH greatly. For example, when there is strong spatial dependency, similar values are more easily observed in the same regions. Stratification plays an important role in identifying SSH, though. For example, a stratification might separate the study area into two regions with the same probability distribution of d as shown in Figure 1. In the whole study area and each stratum, there is strong spatial dependency but no pattern variation in each stratum; that is, no SSH. Clearly, these two properties cannot ensure the existence or absence of spatial dependency. Therefore, quantifying SSH mainly concentrates on the degree to which the stratification determines the target variable, which is useful in revealing the relations between two variables, for example, finding driving factors for evaluating eco-economy coupling coordination (L. Li et al. 2022) and identifying key factors contributing to metal pollution (Luo et al. 2019). No SSH is not a sign of no spatial dependency, and they play different roles in real-life applications.

Permutation Test of Extended Mutual Information

In real-life applications, as well as quantifying SSH, it is important to test the significance of the existence of SSH. To make as few assumptions as

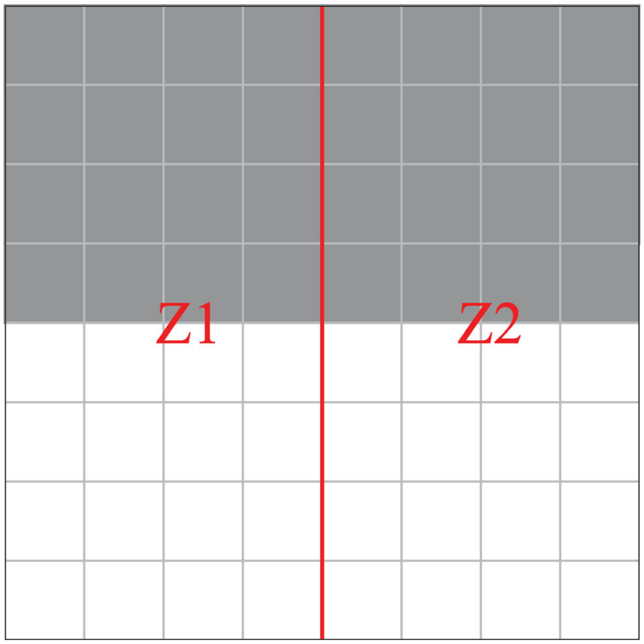

Figure 1. An example of the situation which has no spatial stratified heterogeneity (SSH) but strong spatial dependency.

possible, a permutation test is used to determine the statistical significance of the existence of SSH. Generally, if d_i has the same distribution as d in each stratum s_i, there will be no SSH for the stratification S. For simplicity, such a stratification is referred to as being random in the following. Let H_0 be the null hypothesis that $I_C(d,s)$ or $I_N(d,s)$ is from a random stratification, and H_1 be the alternative hypothesis that $I_C(d,s)$ or $I_N(d,s)$ is not from a random stratification.

Consider j random permutations of the target variable measurements at all positions in the study area. For the ith permutation, the corresponding $I_{Ci}(d,s)$ or $I_{Ni}(d,s)$ is calculated. Under the null hypothesis, each of these values is equally likely. All j extended mutual informations form a reference distribution corresponding to a random stratification. If $I_C(d,s)$ or $I_N(d,s)$ is less than j' of j values, then the p value of rejecting the H_0 hypothesis is $(j'+1)/(j+1)$. If the p value is less than a given threshold, the extended mutual information is statistically significantly different from that calculated from a random stratification. In this case, there exists statistically significant SSH.

Experiments

Illustrative Examples

Six illustrative data sets are used to clarify the procedure and demonstrate the effectiveness of measuring SSH using the extended mutual information. These two measures are implemented using Python version 3.7.9, and the experiments are performed on a computer with an AMD Ryzen 7 4800 Hz CPU and 16 GB memory. The operating system is Ubuntu 20.04. The mutual information and information entropy for nominal variables are calculated using the scikit-learn (Pedregosa et al. 2011) and scipy (Virtanen et al. 2020) packages.

Figure 2 shows the six illustrative data sets. The first to fourth data sets are from J. F. Wang, Zhang, and Fu (2016) to facilitate a comparison with the q statistic. I_N is used for these four data sets, as they all have two categories, black and white. The degree of SSH in these four data sets differs. Figure 2A has no statistically significant SSH. Figure 2B has a small degree of statistically significant SSH at a significance level of 0.05. The third data set has obvious SSH between strata and the fourth data set is fully stratified.

Taking Figure 2C as an example, the extended mutual information between the strata and the target variable is $I(d,s) = 0.4030$ and the entropy of the target variable $H(d)$ is 0.6931. Consequently, $I_N(d,s) = 0.5814$. Finally, the target variable values are reshuffled in all sixty-four cells to perform the permutation test. The p value is less than 0.001. In Figure 2D, the target variable is fully stratified by Z_1 and Z_2. As a result, both q and I_N are one and pass the significance tests. The extended mutual information result is consistent with that from the q statistic, although the q statistic treats the nominal target variable as a continuous one. When there are more than two categories, however, it is hard to calculate the variance. Therefore, the q statistic is no longer appropriate.

The fifth data set in Figure 2E has a continuous target variable. It contains two strata, Z_1 and Z_2, in which the mean of the target variable is 0.004 and 0.0660, respectively. The variance in Z_1 and Z_2 is 1.8868 and 0.0091, respectively. The variance in the whole study area is 0.1839. Obviously, there is significant SSH in this illustrative data set. In terms of Equation 2, the relative entropy $RelE\left(f_{d_{z_1}} \| f\right) = 0.2631$ and $RelE\left(f_{d_{z_2}} \| f\right) = 0.4275$. Therefore, $I_C(d,s) = 0.5 \times \arctan(0.2631)/(\pi/2) + 0.5 \times \arctan(0.4275)/(\pi/2) = 0.2105$. The permutation test shows that the p value is less than 0.001. This result indicates that there is obvious SSH. The q statistic is 2.2744×10^{-7} with a p value of 0.9970158, however. That is,

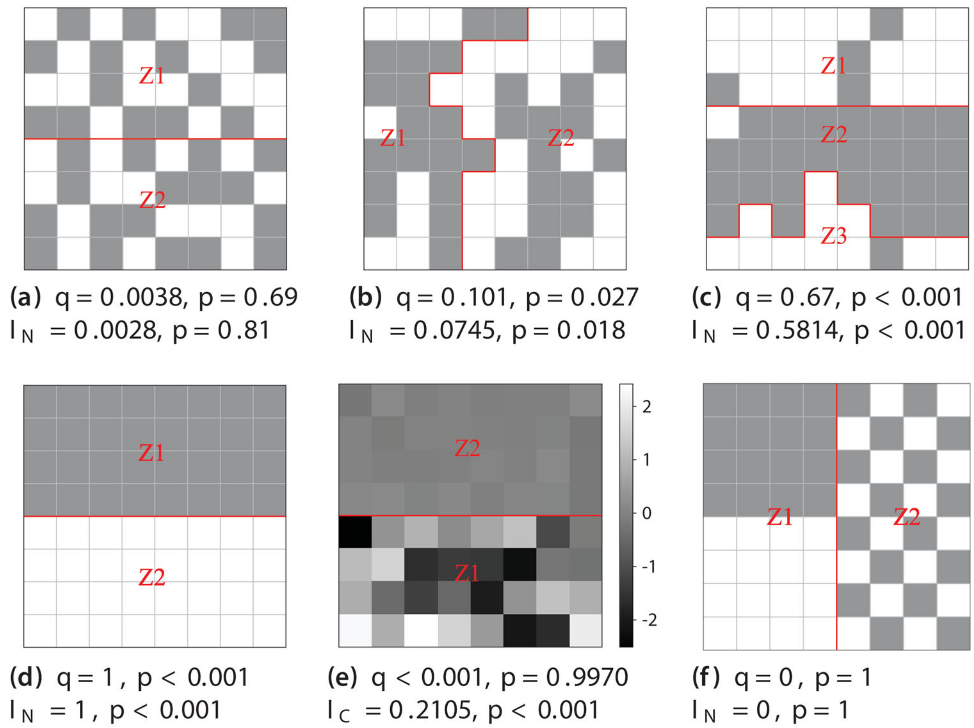

Figure 2. Illustrative examples with different spatial patterns.

the q statistic fails to detect this SSH. In fact, when the within-strata sum of squares is equal to the total sum of squares, the q statistic is not appropriate for detecting the existence of SSH, even if there exists significant SSH.

Properties 1 and 2 ensure that the probability distribution of the target variable in each stratum is the same as the global distribution when the extended mutual information is zero. With respect to the nominal variable, $I_N(d, s) = 0$ indicates that d and S are independent of each other. As a result, $P(d = xs_i \in S) = P(d = x)$ for any $x \in V_d$. For the continuous variable, $I_C(d, s) = 0$ implies that $f_d(x)$ and $f_{d_i}(x)$ are identical almost everywhere in any $s_i \in S$. This means that the density function of d in any strata $s_i \in S$ is the same as the global density function, except in a set with zero measure. Briefly, $f_{d_i}(x) = f_d(x)$ almost everywhere. This ensures that the stratification S introduces no heterogeneity and the target variable is homogeneous in all strata. The q statistic has no such properties, however. Indeed, $q = 0$ only shows that the mean of the target variable in each stratum is the same as the global mean (J. F. Wang, Zhang, and Fu 2016). This does not mean the probability distribution remains unchanged after stratification and ignores higher order moments of the target variable.

Finally, Figure 2F uses an example to show the limitation of both the q statistic and extended mutual information. Clearly, the black and white categories form two separate groups in Z_1 and have a chessboard pattern in Z_2; that is, the spatial configurations between strata are different. Both q statistics and I_N are zero, however, and fail in detecting the heterogeneity of spatial configurations between strata. This is because the q statistic and I_N only quantify SSH from the perspective of the probability distribution changes of the target variable. To inspect the spatial pattern changes between strata, further explanatory analyses that are sensitive in detecting such variation should be undertaken.

Real-Life Examples

Three publicly accessible real-life data sets are now used to illustrate the effectiveness of the extended mutual information. Two of these data sets have nominal target variables and the remaining one has a continuous target variable. Both the q statistic and extended mutual information are used to measure the SSH for comparison. The results show that the extended mutual information is consistent with the q statistic.

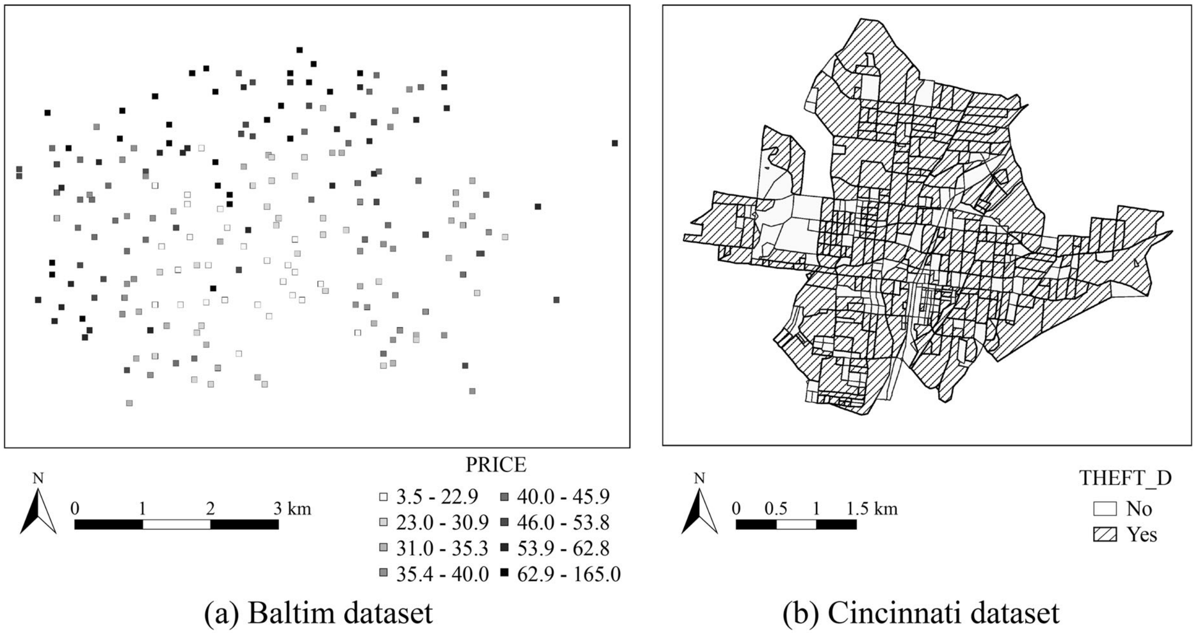

Figure 3. Maps of the Baltim and Cincinnati data sets.

The first data set consists of Baltimore home sale prices and hedonics (Anselin, Syabri, and Kho 2006),[1] which is referred to as Baltim for simplicity. Baltim consists of point-pattern data, as shown in Figure 3A. In total, there are 221 instances. Five attributes of homes are selected to construct the strata of the study area: whether it is a detached unit (DWELL), whether it has a patio (PATIO), whether it has a fireplace (FIREPL), whether it has air conditioning (AC), and whether the dwelling is in Baltimore County (CITCOU). The target variable is the sale price of the home (PRICE). This is a continuous variable. Therefore, I_C is used to detect SSH.

Table 2 presents the degree of SSH as calculated using I_C and the q statistic. The significance level is set to 0.01. Clearly, the results using the extended mutual information are consistent with those from the q statistic. Both measures successfully detect the existence of SSH. The difference is that the two types of measures inspect the SSH using different quantification methods. The q statistic quantifies SSH using within-strata and global variances, whereas I_C measures SSH using the normalized distances between the within-strata and global probability density functions.

Table 2. I_C and q statistic for each attribute of houses in the Baltim data set

Index	DWELL	PATIO	FIREPL	AC	CITCOU
I_C	0.1184	0.0518	0.0895	0.0577	0.1056
q	0.2769	0.2064	0.2760	0.1769	0.1913

The second data set is taken from the 2008 Cincinnati Crime + Socio-Demographics data set,[2] and is referred to as Cincinnati. This data set contains spatial data on an irregular lattice, as shown in Figure 3B. There are 457 objects in the data set. The male population (MALE), female population (FEMALE), median age (MEDIAN_AGE), average family size (AVG_FAMSIZ), and population density (DENSITY) are selected as the explanatory variables. These continuous-valued explanatory variables are discretized into five categories using the equal-width discretization method to construct a stratification. The existence of theft (THEFT_D) is used as the target variable. The target variable is a nominal variable with two categories. Therefore, I_N is used to detect SSH.

Table 3 presents the degree of SSH calculated using I_N and the q statistic for the Cincinnati data set. The significance level is 0.01. For MEDIAN_AGE, AVG_FAMSIZ, and DENSITY, the results from the extended mutual information are consistent with those given by the q statistic. I_N successfully detects statistically significant SSH in the data set for MALE and FEMALE, however, whereas the q statistic detects only insignificant SSH. This is because THEFT_D is intrinsically a nominal variable. The q statistic uses a normal distribution to

Table 3. I_N and q statistic for each explanatory variable of the Cincinnati data set

Index	MALE	FEMALE	MEDIAN_AGE	AVG_FAMSIZ	DENSITY
I_N	0.0400	0.0212	0.0455	0.0613	0.0085[a]
q	0.0370[a]	0.0231[a]	0.0624	0.0831	0.0114[a]

[a]Failed to pass the permutation test.

test its significance, which could lead to misinterpretation. I_N, however, is designed for nominal variables, and successfully detects the significant SSH hidden in the data set from the perspective of the consistency between the explanatory variables and target variables.

The final data set is from the geodetector package in CRAN.[3] We refer to this as Collectdata. The data set has 185 instances. The watershed attribute is used to construct the strata of the study area and soiltype is used as the target variable. As the target variable is a nominal variable with more than two categories, I_N is used to detect SSH. The q statistic is not appropriate for this data set. $I_N = 0.5510$ and the corresponding p value is less than 0.001. Therefore, there is significant SSH, and the degree to which watershed explains the soiltype is 0.5510. When the watershed is used as the target variable and soiltype is used as the explanatory variable, statistically significant SSH also is again observed. $I_N = 0.3887$, however; that is, soiltype explains watershed to a smaller degree. This indicates that $I_N(d, s)$ is not symmetric.

Conclusion and Future Work

SSH plays an important role in spatial analysis. The q statistic has been proven to be an effective tool for quantifying SSH for spatial data. It has two deficiencies, though: It is unable to process nominal target variables and becomes ineffective when the within-strata sum of squares approaches the total sum of squares; that is, between-strata sum of squares is zero (J. F. Wang, Zhang, and Fu 2016). To address these two issues, two new measures for SSH have been proposed by extending mutual information and relative entropy in information theory.

The experimental results presented in this article show that the proposed measures are suitable for both continuous and nominal target variables. In most cases, the normalized mutual information is consistent with the q statistic. I_N, however, is better suited for dealing with nominal variables and is more effective in detecting SSH in two-category situations than the q statistic, because the latter treats two-category variables as a continuous variable approximately. In addition, I_C is effective in identifying statistically significant SSH for continuous target variables when the q statistic approaches 0, as illustrated in Figure 2E.

Although the extended mutual information effectively quantifies the SSH of target variables, it might require further reinforcement to satisfy the requirements of real-life applications. For example, continuous-valued explanatory variables are discretized into different categories regardless of whether the q statistic or extended mutual information is used. This will lead to information loss, and inevitably influences the qualification of SSH. In future work, it is important to explore appropriate measures for processing continuous-valued explanatory variables with no information loss. Additionally, SSH should be inspected from other perspectives. As is shown in Figure 2F, the spatial configuration could have significant differences between strata besides the probability distribution of the target variable. It is instructive to develop new SSH measures for detecting such heterogeneity between strata for analyzing and modeling underlying geographic processes accurately.

Notes

1. See https://geodacenter.github.io/data-and-lab/baltim/.
2. See https://geodacenter.github.io/data-and-lab/walnut_hills/.
3. See https://cran.r-project.org/web/packages/geodetector/index.html.

Funding

This work is supported by the National Natural Science Foundation of China (Grant Nos. 41871286, 42230110, 62072294, 41725006).

References

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Kluwer Academic.

Anselin, L. 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27 (2):93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x.

Anselin, L., L. Syabri, and Y. Kho. 2006. Geoda: An introduction to spatial data analysis. *Geographical Analysis* 38 (1):5–22. doi: 10.1111/j.0016-7363.2005.00671.x.

Bai, H., D. Li, Y. Ge, and J. Wang. 2016. Detecting nominal variables spatial associations using conditional probabilities of neighboring surface objects' categories. *Information Sciences* 329:701–18. doi: 10.1016/j.ins.2015.10.003.

Bartlett, M. S. 1937. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London: Series A. Mathematical and Physical Sciences* 160 (901):268–82.

Batty, M. 1974. Spatial entropy. *Geographical Analysis* 6 (1):1–31. doi: 10.1111/j.1538-4632.1974.tb01014.x.

Batty, M. 2010. Space, scale, and scaling in entropy maximizing. *Geographical Analysis* 42 (4):395–421. doi: 10.1111/j.1538-4632.2010.00800.x.

Boots, B. 2003. Developing local measures of spatial association for categorical data. *Journal of Geographical Systems* 5 (2):139–60. doi: 10.1007/s10109-003-0110-3.

Chen, W., Y. Liu, E. M. Bakker, and M. S. Lew. 2021. Integrating information theory and adversarial learning for cross-modal retrieval. *Pattern Recognition* 117:107983. doi: 10.1016/j.patcog.2021.107983.

Cliff, A. D., and J. K. Ord. 1973. *Spatial autocorrelation*. London: Pion.

Cliff, A. D., and J. K. Ord. 1981. *Spatial processes: Models and applications*. London: Pion.

Costanzo, C. M. 1983. Statistical inference in geography: Modern approaches spell better times ahead. *The Professional Geographer* 35 (2):158–65. doi: 10.1111/j.0033-0124.1983.00158.x.

Cover, T. M., and J. A. Thomas. 2006. *Elements of information theory*. Hoboken, NJ: Wiley-Interscience.

Dutilleul, P. 2011. *Spatio-temporal heterogeneity: Concepts and analysis*. Cambridge, UK: Cambridge University Press.

Dutilleul, P., and P. Legendre. 1993. Spatial heterogeneity against heteroscedasticity: An ecological paradigm versus a statistical concept. *Oikos* 66 (1):152–71. doi: 10.2307/3545210.

Finn, J. T. 1993. Use of the average mutual information index in evaluating classification error and consistency. *International Journal of Geographical Information Systems* 7 (4):349–66. doi: 10.1080/02693799308901966.

Foithong, S., O. Pinngern, and B. Attachoo. 2012. Feature subset selection wrapper based on mutual information and rough sets. *Expert Systems with Applications* 39 (1):574–84. doi: 10.1016/j.eswa.2011.07.048.

Fotheringham, A. S., M. E. Charlton, and C. Brunsdon. 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A: Economy and Space* 30 (11):1905–27. doi: 10.1068/a301905.

Ge, Y., J. Yan, A. Stein, Y. Chen, J. Wang, J. Wang, Q. Cheng, H. Bai, M. Liu, and P. M. Atkinson. 2019. Principles and methods of scaling geospatial earth science data. *Earth-Science Reviews* 197:102897. doi: 10.1016/j.earscirev.2019.102897.

Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24 (3):189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x.

Goodchild, M. F. 2003. The fundamental laws of GIScience. Paper presented at the Summer Assembly of the University Consortium for Geographic Information Science, Pacific Grove, CA. Accessed May 7, 2023. http://csiss.ncgia.ucsb.edu/aboutus/presentations/files/goodchild_ucgis_jun03.pdf.

Goovaerts, P. 1997. *Geostatistics for natural resources evaluation*. Oxford, UK: Oxford University Press.

Griffith, D. A. 2000. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2 (2):141–56. doi: 10.1007/PL00011451.

Griffith, D. A. 2005. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers* 95 (4):740–60. doi: 10.1111/j.1467-8306.2005.00484.x.

Griffith, D. A., Y. Chun, and J. Hauke. 2022. A Moran eigenvector spatial filtering specification of entropy measures. *Papers in Regional Science* 101 (1):259–79. doi: 10.1111/pirs.12646.

Griffith, D., and J. Paelinck. 2011. *Non-standard spatial statistics and spatial econometrics*. Berlin, Germany: Springer.

Griffith, D. A., and R. E. Plant. 2022. Statistical analysis in the presence of spatial autocorrelation: Selected sampling strategy effects. *Stats* 5 (4):1334–53. doi: 10.3390/stats5040081.

Gustafson, E. J. 1998. Quantifying landscape spatial pattern: What is the state of the art? *Ecosystems* 1 (2):143–56. doi: 10.1007/s100219900011.

Heikkila, E. J., and L. Hu. 2006. Adjusting spatial-entropy measures for scale and resolution effects. *Environment and Planning B: Planning and Design* 33 (6):845–61. doi: 10.1068/b31126.

Jiang, Y., J. Gao, L. Yang, S. Wu, and E. Dai. 2021. The interactive effects of elevation, precipitation and lithology on karst rainfall and runoff erosivity. *Catena* 207:105588. doi: 10.1016/j.catena.2021.105588.

Kabos, S., and F. Csillag. 2002. The analysis of spatial association on a regular lattice by join-count statistics without the assumption of first-order homogeneity. *Computers & Geosciences* 28 (8):901–10. doi: 10.1016/S0098-3004(02)00007-9.

Keuper, M., and T. Brox. 2016. Point-wise mutual information-based video segmentation with high temporal consistency. In *Computer Vision–ECCV 2016 workshops*, ed. G. Hua and H. Jégou, 789–803. Cham, Switzerland: Springer International.

Krzyzanowski, B., and S. Manson. 2022. Regionalization with self-organizing maps for sharing higher resolution protected health information. *Annals of the American Association of Geographers* 112 (7):1866–89. doi: 10.1080/24694452.2021.2020617.

Lahiri, S. N. 2003. *Scope of resampling methods for dependent data*. New York: Springer.

Levene, H. 1960. Robust tests for equality of variances. In *Contributions to probability and statistics*, ed. I. Olkin, 278–92. Palo Alto, CA: Stanford University Press.

Li, H., and J. F. Reynolds. 1995. On definition and quantification of heterogeneity. *Oikos* 73 (2):280–84. doi: 10.2307/3545921.

Li, L., F. Zhang, F. Wu, Y. Chen, and K. Qin. 2022. Coupling coordination degree spatial analysis and driving factor between socio-economic and eco-environment in northern China. *Ecological Indicators* 135:108555. doi: 10.1016/j.ecolind.2022.108555.

Longley, P. A., and C. Tobón. 2004. Spatial dependence and heterogeneity in patterns of hardship: An intra-urban analysis. *Annals of the Association of American Geographers* 94 (3):503–19. doi: 10.1111/j.1467-8306.2004.00411.x.

Luo, L., K. Mei, L. Qu, C. Zhang, H. Chen, S. Wang, D. Di, H. Huang, Z. Wang, F. Xia, et al. 2019. Assessment of the geographical detector method for investigating heavy metal source apportionment in an urban watershed of eastern China. *The Science of the Total Environment* 653:714–22. doi: 10.1016/j.scitotenv.2018.10.424.

McRoberts, R. E., D. G. Wendt, M. D. Nelson, and M. H. Hansen. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment* 81 (1):36–44. doi: 10.1016/S0034-4257(01)00330-3.

Mobley, L. R., T. M. Kuo, M. Urato, S. Subramanian, L. Watson, and L. Anselin. 2012. Spatial heterogeneity in cancer control planning and cancer screening behavior. *Annals of the Association of American Geographers* 102 (5):1113–24. doi: 10.1080/00045608.2012.657494.

Moran, P. A. P. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)* 10 (2):243–51. doi: 10.1111/j.2517-6161.1948.tb00012.x.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–30.

Schwanghart, W., J. Beck, and N. Kuhn. 2008. Measuring population densities in a heterogeneous world. *Global Ecology and Biogeography* 17 (4):566–68. doi: 10.1111/j.1466-8238.2008.00390.x.

Shaver, G. 2005. Spatial heterogeneity: Past, present, and future. In *Ecosystem function in heterogeneous landscapes*, ed. G. Lovett, M. Turner, C. Jones, and K. Weathers, 443–49. New York: Springer.

Sokal, R. R., and N. L. Oden. 2008a. Spatial autocorrelation in biology: 1. Methodology. *Biological Journal of the Linnean Society* 10 (2):199–228. doi: 10.1111/j.1095-8312.1978.tb00013.x.

Sokal, R. R., and N. L. Oden. 2008b. Spatial autocorrelation in biology: 2. Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society* 10 (2):229–49. doi: 10.1111/j.1095-8312.1978.tb00014.x.

Stevens, D. L., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99 (465):262–78. doi: 10.1198/016214504000000250.

Theodoridis, S. 2020. *Machine learning: A Bayesian and optimization perspective*. London: Academic Press.

Tian, J., Q. Wang, B. Yu, and D. Yu. 2013. A rough set algorithm for attribute reduction via mutual information and conditional entropy. Paper presented at the 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), ed. J. Chen, X. Wang, L. Wang, J. Sun, and X. Meng, 567–71. Piscataway, NJ: IEEE.

Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46 (2):234–40. doi: 10.2307/143141.

Trepanier, K. E., B. D. Pinno, and R. C. Errington. 2021. Dominant drivers of plant community assembly vary by soil type and time in reclaimed forests. *Plant Ecology* 222 (2):159–71. doi: 10.1007/s11258-020-01096-z.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* 17 (3):261–72. doi: 10.1038/s41592-020-0772-5.

Wang, B., X. Wang, and Z. Chen. 2012. Spatial entropy based mutual information in hyperspectral band selection for supervised classification. *International Journal of Numerical Analysis and Modeling* 9 (2):181–92.

Wang, H., L. Feng, X. Meng, Z. Chen, L. Yu, and H. Zhang. 2017. Multi-view metric learning based on kl-divergence for similarity measurement. *Neurocomputing* 238 (C):269–76. doi: 10.1016/j.neucom.2017.01.062.

Wang, J., X. Li, G. Christakos, T. Liao, T. Zhang, X. Gu, and X. Zheng. 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *International Journal of Geographical Information Science* 24 (1):107–27. doi: 10.1080/13658810802443457.

Wang, J. F., T. L. Zhang, and B. J. Fu. 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators* 67 (Suppl. C):250–56. doi: 10.1016/j.ecolind.2016.02.052.

Wasserman, L. 2010. *All of statistics: A concise course in statistical inference*. New York: Springer.

Womble, W. 1951. Differential systematics. *Science* 114 (2961):315–22. doi: 10.1126/science.114.2961.315.

Xiao, J. 2021. Spatial aggregation entropy: A heterogeneity and uncertainty metric of spatial aggregation. *Annals of the American Association of Geographers* 111 (4):1236–52. doi: 10.1080/24694452.2020.1807309.

Xue, X., Q. Shen, H. Li, W. J. O'Brien, and Z. Ren. 2009. Improving agent-based negotiation efficiency in construction supply chains: A relative entropy method. *Automation in Construction* 18 (7):975–82. doi: 10.1016/j.autcon.2009.05.002.

Ye, S., S. Ren, C. Song, C. Cheng, S. Shen, J. Yang, and D. Zhu. 2022. Spatial patterns of county-level arable land productive-capacity and its coordination with land-use intensity in mainland China. *Agriculture, Ecosystems & Environment* 326:107757. doi: 10.1016/j.agee.2021.107757.

Yuan, X., H. Chen, Y. Song, X. Zhao, Z. Ding, Z. He, and B. Long. 2021. Improving sequential recommendation consistency with self-supervised imitation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, ed. Z. Zhou, 3321–27. Montreal, Canada: International Joint Conferences on Artificial Intelligence.

Zhang, W., Y. Ge, H. Bai, Y. Jin, A. Stein, and P. Atkinson. 2023. Spatial association from the perspective of mutual information. *Annals of the American Association of Geographers*. Advance online publication. doi: 10.1080/24694452.2023.2209629.

Zhao, D., H. Lang, X. Zhang, J. Meng, and Laiquan. 2015. Sea clutter modelling by statistical majority consistency for ship detection in SAR imagery. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), ed. K. Sarabandi, V. Pascazio, and S. B. Serpico, 3695–98. Piscataway, NJ: IEEE.

Zolnik, E. 2021. Geographically weighted regression models of residential property transactions: Walkability and value uplift. *Journal of Transport Geography* 92:103029. doi: 10.1016/j.jtrangeo.2021.103029.

HEXIANG BAI [co-corresponding author] is a Professor in the School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, China, 030006. E-mail: baihx@sxu.edu.cn. His research interests include spatial statistics and rough sets theory-based spatial data mining.

HUI WANG is a Postgraduate Student in the School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, China, 030006. E-mail: baihx@sxu.edu.cn. His research interests include knowledge-based recommender systems.

DEYU LI is a Professor in the School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi, China, 030006. E-mail: lidy@sxu.edu.cn. His research interests include data mining and knowledge discovery.

YONG GE [co-corresponding author] is a Full Professor at the Institute of Geographical Science and Natural Resources Research, Chinese Academy of Science, Beijing, China, 100045. E-mail: Gey@lreis.ac.cn. Her research interests include spatial statistics and spatial data science, including machine learning. Applications concern poverty, land-use/land-cover change detection, and scaling Earth science data.

Appendix. List of symbols

q	q value in q statistic
N	Number of observations in the study area.
N_i	Number of observations in the ith stratum
σ	Standard deviation in the study area
σ_i	Standard deviation in the ith stratum
$\sum$	Summation: sum of all values in the specified range
D	Target variable
V_D	Range of D
x	A value in the range of the target variable
d	The random variable of the target variable value for an observation
d_i	The random variable of the target variable value for an observation in the ith stratum
f	The density function of a continuous d
f_{d_i}	The density function of a continuous d in the ith stratum
S	The set of all strata in a stratification
s_i	The ith stratum in a stratification
s	The random variable corresponding to the stratum where an observation is collected
U	The set of all observations
u	An observation in U
$D(u)$	The target variable value of u
$S(u)$	The stratum where u is collected
$+\infty$	Positive infinity
$\in$	A member of; for example, $u \in U$ indicates that u is a member of U
$\log()$	The natural logarithm function
$p(x)$	The probability of observing x
$p(x\|s_i)$	The probability of observing x given the stratum s_i
$p(s_i, x)$	The probability of observing s_i and x
$p(s_i)$	The probability of observing s_i in U
$H(d)$	The entropy of random variable d
$H(d\|s)$	The conditional entropy of d conditioned to s
$I(d, s)$	The mutual information between d and s
$I_N(d, s)$	The normalized mutual information between d and s for nominal target variables
π	Pi
$RelE(f_{d_i}\|\|f)$	The relative entropy of f_{d_i} and f
$\arctan()$	The inverse tangent function
$I_C(d, s)$	The normalized mutual information between d and s for continuous target variables